



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

**GENÉTICA DA CONSERVAÇÃO DE SUÍNOS LOCALMENTE ADAPTADOS NO  
BRASIL: USO DE FERRAMENTAS GENÔMICAS E GEOGRÁFICAS**

**ELIZABETE CRISTINA DA SILVA**

**TESE DE DOUTORADO EM CIÊNCIAS ANIMAIS**

**BRASÍLIA/DF  
NOVEMBRO DE 2014**



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

**GENÉTICA DA CONSERVAÇÃO DE SUÍNOS LOCALMENTE ADAPTADOS NO  
BRASIL: USO DE FERRAMENTAS GENÔMICAS E GEOGRÁFICAS**

**ALUNA: ELIZABETE CRISTINA DA SILVA**

**ORIENTADOR: SAMUEL REZENDE PAIVA, Ph.D.**

**CO-ORIENTADOR: CONCEPTA M. MCMANUS PIMENTEL, Ph.D.**

**TESE DE DOUTORADO EM CIÊNCIAS ANIMAIS**

**PUBLICAÇÃO: 113D/2014**

**BRASÍLIA/DF  
NOVEMBRO DE 2014**

## REFERÊNCIA BIBLIOGRÁFICA E CATALOGAÇÃO

SILVA, E.C da. **Genética da conservação de suínos localmente adaptados no Brasil: uso de ferramentas genômicas e geográficas.** Brasília: Faculdade de Agronomia e Medicina Veterinária, Universidade de Brasília, 2014, 157 p. Tese de Doutorado

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

### FICHA CATALOGRÁFICA

SILVA, Elizabete Cristina da. **Genética da conservação de suínos localmente adaptados no Brasil: uso de ferramentas genômicas e geográficas.** Brasília: Faculdade de Agronomia e Medicina Veterinária, Universidade de Brasília, 2014, 157 p. Tese (Doutorado em Ciências Animais) - Faculdade de Agronomia e Medicina Veterinária. 2014.

1. Filogeografia. 2. Sequenciamento de nova geração. 3. Receptores de sabor. 4. Suínos. I. Paiva, S. R. II. Ph.D.

**CDD ou CDU**  
**Agris/FAO**

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

**GENÉTICA DA CONSERVAÇÃO DE SUÍNOS LOCALMENTE ADAPTADOS NO  
BRASIL: USO DE FERRAMENTAS GENÔMICAS E GEOGRÁFICAS**

**ELIZABETE CRISTINA DA SILVA**

**TESE DE DOUTORADO SUBMETIDA AO  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
CIÊNCIAS ANIMAIS, COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA OBTENÇÃO  
DO TÍTULO DE DOUTOR EM CIÊNCIAS  
ANIMAIS.**

**APROVADA POR:**

---

**SAMUEL REZENDE PAIVA, Ph.D. (EMBRAPA)  
(Orientador)**

---

**ALEXANDRE RODRIGUES CAETANO, Ph.D. (EMBRAPA Recursos Genéticos e  
Biotecnologia)**

---

**UBIRATAN PIOVEZAN, Ph.D. (EMBRAPA Pantanal)**

---

**RINALDO WELLERSON PEREIRA, Ph.D. (Universidade Católica de Brasília)**

---

**MÔNICA CORRÊA LEDUR, Ph.D. (EMBRAPA Suínos e Aves)**

**BRASÍLIA/DF, 13 de NOVEMBRO de 2014**

“De tudo, ficaram três coisas:  
A certeza de que estamos sempre começando...  
A certeza de que precisamos continuar...  
A certeza de que seremos interrompidos antes de terminar...  
Portanto devemos:  
Fazer da interrupção um caminho novo ...  
Da queda um passo de dança...  
Do medo, uma escada...  
Do sonho, uma ponte...  
Da procura, um encontro...”

**Fernando Sabino**

## AGRADECIMENTOS

Ao comitê de orientação, Dr. Samuel Paiva e Professora Dr<sup>a</sup>. Connie McManus, pela oportunidade de realizar esse curso e pelas orientações para o desenvolvimento dessa tese.

Ao pesquisador Dr. Miguel Pérez-Enciso pelos seus ensinamentos, orientações e pela oportunidade de trabalhar no *Centre de Recerca en Agrigenómica* (CRAG), Barcelona, Espanha.

Ao pesquisador Dr. Sebastián Ramos-Onsins pelos ensinamentos e disposição em ajudar com as análises estatísticas dos trabalhos executados no CRAG.

Ao Professor Dr. Eugeni Roura, da Universidade de Queensland, Austrália, pelas contribuições no trabalho sobre os genes de receptores de nutrientes e do gosto.

Ao pesquisador, da Embrapa Pantanal, Dr. Ubiratan Piovezan por fornecer as amostras de suínos Monteiros e as informações necessárias para o estudo de Filogeografia.

Aos meus amigos do Laboratório de Genética Animal da Embrapa Recursos Genéticos e Biotecnologia-EMBRAPA-CENARGEN, em especial, Ronyere Olegário, Luciana Villela e Lilian Cavalcante, pela amizade, conselhos e momentos de distração.

Aos amigos do CRAG Erica Bianco e William Burgos pela amizade e por compartilharem seus conhecimentos comigo.

À UnB e ao Programa de Pós-Graduação em Ciências Animais por dar suporte para realização desta Pós-Graduação.

À Embrapa Recursos Genéticos e Biotecnologia por permitir a realização dessa pesquisa no interior de suas instalações e por fornecer as amostras de DNA de suínos, materiais e equipamentos necessários para esse estudo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de formação durante o curso de doutorado.

Ao Programa Ciência sem Fronteiras-CNPq pela bolsa de Doutorado-Sanduíche, que foi realizado no Centre de Recerca en Agrigenómica (GRAG), em Barcelona, Espanha.

À minha família pela força e carinho sempre que preciso. Ao meu noivo e melhor amigo, Victor Hugo, por sua paciência, confiança e incentivo durante todos esses anos de estudos.

A todos que me ajudaram direta e indiretamente a completar esse estudo, pois tenham certeza que sem a contribuição de cada um seria quase impossível concretizá-lo.

## **BIOGRAFIA**

ELIZABETE CRISTINA DA SILVA, filha de Manoel Otávio da Silva e Nerice Pereira da Silva, nasceu em 04 de dezembro de 1981, em Palmares-PE. Em abril de 2003 iniciou o curso de Zootecnia na Universidade Federal Rural de Pernambuco (UFRPE) concluindo-o em agosto de 2008. Neste mesmo período ingressou no Programa de Pós-Graduação em Zootecnia, área de concentração em Produção Animal, da UFRPE, o qual concluiu em 12 de Agosto de 2010. Em abril de 2011 ingressou no Programa de Pós-Graduação em Ciências Animais da Universidade de Brasília-UnB, Campus Darcy Ribeiro-DF. Submeteu-se a defesa de tese para obtenção do título de doutor em Ciências Animais, na linha de pesquisa Caracterização e Conservação de Recursos Genéticos Animais, em 13 de novembro de 2014.



## SUMÁRIO

ÍNDICE DE TABELAS .....	xi
ÍNDICE DE FIGURAS .....	xiii
RESUMO .....	xv
ABSTRACT .....	xvii
<b>CAPÍTULO 1.....</b>	<b>1</b>
1. INTRODUÇÃO.....	1
1.1 Problemática e relevância.....	3
1.2 Objetivos.....	5
2. REVISÃO DE LITERATURA .....	6
2.1 Importância das raças suínas localmente adaptadas .....	6
2.2 Diversidade genética e uso de ferramentas moleculares .....	8
2.2.1 Microssatélites .....	9
2.1.2 DNA mitocondrial (mtDNA) .....	11
2.1.3 Polimorfismo de nucleotídeo único (SNP).....	12
2.2 Sequenciamento de nova geração (NGS) .....	13
2.3 Filogeografia.....	14
2.4 Genes de receptores de sabor e de nutrientes .....	15
3 REFERÊNCIAS BIBLIOGRÁFICAS .....	16
<b>CAPÍTULO 2 FILOGEOGRAFIA DO SUÍNO MONTEIRO NO PANTANAL DE NHECOLÂNDIA, MATO GROSSO DO SUL, BRASIL.....</b>	<b>26</b>
1. INTRODUÇÃO.....	28
2. MATERIAL E MÉTODOS.....	30
3. RESULTADOS E DISCUSSÃO .....	36
4. CONCLUSÕES .....	48
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	49
<b>CAPÍTULO 3 VARIAÇÃO GENÉTICA DO GENOMA MITOCONDRIAL EM SUÍNOS E JAVALIS .....</b>	<b>53</b>
1. INTRODUÇÃO.....	57
2. MATERIAL E MÉTODOS.....	59
3. RESULTADOS E DISCUSSÃO .....	62

4. CONCLUSÕES .....	72
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	73
<b>CAPÍTULO 4 VARIABILIDADE EM GENES DE RECEPTORES DE NUTRIENTES E DO GOSTO EM SUÍNOS .....</b>	<b>77</b>
1. INTRODUÇÃO.....	79
2. MATERIAL E MÉTODOS.....	81
3. RESULTADOS E DISCUSSÃO .....	85
4. CONCLUSÕES .....	99
5. REFERÊNCIAS BIBLIOGRÁFICAS .....	100
<b>CAPÍTULO 5 .....</b>	<b>103</b>
1. CONCLUSÕES .....	103
<b>ANEXOS .....</b>	<b>105</b>

## ÍNDICE DE TABELAS

<b>CAPÍTULO 2 FILOGEOGRAFIA DO SUÍNO MONTEIRO NO PANTANAL DE NHECOLÂNDIA, MATO GROSSO DO SUL, BRASIL</b> .....	26
<b>Tabela 1</b> Populações, número de suínos monteiros amostrados de cada ponto geográfico e as coordenadas geográficas.....	31
<b>Tabela 2</b> Nome dos loci, cromossomos (Cr.) sequências dos primers e tamanho esperado (Te) obtido neste trabalho, tipo de marcação fluorescente (Marc.) e multiplex (M). .....	34
<b>Tabela 3</b> Parâmetros de diversidade intrapopulacional para o conjunto de 19 locos de microssatélites.....	37
<b>Tabela 4</b> Estrutura genética analisada pela Análise de Variância Molecular (AMOVA). .....	42
<b>Tabela 5</b> Distribuição dos 22 haplótipos por raça ou grupo genético de suíno e por geografia para o fragmento de 363 pb da região D-loop do mtDNA. Os hífen indicam deleções e os pontos indicam ausência de polimorfismo. ....	44
<b>ANEXO A: CAPÍTULO 2</b> .....	105
<b>Tabela A1</b> Quantidades de reagentes utilizados nas reações de PCR e de sequenciamento. ....	105
<b>Tabela A2</b> Parâmetros de diversidade genética analisados em 19 locos de microssatélites nos suínos monteiros do Pantanal. ....	106
<b>Tabela A3</b> Frequência (Freq) dos alelos diagnósticos (AD) por população e locos. ....	107
<b>Tabela A4</b> Designação dos suínos monteiros à população de origem.....	107
<b>Tabela A5</b> Matriz de distância geográfica (km) entre os 10 pontos de coletas representados pelas populações (Pop) amostradas no Pantanal-MS. ....	108
<b>Tabela A6</b> Estimativas de $F_{ST}$ par a par abaixo na diagonal e valores de P acima na diagonal com os 19 locos de microssatélites.....	105
<b>Tabela A7</b> Distâncias genéticas D de Nei (Nei, 1972) com os 19 locos de microssatélites..	109
<b>Tabela A8</b> Proporção de cada população designada à cada cluster para K2 e K4. ....	110
<b>Tabela A9</b> Índices de diversidade genética para as sequências da região controle do mtDNA de suínos de cada população.....	111
<b>Tabela A10</b> Índices de diversidade para as sequências da região controle do mtDNA de suínos para as quatro regiões fisiográficas amostradas. ....	111
<b>Tabela A11</b> Estimativa de fluxo gênico baseado em $F_{ST}$ pelo método pairwise entre as cinco populações deste estudo abaixo na diagonal. Os valores de P estão acima na diagonal e os asteriscos indicam valores significativos ( $P < 0,05$ ). ....	112
<b>Tabela A12</b> Estimativa de fluxo gênico baseado em $F_{ST}$ pelo método pairwise entre as quatro regiões fisiográficas do Brasil abaixo na diagonal. Os valores de P estão acima na diagonal e os asteriscos indicam valores significativos ( $P < 0,05$ ). Análise realizada com a região controle do mtDNA .....	112

<b>CAPÍTULO 3 VARIAÇÃO GENÉTICA DO GENOMA MITOCONDRIAL EM SUÍNOS E JAVALIS</b> .....	53
<b>Tabela 1</b> Parâmetros de diversidade nucleotídica e teste de neutralidade (D de Tajima) por população suína usando sequências de mtDNA completo e região controle. ....	66
<b>Tabela 2</b> Teste de McDonald–Kreitman (MKT) e Índice de Neutralidade (NI) por população suína e por cada gene do mtDNA. Os P-values foram obtidos pelo teste exato de Fisher.....	69
<b>ANEXO B: CAPÍTULO 3</b> .....	114
<b>Tabela B13</b> Valores de distância genética e erro padrão ( $d \pm SE$ ) entre populações obtidos com o DNA mitocondrial completo (abaixo na diagonal) e a região controle (acima na diagonal).....	114
<b>Tabela B14</b> Posição de cada mutação não-sinônima deletéria e frequência nas populações suínas. Essa frequência não incluiu mutações polimórficas ou fixadas no outgroup.....	115
<b>Tabela B15</b> D de Tajima para mutações sinônimas ( $T_s$ ) e não-sinônimas ( $T_a$ ) por gene e por população. Valores em negrito ou sublinhado foram significativos a 5% e 1%, respectivamente. ....	117
<b>CAPÍTULO 4 VARIABILIDADE EM GENES DE RECEPTORES DE NUTRIENTES E DO GOSTO EM SUÍNOS</b> .....	77
<b>Tabela 1</b> Raças, códigos e distribuição dos 79 suínos analisados. ....	86
<b>Tabela 2</b> Informações gerais sobre genes de receptores do gosto em suínos: cromossomo, posição inicial e final, SNPs totais e distribuição por região genômica, consequências funcionais e percentuais de dados perdidos (%DP). As análises foram realizadas em 77 suínos (os dois javalis de Sumatra não foram incluídos).....	89
<b>Tabela 3</b> Diversidade nucleotídica por região genômica ( $\pi$ ), sinônima ( $\pi_s$ ) e não-sinônima ( $\pi_a$ ) e índice de fixação ( $F_{ST}$ ) para cada gene dos receptores do gosto e um pseudogene analisados na população suína total excluindo os dois javalis de Sumatra ( $n=77$ ).....	90
<b>Tabela 4</b> Diversidade nucleotídica na região total ( $\pi_t$ ) e na região do gene ( $\pi_g$ ) por população e por grupo de genes. ....	91
<b>Tabela 5</b> SNPs não-sinônimos (nsSNPs) que foram preditos como deletérios (SIFT score $\leq 0.05$ ) para funcionalidade da proteína por SIFT.....	95
<b>ANEXO C: CAPÍTULO 4</b> .....	121
<b>Tabela C16</b> Números de acesso do GenBank das sequências incluídas nas análises. ....	121
<b>Tabela C17</b> Diversidade nucleotídica total ( $\pi_t$ ) e região do gene ( $\pi_g$ ) por população. ....	122
<b>Tabela C18</b> Distribuição de mutações não-sinônimas por gene, posição no cromossomo, mudança de alelo, identificação de SNP do dbSNPs, SIFT e frequência alélica por população. Escore SIFT $\leq 0.05$ é considerado como deletério na função da proteína e valores $> 0.05$ são considerados tolerados sem nenhum dano à função da proteína. ....	123
<b>ANEXO D: CAPÍTULO 4</b> .....	129
Arquivo no formato GFF3 que foi utilizado na anotação das variantes.....	129

## ÍNDICE DE FIGURAS

<b>CAPÍTULO 2 FILOGEOGRAFIA DO SUÍNO MONTEIRO NO PANTANAL DE NHECOLÂNDIA, MATO GROSSO DO SUL, BRASIL</b> .....	26
<b>Figura 1</b> Distribuição das populações amostradas de suínos monteiros do Pantanal-MS.....	31
<b>Figura 2</b> Árvore filogenética construída pelo método Neighbor-Joining baseando-se na distância D de Nei (1972). Grupos: I (fisicamente próximos e pouco diferenciadas); II (fisicamente distantes e mais diferenciadas) e III (geograficamente distantes e pouco diferenciadas).....	40
<b>Figura 3</b> Teste de Mantel entre distâncias genéticas de Nei e geográficas (A) e entre $F_{ST}$ e distâncias geográficas (B).....	40
<b>Figura 4</b> Proporção de cada população designada a cada cluster para K2 e K4. As quatro cores são referentes às populações definidas ( <i>clusters</i> ) com os resultados do programa STRUCTURE (Pritchard et al., 2000) e visualizadas com o programa DISTRUCT (Rosemberg, 2004): verde=cluster1; vermelho=cluster 2; amarelo=cluster 3 e azul=cluster 4.....	42
<b>ANEXO A: CAPÍTULO 2</b> .....	105
<b>Figura A1</b> Teste de parentesco individual (relatedness) para as onze populações de ..... Monteiros deste estudo baseado em dados de microssatélites. ....	112
<b>Figura A2</b> Estatística Delta K baseada no método de clusterização de Evanno et al. (2005).....	113
<b>Figura A3</b> Relação genética e geográfica entre os 22 haplótipos retratada pela árvore construída pelo método Neighbour-Joining. ....	113
<b>CAPÍTULO 3 VARIAÇÃO GENÉTICA DO GENOMA MITOCONDRIAL EM SUÍNOS E JAVALIS</b> .....	53
<b>Figura 1</b> Árvore consenso obtida pela análise Bayesiana a partir de 108 sequências de mtDNA completo. As cores rosa e cinza indicam os suínos de origem asiática e europeia, respectivamente. Dentro do clado europeu, os ramos de cor azul claro indicam amostras europeias, os ramos de cor verde as amostras americanas. No clado asiático os suínos americanos com haplótipos asiáticos estão destacados pela cor azul e a cor vermelha destaca os javalis do norte da China.....	64
<b>Figura 2</b> Diversidade nucleotídica por população (A) para todos os genes codificadores do mtDNA por genes e em todas as cinco populações suína (B). ....	67
<b>Figura 3</b> Diversidade nucleotídica sinônima e não-sinônima (A e D), divergência (B e E) e razões (C e F) para cada população suína e cada gene do mtDNA. ....	68
<b>ANEXO B: CAPÍTULO 3</b> .....	114
<b>Figura B4</b> Árvore consenso Bayesiana de 341 sequências da região controle do mtDNA. As cores dos ramos indicam designações diferentes dos cladros (preto são europeus e vermelho e azul são ASD e ASWB, respectivamente). Cinza claro representa as sequências usadas para o estudo de mtDNA completo. ....	118
<b>Figura B5</b> Relação entre 62 haplótipos pelo método de network Median-Joining usando 575 pb da região controle do mtDNA. Os diferentes círculos coloridos correspondem às populações e os tamanhos proporcionais às frequências das amostras.. ....	119

**Figura B6** Número de sítios sinônimos e não-sinônimos (a) ocorrendo em cada um dos 13 genes codificadores de proteína. Distribuição das mutações não-sinônimas de cada gene em tolerado e deletério que foram determinado por escores SIFT (b). ..... 120

#### **CAPÍTULO 4 VARIABILIDADE EM GENES DE RECEPTORES DE NUTRIENTES E DO GOSTO EM SUÍNOS** .....77

**Figure 1** Análises de Componentes Principais (PCA) e estrutura. PCA foi realizada com todos os SNPs dos 19 genes, excluindo o pseudogene (A) e para diferentes subconjuntos: SNPs totais (B) e não-sinônimos para o sabor amargo (C). Análise de estrutura com todos os SNPs (D) e SNPs não-sinônimos para o gosto amargo (E). Cada indivíduo está representado por uma linha vertical com as proporções de atribuição a cada cluster mostradas no topo, e as diferentes cores são referentes aos diferentes grupos: internacional (INT) = azul; asiáticos (ASD e ASWB) = verde; e europeus (IB e EUWB) = vermelho. .... 94

#### **ANEXO C: CAPÍTULO 4** ..... 121

**Figura C7** Resultados da análise de STRUCTURE para SNPs de genes de receptores do gosto amargo: o melhor K de 1 a 15 pela estatística Delta K para o conjunto de SNPs (A) e para SNPs não-sinônimos (B)..... 128

# GENÉTICA DA CONSERVAÇÃO DE SUÍNOS LOCALMENTE ADAPTADOS NO BRASIL: USO DE FERRAMENTAS GENÔMICAS E GEOGRÁFICA

Autor: Elizabete Cristina da Silva  
Orientador: Samuel Rezende Paiva, Ph.D.

## RESUMO

A espécie suína é uma das que mais perdeu diversidade genética ao longo dos anos por meio da especialização de seus grupos genéticos. Esta tese teve como objetivo geral analisar as raças suínas localmente adaptadas do Brasil e de outros países da América Latina e Caribe, além de raças domésticas e javalis europeus e asiáticos por meio de ferramentas genômicas e geográficas, como uma maneira de avançar o conhecimento na espécie e poder propor futuros caminhos para sua conservação. Dessa forma, foram realizados três estudos. No primeiro estudo foi realizada uma análise de filogeografia com 181 amostras de suínos Monteiro de 10 pontos geográficos do Pantanal-MS, Brasil, agregando informações moleculares de 19 microssatélites e de um fragmento da região controle do DNA mitocondrial (mtDNA), constatou-se que as populações são geneticamente próximas, independente da distribuição geográfica, como demonstrado pelos valores de  $F_{ST}$  que variaram entre 0,009 a 0,063. O teste de Mantel indicou que existe correlação entre distância geográfica e distância genética variando de 23,09% ( $P=0,06$ ), usando a distância genética  $D$ , a 24,66% quando se utilizou a  $F_{stP}$  ( $P=0,055$ ). Verificou-se que os Monteiros pertencem a um único haplótipo europeu, mas que esses animais não possuem um haplótipo exclusivo por compartilharem seu principal haplótipo com outras populações suínas (comerciais e outras raças locais) que é uma consequência do fluxo gênico entre elas. No segundo estudo foram realizadas análises de variabilidade em 25 raças suínas (incluindo as raças brasileiras Piau, Monteiro e Moura) e 59 javalis, a partir de sequências completas do genoma mitocondrial ( $n=108$ ) e da região controle ( $n=342$ ). Os suínos americanos apresentaram alta frequência de haplótipos europeus (78,6%), com alguns haplótipos asiáticos. A percentagem de introgressão de asiáticos em raças europeias como Pietrain, Large White, Hampshire e Duroc foi de 11,6%. Esta introgressão foi responsável por um aumento considerável na variabilidade, que foi 0,63 e 0,11 com e sem haplótipos asiáticos, respectivamente. A diversidade nucleotídica estimada nos 13 genes

codificadores do mtDNA variou entre as populações e foi mais alta em javalis europeus, seguidos pelos suínos europeus e americanos. A diversidade nucleotídica variou de 0,003 para o *COX1* e *COX3* a 0,008 para o *NADH4L*. O teste McDonald–Kreitman detectou seleção positiva ou negativa em seis genes (*NADH1*, *NADH2*, *COX3*, *NADH3*, *NADH4* e *NADH5*). No terceiro estudo foram realizadas análises de sequências de 21 genes de receptores do gosto e de nutrientes em 79 genomas representando 14 diferentes raças/populações suínas, incluindo as raças brasileiras Piau, Monteiro e Moura. O subconjunto de genes de receptores do gosto amargo (Tas2R) apresentou mais variabilidade que o subconjunto de não-Tas2R nas diferentes raças de suínos baseando-se nos polimorfismos não-sinônimos de nucleotídeo único (nsSNP). Em particular, Tas2R39 mostrou maior número (19) de nsSNPs, enquanto GPR120 (sensor de PUFA) não apresentou nenhum nsSNP.

**PALAVRAS-CHAVE:** filogeografia, sequenciamento de nova geração, receptores de sabor, suínos



## GENETIC OF CONSERVATION OF LOCALLY ADAPTED PIGS IN BRAZIL: USE OF GENOMICS AND GEOGRAPHIC TOOLS

Author: Elizabete Cristina da Silva

Advisor: Samuel Rezende Paiva, Ph.D.

### ABSTRACT

The swine is one that lost more genetic diversity over the years through specialization of their genetic groups. The main goal of this thesis was to analyze locally adapted pigs from Brazil and other countries using genomic and geographical tools. We used 181 samples of Monteiro pigs collected at 10 geographic points in the Pantanal-MS, Brazil, and adding molecular information from 19 microsatellites and from of a fragment of the mitochondrial DNA control region of to perform a phylogeographic study. There is little genetic difference between Monteiro groups independent of geographic distribution, as demonstrated by  $F_{ST}$  values ranging from 0.009 to 0.063. Mantel's test indicated that there was a low correlation with the geographical distance ranging from 23.09% ( $P = 0.06$ ), using the genetic distance  $D$ , and 24.66% using the  $F_{ST}$  ( $P = 0.055$ ). In the analyzes performed with the mtDNA control region, the Monteiro was shown to belong to a single European haplotype, but these animals do not possess a unique haplotype, as they share them main haplotype with other (commercial and other local breeds) pig populations, as a consequence of gene flow between them. In another study, analysis of variability were performed in pigs (25 pig breeds, including the Brazilian Piau breeds, and Moura Monteiro) and 59 wild boars using complete sequences of the mitochondrial genome and its control region. American pigs showed a high frequency of European haplotypes (78.6%), and also presented some Asian haplotypes. The percentage of Asian introgression into European breeds (Pietrain, Large White, Duroc and Hampshire) was 11.6%. This introgression was responsible for a considerable increase in variability, which was 0.63 and 0.11 when we included or not Asian haplotypes, respectively. The estimated nucleotide diversity in 13 genes encoding mtDNA varied among populations; this was higher in European wild boars, followed by European and American pigs. The nucleotide diversity ranged from 0.003 (for COX1 and COX3) to 0.008 (for NADH4L). In addition, ATPase6 and NADH4 genes showed an excess of non-synonymous differences in these populations. The

McDonald-Kreitman test performed by population detected positive or negative selection in six genes (NADH1, NADH2, COX3, NADH3, NADH4 and NADH5) across populations. Finally, sequence analysis of 21 genes for receptors and nutrients in 79 genomes representing 14 different breeds/pig populations, including the Brazilian Piau, Monteiro and Moura breeds were performed. The analyzes revealed that the subset of genes for bitter taste receptors (Tas2R) showed higher variability than the subset of non-Tas2R in different pig breeds based on polymorphisms in non-synonymous single nucleotides (nsSNP). In particular, Tas2R39 showed greater number (19) of non-synonymous SNPs while GPR120 sensor (PUFA) showed no nsSNP. The results showed that differences in taste receptor genes, between Asian and European pig breeds, only a small part of the variability is not consistent with previous findings involving data from complete genome.

**KEY-WORDS:** phylogeography, next generation sequencing, pigs

## CAPÍTULO 1

### 1. INTRODUÇÃO

A suinocultura durante muitos anos tem sido destaque no setor agropecuário mundial. A carne suína é a mais consumida no mundo, no ano de 2013 foram consumidos 107,5 milhões de toneladas. A China segue liderando no ranking de produção mundial com 53,8 milhões de toneladas, seguida pela União Européia com 22,45 milhões de toneladas, Estados Unidos com 10,51 milhões de toneladas e em quarto lugar o Brasil com uma produção de carne de 3,3 milhões de toneladas que representa menos de 4% da produção total (USDA, 2013). A espécie suína é uma das espécies de animais domésticos que mais se destaca nos estudos genômicos, principalmente, na busca de loci ou QTLs (*Quantitative Trait Loci*) que exercem influência sobre as características de interesse econômico (Pires et al., 2006; Pinto et al., 2010).

A espécie suína também tem adquirido importância nos estudos de conservação de recursos genéticos devido à perda de material genético que vem ocorrendo durante décadas (Mariane et al., 2003). Isto pode ser notado pela falta de informações a respeito das raças localmente adaptadas, em vários países do mundo, principalmente na América Latina e Caribe. A escassez de informações pode estar conectada a falta de interesse pelas raças locais de cada país, uma vez que a maioria dos rebanhos suínos está composta por raças de alto potencial genético para produção de carne, como Landrace, Large White, Duroc, Pietrain e Hampshire (FAO, 2007).

As raças suínas mais utilizadas na suinocultura intensiva são Duroc, Landrace, Large White, Pietrain, Hampshire, além de uma grande variedade de linhagens modernas de suínos que foram desenvolvidas para atender a demanda atual de animais com maior potencial genético para deposição de proteína na carcaça e fêmeas hiperprolíficas (Irgang et al., 1992; Fávero & Figueiredo, 2009). Contudo, ainda existem as raças localmente adaptadas ou crioulas que representam uma pequena parcela do material genético dentro da suinocultura.

Estas raças ou grupos genéticos têm sido encontrados nas propriedades de base familiar e se tornaram um meio de vida para o homem do campo (Castro et al. 2002; Silva et al., 2011).

As raças suínas localmente adaptadas no Brasil exerciam um papel importante na suinocultura local até a década de 50. Depois deste período, com o surgimento das indústrias de óleos vegetais e mudanças nos hábitos alimentares dos consumidores, esses animais que apresentam características para alta deposição de gordura foram aos poucos substituídos por raças importadas (Vianna, 1986). A grande disponibilidade de técnicas moleculares proporcionou a realização de vários estudos em raças ameaçadas de extinção utilizando as informações dos marcadores baseados em polimorfismos do DNA (Maudet et al., 2002; Tadano et al., 2007). Os microssatélites e mtDNA são os marcadores mais utilizados para estudar a diversidade e estrutura genética e a origem de raças suínas brasileiras (Sollero et al., 2009; Souza et al., 2009; Cavalcante Neto, 2010, Silva et al., 2011) e também de outros países (Ramirez et al. 2009; Luetkemeier et al. 2010; Yu et al. 2013).

Os avanços no campo da genética da conservação têm permitido a realização de estudos que englobam a área de filogeografia que geram informações sobre a distribuição da variação genética numa escala espacial de várias espécies de animais e plantas utilizando marcadores moleculares combinados com dados demográficos e geográficos (Foll & Gaggiotti, 2006; Taberlet & Bouvet, 1994). A filogeografia contribui para a compreensão dos processos microevolucionários, como fluxo de genes, seleção natural, a existência de isolamento por distância e gargalo de garrafa (*bottleneck*), que podem conduzir a perda de variabilidade genética ou mesmo a extinção de uma espécie e/ou população (Mantel, 1967; Manel et al., 2003).

Atualmente são esperados maiores avanços nos estudos de genética da conservação e no melhoramento genético dos animais domésticos, uma vez que as novas tecnologias de sequenciamento de alto desempenho estão cada vez mais acessíveis e várias espécies de animais domésticos têm o genoma sequenciado (International Chicken Genome Sequencing Consortium, 2004; Bovine Genome Sequencing and Analysis Consortium et al., 2009), incluindo a espécie suína (Archibald et al., 2010). Desde que foi iniciado o sequenciamento do genoma da espécie suína, em 2005, houve um aumento de informações disponíveis na literatura sobre genes que codificam proteínas (Bosse et al., 2014), *QTLs* relacionados com características de importância econômica para suinocultura industrial (Soma et al., 2011; Becker et al., 2013; Hu et al., 2013), o desenvolvimento e utilização de chips contendo um grande número de SNPs (Ramos et al., 2009; Ramayo-Caldas et al., 2010) e

estudos de associação global do genoma ou Genome- Wide Association-GWAS (Nonneman et al., 2011; Do et al., 2013; Manunza et al., 2014).

### **1.1 Problemática e relevância**

Os suínos localmente adaptados no Brasil apresentam um patrimônio genético com uma diversidade muitas vezes superior às raças comerciais (Sollero et al., 2009; Silva et al., 2011). Animais adaptados às condições tropicais são mais resistentes a várias doenças e parasitas do que raças de origem européia (Oliveira et al. 2009, Abelli et al. 2010). Os suínos localmente adaptados no Brasil são conhecidos pela rusticidade, por terem uma conformação corporal para deposição de gordura e por serem menos exigentes em manejo se comparados às raças suínas melhoradas (Ferreira & Lima, 2001; Mariante et al., 2003). De acordo com Mariante et al.(2003), os suínos locais também são capazes de sobreviver em diferentes ecossistemas das regiões brasileiras mesmo em condições desfavoráveis de criação.

Em geral, como acontece na maioria dos países, as raças suínas locais são encontradas apenas em pequenos criatórios que estão distribuídos nas diferentes regiões do País (Castro, 2002; Silva Filha et al., 2005; Silva et al., 2011). Estudos realizados com essas raças focaram na caracterização fenotípica (Castro, 2002; Silva Filha, 2005 e McManus et al., 2010), caracterização genética baseada em marcadores moleculares (Sollero et al., 2009; Souza et al., 2009; Cavalcante Neto, 2010; Silva et al., 2011) e prospecção de QTLs (Pires et al., 2006; Pinto et al., 2010). Contudo, não existem ou são escassas informações de como ocorre a distribuição da variabilidade genética nas regiões em que as raças suínas locais foram adaptadas.

Em suínos, como também em muitas espécies, diversas análises de mtDNA estão sendo realizadas, principalmente, com base na região não codificadora (região controle ou D-loop) ou citocromo b (Souza et al., 2009; Luetkemeier et al., 2010; Wang et al., 2010; Yu et al., 2013) e poucos estudos analisaram o genoma mtDNA completo (Yang et al., 2003; Wu et al., 2007; Fernandez et al., 2008). Estudos de filogeografia também foram realizados com várias espécies de plantas e animais domésticos (Foll & Gaggiotti, 2006; Taberlet & Bouvet, 1994). Um estudo desta natureza esclareceu que os suínos tiveram múltiplos centros de domesticação em toda a Eurásia e que os javalis originaram os porcos domésticos europeus (Larson et al., 2005). García et al. (2011), numa análise filogeográfica com mtDNA,

identificaram elevados níveis de diferenciação e baixa taxa de migração entre populações de javalis colombianos de diferentes localidades, sugerindo que cada população do estudo fosse considerada uma única unidade de manejo genético.

O surgimento das novas tecnologias de sequenciamento permitiu grandes avanços nos estudos genômicos de diversas espécies de animais domésticos de interesse econômico (International Chicken Genome Sequencing Consortium, 2004; Bovine Genome Sequencing and Analysis Consortium et al., 2009), incluindo a suína que teve o genoma completo de um animal da raça Duroc sequenciado pelo Consórcio Internacional de Sequenciamento do Genoma Suíno (SGSC) (Archibald, Bolund et al., 2010).

Na versão atual do ensembl v75 existem 25.322 genes codificadores anotados no *S. scrofa assembly* 10.2 ([http://www.ensembl.org/Sus\\_scrofa/Info/](http://www.ensembl.org/Sus_scrofa/Info/)). No entanto, ainda são excassas as informações sobre todos os genes de receptores de nutrientes e do gosto presentes na cavidade oral de suínos que apesar de serem bem documentados em outras espécies de vertebrados (Conte et al., 2002; Conte et al., 2003; Li & Zhang, 2014) foram pouco estudados na espécie suína (Groenen et al. 2012). Os genes de receptores de sabor apresentam uma grande importância para a suinocultura, uma vez que estão envolvidos na escolha do tipo de alimentos pelos animais, além disso, desempenham um importante papel na detecção de toxinas que podem comprometer a saúde e até por em risco a vida dos animais (Wong et al., 1996; Dong et al., 2009; Li & Zhang (2014).

## 1. 2 OBJETIVOS

### 1.2.1 Objetivo Geral:

Diante do exposto, o presente trabalho visou, por meio de ferramentas genômicas e geográficas, realizar um estudo que engloba as áreas, da filogeografia e as tecnologias de sequenciamento de nova geração (NGS) para discernir o padrão de variabilidade nucleotídica do genoma mitocondrial e de genes nucleares em raças suínas localmente adaptadas no Brasil, crioulas dos países Argentina, Cuba, Peru, Colombia, Guatemala, México, Estados Unidos, e outras raças domésticas e javalis europeus e asiáticos.

### 1.2.2 Objetivos Específicos:

**Capítulo 1** - Caracterizar e comparar os padrões de distribuição geográfica da variabilidade genética em populações de suínos Monteiros no Pantanal da Nhecolândia, Mato Grosso do Sul, Brasil, utilizando polimorfismos de microssatélites e da região controle do mtDNA.

**Capítulo 2** - Estudar a variabilidade e identificar assinaturas de seleção no genoma mitocondrial de suínos domésticos e javalis.

**Capítulo 3** - Quantificar e comparar a variabilidade de genes de receptores de nutrientes e do gosto em diferentes raças suínas domésticas e javalis de vários países.

## 2. REVISÃO DE LITERATURA

### 2.1 Importância das raças suínas localmente adaptadas

Embora com a origem incerta, é sabido que os grupos genéticos de suínos localmente adaptados no Brasil, como Piau, Canastra, Canastrão, Nilo, Caruncho, Pirapetinga, Moura e Monteiro foram formados a partir das raças ibéricas. Estudos realizados por Grossi et al. (2006) e Souza et al. (2009) confirmaram tal informação por meio de marcadores no mtDNA. As informações sobre as principais características morfológicas dessas raças podem ser consultadas nas obras dos autores Athanasof (1947), Cavalcanti (1985), Vianna (1985) e McManus et al. (2010).

Quanto à situação das raças suínas localmente adaptadas no Brasil, das treze raças e/ou ecótipos identificados apenas dois foram registradas na Associação Brasileira de Criadores de Suínos (ABCS): a raça Piau e a raça Moura. Além dessas duas raças, o porco Monteiro desempenha um papel importante na suinocultura local do Pantanal brasileiro. Essas três raças foram escolhidas para esse estudo e a seguir será dada uma breve descrição de suas características e situação no País.

A raça Piau foi a primeira raça nacional registrada no Pig Book brasileiro, em 1989, com 1250 suínos registrados entre 1989 e 1995. A principal qualidade do Piau é a rusticidade e os animais podem ser identificados pela pelagem ovejuna (branca/creme com manchas pretas difusas), orelha do tipo asiática a ibérica e perfil fronto-nasal retilíneo a subconcavilíneo. A média de leitões nascidos por leitegada é de 8,06 e ao desmame é de 7,41 leitões, e o peso médio da leitegada aos 21 dias é de 33,43 kg. Essa raça é considerada a raça brasileira mais importante (Cavalcanti, 1985; Vianna, 1985). Segundo Cavalcanti (1985), a raça Piau foi a que sofreu maior processo de seleção e melhoramento genético por serem animais hiperprolíficos e apresentarem características desejáveis para produção de carne. Essa



raça foi submetida à seleção por muitos anos e alguns exemplares podem ser encontrados nos estados São Paulo, Minas Gerais, Paraná, Rio Grande do Sul e Bahia, com maior concentração na bacia do rio Paranaíba, (Vianna, 1985; Mariante et al., 2003).

O suíno Monteiro encontrado no Pantanal é uma raça asselvajada e se originou de suínos trazidos pelos colonizadores para a região de Corumbá na segunda metade do século XVIII. Esse animal apresenta características muito semelhantes as do javali, apresentando conformação do corpo e cabeça em formas de cunha, coloração preto acinzentado ou marrom, orelhas pequenas e eretas, membros fortes e ágeis (Herrera et al., 1996). No Pantanal, o porco Monteiro é muito visado para caça, constituindo uma fonte protéica para as famílias que habitam a região. Cruzamentos de Monteiro com javali já foram realizados objetivando a obtenção de carne mais saudável e saborosa (Rosa et al., 2000).

A raça Moura foi originada do cruzamento entre as raças localmente adaptadas no Brasil, como Canastrão e Canastra e a raça exótica Duroc (Egito et al., 2002). Essa raça é considerada uma das mais antigas (Vianna, 1985) e preservadas até hoje em pequenos plantéis da Universidade Federal do Paraná (UFPR), Embrapa Suínos e Aves e Universidade Estadual de São Paulo, além de algumas fazendas em Santa Catarina e Paraná (Silva 2010).

As características marcantes para criação da raça Moura são a sua rusticidade, prolificidade, o comprimento de carcaça e o marmoreio na carne. Os animais possuem pelagem preta entremeada de pelos brancos, orelhas do tipo ibéricas grandes e firmes e perfil fronto-nasal retilíneo a subcôncavilíneo. A média de leitões nascidos por leitegada é de 8,32 e de leitões desmamados 7,57, o peso médio da leitegada aos 21 dias é de 33,60 kg. Essas três raças e os demais ecótipos podem ser encontrados em pequenos rebanhos distribuídos pelo Brasil, sendo utilizados em cruzamentos com raças exóticas, por exemplo, excelentes matrizes da linha fêmea MO25C (resultado do cruzamento de Moura com Landrace e Large White) foram lançadas esse ano pela Embrapa Suínos e aves (<https://www.embrapa.br/suinos-e-aves/busca-de-produtos-processos-e-servicos/-/produto-servico/1558/femea-embrapa-mo25c>, 2014; Castro et al., 2002; Silva Filha et al., 2005; Cavalcante Neto et al., 2007).

As raças suínas localmente adaptadas no Brasil são conhecidas por possuírem características importantes, como resistência às doenças e ectoparasitas, boas qualidades maternas e adaptadas às condições de manejos precários (Cavalcanti, 1985; Mariante & Cavalcante, 2006). No entanto, raros estudos exploraram suas características peculiares. Apenas as raças Piau, Moura e Monteiro são até hoje inseridas em trabalhos de pesquisas em instituições públicas brasileiras, como Universidade Federal de Viçosa (UFV, Minas Gerais) e Embrapa Suínos e Aves e Embrapa Pantanal, respectivamente. A raça Piau, por exemplo, foi

muito utilizada nos cruzamentos divergentes para prospecção de *QTLs* (*Quantitative Trait Loci*) para características de desempenho e qualidade de carne (Pires et al., 2006; Pinto et al., 2010). A raça Moura apresenta índices produtivos inferiores às raças exóticas, porém pode ser utilizada para produção de presunto curado (Fávero et al., 2007) e para melhorar qualidade de carne (marmoreio) quanto usadas nos cruzamentos com raças comerciais (linha MO25C lançada pela Embrapa Suínos e Aves em 2014) e a raça Monteiro demonstrou reagir bem a *Leptospiras* patogênicas (Girio et al., 2004). Para estas e as demais raças suínas brasileiras, estudos de levantamento populacional e de caracterização morfológica e genética visando à conservação podem ser encontrados nas publicações de Silva et al. (2002); Gonela (2003); Silva Filha et al. (2006); Sollero et al. (2008); Cavalcante Neto (2010); McManus et al. (2010) e Silva et al. (2011).

A perda de diversidade genética tem sido observada nas populações suínas brasileiras segundo reportado por Sollero et al. (2009) e Silva et al. (2011). Os autores identificaram elevado nível de consanguinidade intra-populacional causado pelos acasalamentos entre animais aparentados, que é uma prática comum nos pequenos criatórios de suínos do país.

## **2.2 Diversidade genética e uso de ferramentas moleculares**

A diversidade genética de uma espécie compreende a variação de genes no indivíduo dentro de uma população ou variação entre as populações geográficas. O nível de diversidade genética tende a ser diferente de um indivíduo para outro dentro de uma população e, portanto, é responsável pelas diferenças entre populações de uma mesma espécie (Halliburton, 2004) que são resultantes do processo evolutivo e permitem a adaptação às diferentes condições de vida (Ayala, 1982).

A diversidade genética pode aparecer espacialmente estruturada em diferentes escalas, como população, subpopulação ou em um grupo de indivíduos próximos, e sua distribuição pode ser influenciada pelas forças evolutivas, como seleção natural, migração e deriva, e pelos processos demográficos e históricos. Todos esses fatores afetam o grau de expansão, contração e fragmentação das populações e podem causar o isolamento das espécies (Beaumont, 1999).

A variação intrapopulacional pode ser facilmente analisada utilizando parâmetros, como número de alelos, heterozigosidade média e proporção de locos polimórficos por indivíduo (Hedrick, 1999). Essa variação é a condição necessária para a seleção natural atuar nas diferenças que ocorrem dentro das populações (Toggler et al., 1995) e quanto mais variabilidade genética maior a chance de sobrevivência diante de mudanças ambientais (Flashman et al., 2001).

Um grande passo nos estudos de caracterização dos recursos genéticos tem sido dado nas últimas décadas. Os avanços na área da genética contribuíram para melhoria nos estudos que envolvem o campo da genética da conservação. Há muitos marcadores em DNA (RFLP, AFLP, RAPD, microssatélites, mtDNA, Y, SNPs) que foram aplicados para estudar a diversidade genética de animais e plantas (Spritze et al., 2003; Albuquerque et al., 2006). Neste trabalho serão dadas descrições dos marcadores escolhidos para o presente estudo, como os microssatélites, o mtDNA e os SNPs.

### 2.2.1 Microssatélites

Os Microssatélites ou SSR (*simple sequence repeats* ou *STR*) são sequências curtas que consistem em elementos repetidos *in tandem* cujo comprimento pode variar de 1 a 6 pb e apresentam características desejáveis para serem utilizados em estudos de genética de populações e identificação individual (Powel et al., 1996; Werner et al., 2004). Esses marcadores são altamente polimórficos podendo contribuir para detectar com maior precisão e eficiência as diferenças genéticas entre indivíduos. Tratando-se da neutralidade seletiva, os microssatélites podem estar relacionados com a codificação de elementos e regulação da expressão do gene (Kashi & Soller, 1999).

Os microssatélites podem ser classificados de acordo com o comprimento da unidade de base repetida (mono-, di-, tri-, tetra-, penta-e hexa-nucleotídeos) e estrutura ("perfeito" e "imperfeito") e representam as regiões instáveis do genoma devido às alterações mutacionais causadas pela perda ou adição de unidades repetitivas. Essas mutações são causadas por mecanismo intramolecular denominado de *slippage* ("escorregão") da DNA polimerase ou *crossing over* desigual (Schlotterer & Tautz, 1992).

O número de repetições dos microssatélites, a sequência da unidade de repetição, o comprimento da unidade de repetição, as sequências que flanqueiam, a presença

de imperfeições, a taxa de recombinação e de transcrição, e a eficiência do sistema reparador são fatores que influenciam na dinâmica evolutiva desses marcadores (Schlotterer & Tautz, 1992).

Os microssatélites espécie-específicas estão descritos na literatura para várias espécies de animais e plantas (Tautz, 1989), e a utilização dos mesmos está muito difundida nos estudos de diversidade genética de animais ameaçados de extinção, pois além de serem facilmente amplificados por reação em cadeia da polimerase-PCR, várias literaturas disponibilizam informações úteis que têm contribuído para as diversas pesquisas empregando esses marcadores (Ferreira et al., 2014; Silva et al., 2014).

Um entrave para o uso dos marcadores microssatélites é a dificuldade de comparar os resultados obtidos por diferentes grupos. Diante disso, a FAO e a *International Society of Animal Genetics* (ISAG) recomendam desde 2007 painéis de marcadores microssatélites para estudar a diversidade genética de várias espécies de animais domésticos. Para a espécie suína existe um painel composto por 27 locos de microssatélites que permitem quantificar a variabilidade genética dessa espécie (FAO, 2007). O objetivo é que nas pesquisas sejam utilizados o mesmo painel para que os resultados possam ser comparados globalmente. No entanto, é mais comum a utilização de marcadores recomendados pela FAO combinados com outros marcadores descritos na literatura (Sollero et al., 2008; Silva et al., 2011).

Diversos estudos sobre a diversidade e estrutura genética de populações suínas foram realizados utilizando marcadores microssatélites tanto no Brasil (Gonela, 2003; Sollero et al., 2009; Silva et al., 2011) como em outros países (Calvo et al., 2000; Lemus-Flores et al., 2001, Chaiwatanasin et al., 2002; Canul et al., 2005; Vicente et al., 2008).

Gonela (2003) utilizou seis locos de microssatélites na caracterização genética de uma pequena população de Porco Monteiro (N=99) e *Tayassu pecari* (N=15) e verificou alta variabilidade (4 a 15 alelos, heterozigosidade de 0,13 a 1,00 e conteúdo de informação polimórfica variando de 0,18 a 0,84).

Sollero et al. (2008) avaliaram, com 28 microssatélites, a diversidade e estrutura genética de cinco populações suínas, sendo três localmente adaptadas no Brasil (Piau, Moura e Monteiro) e duas comerciais (Landrace e MS60). Os resultados desse estudo mostraram que 14% da variação genética total foi devido às diferenças entre raças. Na análise de variabilidade intrarracial, os autores verificaram que a raça Piau apresentou maior valor de heterozigosidade esperada.

A diversidade e estrutura genética de nove grupos genéticos de suínos localmente adaptados no estado de Pernambuco, Brasil (Baé, Caruncho, Canastra, Canastrão, Mamelado, Moura, Nilo, Piau e Sem Raça Definida-SRD) e três raças exóticas (Duroc, Landrace e Large White) foram analisadas com 22 marcadores microssatélites. Os resultados indicaram que 3,6% da variação genética total foram das diferenças entre suínos locais e comerciais. Os suínos locais apresentaram maiores valores para os parâmetros de diversidade genética intrapopulacional (Silva et al. 2011).

### 2.1.2 DNA mitocondrial (mtDNA)

O mtDNA é uma molécula circular de fita dupla com aproximadamente 16 kb que contém 13 genes codificadores de proteínas, 22 genes de RNAs de transferência e 2 genes de RNA ribossômicos (Gray, 1989) e apresenta uma taxa média de substituições sinônimas aproximadamente 20 vezes maior do que no DNA nuclear (Pesole et al. 1999).

A sua aplicação como marcador molecular para estudar a evolução molecular, classificação biológica e estrutura genética é devido às características, como peso molecular pequeno, taxa de evolução rápida e quase exclusivamente de herança materna (haplóide), e ausência de recombinação genética. As duas regiões do mtDNA mais recomendadas nos estudos de estrutura populacional e interpretação filogenética das espécies são a região controle (D-loop) e o gene que codifica para a citocromo b (Awise, 1994).

A região controle desempenha um papel de regular a replicação e a transcrição do mtDNA e compõe cerca de 6% do genoma mitocondrial. Esta região está localizada entre tRNA para prolina e tRNA fenilalanina e contém de 5 a 29 sequências repetidas *in tandem* (SRTs) (Clayton, 1991) com uma taxa de substituição de base de cinco a dez vezes maior que a de outras regiões do mtDNA (Mackay et al. 1986).

Diversos estudos filogenéticos com populações suínas usaram as sequências da região controle (D-loop) e do gene citocromo b. Randi et al. (1996) usaram os polimorfismos do citocromo b nos estudos evolutivos de *Suiformes*. Alves et al. (2003) empregaram citocromo b (1140 pb) e a região controle (707pb) para determinar as relações filogenéticas entre suínos ibéricos antigos e atuais. Alex et al. (2004) encontraram uma mistura de haplótipos asiáticos e europeus no porco preto de Canárias por meio de polimorfismos identificados no gene citocromo b. No Brasil, Souza et al. (2009) utilizaram o citocromo b

para investigar a origem das raças suínas brasileiras e verificaram que as raças suínas brasileiras são descendentes de duas linhagens maternas (europeia e asiática), as raças Nilo, Monteiro, Piau e Tatu compartilharam haplótipos de uma linhagem Ibérica, enquanto, a raça Moura apresentou haplótipos de diferentes linhagens de origem europeia e asiática e, por isso, essa raça apresenta maior divergência em relação às demais raças suínas brasileiras.

### **2.1.3 Polimorfismo de nucleotídeo único (SNP)**

Os SNPs são as variações (A, C, T e G) mais comuns de DNA dentro do genoma (Sachidanandan et al., 2001). Os SNPs apresentam vantagens em relação a outros marcadores genéticos pela possibilidade de serem identificados por meio de técnicas de alto desempenho e a um custo, relativamente, baixo (Wasson et al, 2002; Lindroos et al., 2003; Wiedmann et al., 2008; Shen et al., 2009). A tendência é que esses marcadores venham a substituir os microssatélites nos estudos de identificação genética e testes de paternidade (Fries et al., 2001; Werner et al., 2004), pois os SNPs têm garantido bons resultados nos testes de paternidade (Heaton et al., 2002), de identificação individual (Allen et al., 2010) e nos estudos de marcas de seleção (Nielsen, 2005) e de processos evolutivos (Bouck & Vision, 2007). Os SNPs no cromossomo X de suínos de diversos países, incluindo raças brasileiras, permitiram a separação de dois grupos distintos asiáticos e europeus (Burgos-Paz et al., 2012).

Os SNPs são um dos temas centrais em genômica e, com o surgimento das plataformas de sequenciamento de nova geração, milhões de SNPs estão sendo descobertos em várias espécies de interesse agropecuário (Eck et al., 2009; Wiedmann et al., 2008; Liu, 2009; Allen et al., 2010), pois por serem variações facilmente detectáveis estão sendo preferidas nos estudos que envolvem associação entre genótipo e fenótipo (Hayes & Goddard, 2010; Barris et al., 2012).

## 2. 2 Sequenciamento de nova geração (NGS)

Durante muitos anos o método de sequenciamento de Sanger (1987) foi fundamental na era genômica. Porém, os novos avanços biotecnológicos na área da genética exigem métodos de sequenciamento rápidos e de baixo custo. Na última década, mais precisamente a partir de 2005, a forma de aquisição de dados genômicos tem mudado completamente (Mardis, 2008), graças às tecnologias de sequenciamento de nova geração (SNG) que são capazes de sequenciar milhões de pares de bases em uma única corrida (Dalca & Brudno, 2010).

Esses novos métodos permitem um sequenciamento mais barato e eficiente e com eles é possível sequenciar muitos genomas simultaneamente e em um período de tempo bastante reduzido. Em relação ao custo e tempo de análise, com essas novas tecnologias pode-se sequenciar, por exemplo, um genoma humano completo em uma semana a um custo de 200 vezes menor do que os métodos de sequenciamento anteriores (Dalca & Brudno, 2010).

No mercado estão disponíveis as plataformas desenvolvidas por diversas companhias: 454 (Roche), HiSeq2000, HiSeq2500 e MiSeq (Illumina), SoLiD (Life Technologies), HeliScope (Helicos BioSciences), Single Molecule Real Time-SMRT (Pacific Bioscience), Ion Torrent-PGM e Proton (Life Technologies). Estas plataformas diferem no método de análise, tempo de execução, tamanho dos fragmentos analisados (*reads*) e custos (Dalca & Brudno, 2010). Estas plataformas estão sendo empregadas para sequenciamento de novo (Li et al., 2010), re-sequenciamento para descobrir variantes genômicas (Ramayo-Caldas et al., 2010; Liu, 2009; Rubin et al., 2012), análises de transcriptoma (Nagalakshmi et al., 2008) e estudos de metilação do genoma (Zhang & Jeltsch, 2010).

Diante do exposto, o sequenciamento de genomas completos está mais acessível, no entanto, para análise das sequências é exigido o emprego de potentes ferramentas de bioinformática (Li et al., 2008, Shen et al., 2010), sendo este um ponto crítico a ser discutido pela necessidade de recursos humanos qualificados para manipulação dos dados de sequenciamento de nova geração (Sboner et al., 2011).

As novas tecnologias de sequenciamento também contribuíram para diminuir os custos de genotipagem de SNP e isso é de extrema importância para o descobrimento de polimorfismos associados às características de interesse econômico (Sborner et al., 2011). Além disso, os polimorfismos acessados podem ser incorporados aos métodos de Seleção

Genômica (Meuwissen et al. 2001) para obter maior progresso genético dos animais nos programas de avaliação e melhoramento animal (Caetano, 2009).

### **2.3 Filogeografia**

A filogeografia contribuiu para o entendimento da distribuição da diversidade genética de muitas espécies de animais e plantas (Eizirik et al., 2001; Miller et al., 2011). Avise (2000) define filogeografia como um campo de estudo preocupado com os princípios e processos que governam as distribuições geográficas de linhagens genealógicas, especialmente aquelas dentro e entre espécies proximamente relacionadas.

A filogeografia surgiu como uma disciplina e seu foco principal é estudar a variação genética em populações de uma espécie ou entre espécies próximas e podem ser realizados com enzimas, RAPDs, RFLPs, cromossomo Y, DNA mitocondrial (mtDNA), SNPs e os microssatélites (Avise et al., 1987; Avise, 1998). O recomendado é a utilização de pelo menos dois tipos de marcadores para obtenção de resultados satisfatórios (Scandura et al., 2011).

Um dos principais pontos discutidos nos trabalhos de filogeografia é a estrutura geográfica e genética populacional das espécies que é definida como o resultado dos processos microevolutivos e demográficos que atuam entre e dentro das populações, assim como da biogeografia histórica das linhagens genéticas (Wenink et al., 1996). As variações geográficas também podem revelar quais os mecanismos históricos, ecológicos, os padrões fenotípicos, assim como a distribuição de muitas espécies de animais domésticos. Entender estes padrões de variação geográfica é o primeiro passo para inferir os processos evolutivos de suas populações naturais (Zink & Remsen, 1986).

Para alcançar as metas propostas nos estudos filogeográficos é necessário estabelecer as relações filogenéticas de um grupo de indivíduos da mesma espécie que ocorre em uma área geográfica (ex.: no território nacional). Isso pode ser feito mediante o uso de sequências de DNA, avaliação dos agrupamentos formados para verificar se os mesmos pertencem à mesma área geográfica e se os grupos mais relacionados são mais próximos geograficamente e, por último, relacionar os indivíduos a outros organismos para verificar se há ou não congruências na distribuição geográfica (Miyaki, 2009).



## 2.4 Genes de receptores de sabor e de nutrientes

Os receptores de sabor estão localizados na cavidade oral e também ao longo do trato gastrointestinal (Colombo et al., 2012). Estes receptores de sabor contribuem para que os animais e seres humanos possam aceitar ou rejeitar certos tipos de alimentos ou nutrientes, pois através dos receptores gustativos podem ser identificados os alimentos/nutrientes benéficos ou nocivos à saúde (Chandrashekar et al., 2000). Existem cinco tipos de receptores de sabores que são mais conhecidos e estudados até o momento: amargo, doce, salgado, ácido ou azedo e recém-aceito umamami (Kinnamon & Cummings 1992; Bachmanov & Beauchamp, 2007). Além desses, existem receptores de novos sabores que foram descritos mais recentemente, como os sabores gordura e metálico (Chaudhari & Roper, 2010).

Os receptores de sabor doce e umami são particularmente interessantes por permitirem detectar fontes de carboidratos e proteínas, que são de extrema importância para manutenção e sobrevivência dos animais (Feng & Zao, 2013). Os dois sabores são mediados por alguns receptores de gosto do tipo I acoplados às proteínas G conhecidos como Tas1R e T1rs (Adler et al., 2000; Meyerhof et al., 2011). Contudo, os receptores de sabor amargo adquiriram mais importância porque permitem que os organismos identifiquem substâncias tóxicas nos alimentos que podem por em risco a saúde e a vida dos organismos (Wong et al., 1996; Dong et al., 2009; Li & Zhang (2014)). Algumas espécies, simplesmente, por possuírem mais receptores para o sabor amargo apresentam mais sensibilidade para detectá-los, por exemplo, nos humanos foram identificados 25 receptores funcionais de sabor amargo (Conte et al. 2002) e, por isso, toleram menos o sabor amargo que os suínos, que possuem menos receptores de sabor amargo (Groenen et al., 2012). A sensibilidade em detectar o sabor amargo dos alimentos é mediada por dezenas de receptores do gosto do tipo II acoplados às proteínas G conhecidos como T2Rs e Tas2Rs (Adler et al., 2000; Meyerhof et al., 2011).

Embora o repertório de genes de receptores de sabor tenha sido descrito em alguns animais, como aves, rãs, cães, humanos, peixes teleósteos (Conte et al., 2002; Conte et al., 2003; Li & Zhang (2014)), ainda está totalmente descrito em suínos (Groenen et al., 2012). Li & Zhang (2014) estudaram os genes do gosto amargo em pelo menos 54 espécies de vertebrados e verificaram que a diversidade genética, observada nesses genes, pode estar relacionada à adaptação das espécies para detectar a presença de toxinas nas dietas.

### 3. REFERÊNCIAS BIBLIOGRÁFICAS

- ABELLI, A.M.G.; NAKATA, L.C.; ANDRÉO, R. et al. mRNA profile of Nellore calves after primary infection with *Haemonchus placei*. **Vet. Parasitol.** n.176, p.195-200, 2010.
- ADLER, E.; HOON, M.A; MUELLER, K.L. et al. A novel family of mammalian taste receptors. **Cell**, v.100, n 6, p.693-702. 2000.
- ALBUQUERQUE, M.S.M.; EGITO, A.A.; PAIVA, S.R. et al. Conservação e caracterização de búfalos no Brasil: uma Revisão da Literatura. **Embrapa Rec. Gen. e Biotec.**, Brasília, DF, 2006.
- ALEX, F.; MARCEL, C.; JOSE-LUIS, A. et al. Estimating the frequency of Asian cytochrome B haplotypes in standard European and local Spanish pig breeds. **Genet. Sel. Evol.** n. 36, 97–104, 2004.
- ALLEN, A. R.; TAYLOR, M. MCKEOWN, B. Compilation of a panel of informative single nucleotide polymorphisms for bovine identification in the Northern Irish cattle population. **BMC Genetics.** v.11, n.5. 2010.
- ALVES, E., OVILO, C. RODRIGUEZ, M.C. et al. Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and other domestic and wild pig populations. **Anim. Genet.** 34: 319–24. 2003.
- ATHANASSOF, N. **Manual do Criador de Suínos.** Piracicaba: Edições Melhoramentos, 1947.
- ARCHIBALD, A. L.; BOLUND, L. et al. Pig genome sequence analysis and publication strategy. **BMC Genomics** 11: 438, 2010.
- AVISE, J.C. **Phylogeography: the History and Formation of Species.** Harvard University Press, Cambridge, MA. 446pp. 2000.
- \_\_\_\_\_. **Molecular markers, natural history and evolution.** New York: Chapman & Hall, 541 p. 1994.
- AVISE, J. C. ARNOLD, J.; BALL, R. M. et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. **Annual Review of Ecology, Evolution and Systematics**, v. 18, p. 489-522, 1987.

- AVISE, J. C.; BALL, R. M.; ARNOLD, J. Current versus historical population sizes in vertebrate species with high gene flow: A comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. **Mol. Biol. Evol.**, n. 5. pp.331- 344. 1998.
- AYALA, F.J. **Population and Evolutionary Genetics: A Primer**. The Benjamin Cummings Pub. Co. Inc, California, 268p, 1982.
- BACHMANOV, A. A.; BEAUCHAMP, G. K. Taste receptor genes. *Annu Rev Nutr*, v. 27, p 389–414, 2007.
- BEAUMONT, M. A. Detecting population expansion cline using microsatellites. **Genetics**. n. 153, 2013-2029. 1999.
- BECKER, D.; WIMMERS, K.; LUTHER, H. et al. A Genome-Wide Association Study to Detect QTL for Commercially Important Traits in Swiss Large White Boars. **PLoS ONE**, v.8, n.2, p.55951, 2013.
- BOSSE, M.; MEGENS, H.-J.; LAURENT, A.F. et al. Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. **Nature Communications**, v.5, n. 4392, 2014.
- BOUCK, A.; VISION, T. The molecular ecologist's guide to expressed sequence tags. **Mol. Ecol.** v.165, 907–924. 2007.
- BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v.324, n.5926, p.522–528, 2009.
- BURGOS-PAZ, W., SOUZA, C. A.; CASTELLO, A. Worldwide genetic relationships of pigs as inferred from X chromosome SNPs. **Animal Genetics**, doi: 10.1111/j.1365-2052.2012.02374.x. 2012.
- CAETANO, A.R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectiva para o futuro. **Revista Brasileira de Zootecnia**, v.38, p.64-71, 2009.
- CALVO, J.H.; LOBERA, J.; OSTA, R. et al. Caracterización genética de la raza porcina chato murciano. **Archivos de Zootecnia**, v. 49, p. 53-58, 2000.
- CANUL, S.M.; SIERRA, V.A.; MARTÍNEZ, M.A. et al. Caracterización genética del cerdo pelón mexicano mediante marcadores moleculares. **Archivos de Zootecnia**, n. 54, p. 267-272, 2005.
- CASTRO, S. T. R.; ALBUQUERQUE, M. S. N.; GERMANO, J. L. Census of Brazilian naturalized swine breeds. **Archivos de Zootecnia**, n. 51, p. 1-5, 2002.
- CAVALCANTE NETO, A. **Origem do suíno casco-de-burro e sua relação genética com populações ibéricas e americanas**. Jaboticabal, 2010. **Tese (doutorado)** – Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2010.

- CAVALCANTI, S. S. **Produção de suínos**. Campinas, SP: Instituto Campeiro de Ensino Agrícola, 1985. 453 p.
- CHAIWATANASIN, W.; SOMCHAI, C.; SRISUWAN, C. et al. Genetic Diversity of Native Pig in Thailand Using Microsatellite Analysis. **Kasetsart Journal Natural Science.**, v. 36, p.133 -137, 2002.
- CHANDRASHEKAR, J.; MUELLER, K.L.; HOON, M.A. et al. T2Rs function as bitter taste receptors. *Cell*, v. 100, n.6, p.703-711, 2000.
- CHAUDHARI, N.; ROPER, S. D. The cell biology of taste. *J. Cell Biol.* v.190, p.285–296, 2010.
- CLAYTON D.A. Replication and transcription of vertebrate mitochondrial DNA. **Annual Review of Cell Biology**, v7, pp. 453–478. 1991.
- COLOMBO, M.; TREVISI, P.; GANDOLFI, G. et al. Assessment of the presence of chemosensing receptors based on bitter and fat taste in the gastrointestinal tract of young pig. **Journal of Animal Science**, v. 90, p.128-130, 2012.
- CONTE, C.; EBELING, M.; MARCUZ, A. et al. Evolutionary relationships of the Tas2r receptor gene families in mouse and human. *Physiol Genomics*, v. 14, n 1, p73-82, 2003.
- CONTE, C.; EBELING, M.; MARCUZ, A. et al. Identification and characterization of human taste receptor genes belonging to the TAS2R family. *Cytogenet Genome Res*, v. 98, n 1, p. 45-53. 2002.
- DALCA, A. V.; BRUDNO, M. Genome variation discovery with high-throughput sequencing data. **Briefings in Bioinformatics Advance**. 2010.
- DO, D.N.; STRATHE, A.B.; OSTERSEN, T. et al. Genome-Wide Association Study Reveals Genetic Architecture of Eating Behavior in Pigs and Its Implications for Humans Obesity by Comparative Mapping. **PLoS ONE**, v. 8, n.8, p.71509, 2013.
- DONG, D.; JONES, G.; ZHANG, S. Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evolutionary Biology*, n 9, 12p. 2009.
- ECK, S. H.; ANNA, B-P., FLISIKOWSKI, K. Whole genome sequencing of a single Bos Taurus animal for single nucleotide polymorphism discovery. **Genome Biology**. v.10, 2009.
- EGITO, A.A.; MARIANTE, A.S.; ALBUQUERQUE, M.S.M. Programa Brasileiro de Conservação de Recursos Genéticos Animais. **Archivos de Zootecnia**, v. 51, pp.39-52, 2002.
- EIZIRIK, E.; KIM, J.; MENNOTI-RAYMOND, M. et al. Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, mammalia, Felidae). **Molecular Ecology**. v. 10, p. 65-79. 2001.
- FAO (2007) The State of the World's Animal Genetic Resources for Food and Agriculture – in brief. Commission on Genetic Resources for Food and Agriculture: FAO, Rome, Italy.

- FÁVERO, J. A. E. P. F.; FEDALTO, L. M.; WOLOSZYN, N. A raça de suínos moura como alternativa para a produção agroecológica de carne. **Revista Brasileira de Agroecologia**, v.2, n.1, 2007.
- FAVERO, J. A.; FIGUEIREDO, E. A. P. Evolução do melhoramento genético de suínos no Brasil. **Revista Ceres**, v. 56, n. 4, pp 420-427, 2009.
- FENG, P.; ZHAO, H. Complex evolutionary history of the vertebrate sweet/umami taste receptor genes. v.58, n.18, p 2198-2204. **Chinese Science Bulletin**, 2013.
- FERNÁNDEZ, A. I.; ALVES, E.; FERNÁNDEZ, A. et al. Mitochondrial genome polymorphisms associated with longissimus muscle composition in Iberian pigs. **Journal of Animal Science**. n. 86, 1283-1290, 2008.
- FERREIRA, A.S; Lima, K.R.S. **As raças nacionais de suínos serão extintas?** Ação ambiental, v. 3, p. 24-26, 2001.
- FERREIRA, J.S.B.; PAIVA, S.R.; SILVA, E.C. et al. Genetic diversity and population structure of different varieties of Morada Nova hair sheep from Brazil. **Genetics and Molecular Research**, v.13, p.2480 - 2490, 2014.
- FLEISHMAN, E; LAUNER, A. E.; SWITY, K. R. Rules and exception in conservation genetics: genetic assessment of the endangered plant *Cordylanthus palmatus* and its implications for management planning. **Biological Conservation**, Essex, v. 98, p.45-53, 2001.
- FOLL, M.; GAGGIOTTI, O.E. Identifying the environmental factors that determine the genetic structure of populations. **Genetics**. n.174. pp.875–891. 2006.doi:10.1534/genetics.106.059451.
- FRIES, R., EGGEN, A. STRAZINGER, G. The bovine genome contains polymorphic microsatellites. **Genomics**.n. 8, pp.403-406, 1990.
- GARCÍA, G.; VERGARA, J.; LOMBARDI, R. Genetic characterization and phylogeography of the wild boar *Sus scrofa* introduced into Uruguay. **Genetics and Molecular Biology**, v.34, n2, pp329-337. 2011.
- GIRIO, S.; PEREIRA F. L. Investigation of antibodies to *Leptospira* spp. in wild and feral animals from the region of Nhecolândia, Mato Grosso do Sul, Brazil. Use of the immunohistochemistry technique for the agent detection. **Ciência Rural**,v. 34, n. 1, 2004.
- GONELA, A. **Aplicação de Marcadores Microsatélites de *Sus scrofa domestica* na caracterização genética de populações de *Sus scrofa* sp (porco-Monteiro) e *Tayassu pecari* (queixada)**. Ribeirão Preto, 2003, 88 f. (Tese de Doutorado). Universidade de São Paulo- SP, 2003.
- GRAY M.W. Origin and evolution of mitochondrial DNA. *Annual Review of Cell Biology*. v.5, p. 25–50. 1989.
- GROENEN, M. A.M.; MEGENS, H.-J.; YALDA ZARE, W. C. W. et al. The development and characterization of a 60K SNP chip for chicken. **BMC Genomics**, v.12, n.274, 2011.

- GROSSI, S. F. et al. Genetic diversity in wild (*Sus scrofa scrofa*) and domestic (*Sus scrofa domestica*) pigs and their hybrids based on polymorphism of a fragment of the D-loop region in the mitochondrial DNA. **Genetics and Molecular Research**, Ribeirão Preto, v. 5, n. 4, p. 564-568, 2006.
- HALLIBURTON, R.. Introduction to Population Genetics. Pearson Education International, **New Jersey**. 650p, 2004.
- HEATON, M. P.; HARHAY, G. P., GARY L. et al. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. **Mammalian Genome**, n. 13, p. 272–281, 2002.
- HEDRICK, P.W. **Genetics of population**. 2nd ed. Jones and Bartlett Publishers. Sudbury, Massachusetts. pp: 553, 1999.
- HERRERA, R. C. S. P., SOUZA, R.; HERRERA, H. M.; MAURO, R. de A. Hábitos alimentares do Porco Monteiro (*Sus scrofa*) no Pantanal da Nhecolândia, Mato Grosso do Sul. In: Simpósio sobre recursos naturais e sócio econômicos do Pantanal, 2; 1996, Corumbá. Manejo e conservação. **anais...**Brasília: Embrapa SPI, p67, 1996.
- HUANG W, MARTH G. EagleView: a genome assembly viewer for next-generation sequencing technologies. **Genome Res.**, v.18, n.9, p.1538-43, 2008.
- HU, Z-L.; PARK, C. A.; WU, X.-L. et al. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, v.41 (D1): D871-D879, 2013.
- IRGANG, R.; FAVERO, J.A.; SCHEID, I. **Heterose e complementariedade entre raças na produção de suínos para o abate**. Suinocultura Dinâmica, Ano I, n.3, 6p, 1992.
- International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution International Chicken Genome Sequencing Consortium. **Nature**, n.432, 695–716, 2004.
- KASHI, Y; SOLER, M. Funcional role of microsatellites. **Oxford University Press**, New York. pp. 10-23, 1994.
- KINNAMON, S. C.; CUMMINGS, T. A. Chemosensory transduction mechanisms in taste. *Annu. Rev. Physiol.* v. 54, p 715–731, 1992.
- LARSON, G.; ALBARELLA, U.; DOBNEY, K. et al. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. **Proceedings of the National Academy of Sciences**, v. 104, n.39, p. 15276-15281, 2007.
- LARSON, G.; DOBNEY, K., ALBARELLA, U. et al. Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication. **Science**, v. 307 no. 5715, p. 1618-1621, 2005.

- LEMUS-FLORES, C.; ULLOA-ARVIZU, R.; RAMOS-KURI, M. et al. Genetic analysis of Mexican hairless pig populations. **Journal Animal Science**, v. 79, p. 3021- 3026, 2001.
- LI, H.; HOMER, N. A survey of sequence alignment algorithms for next-generation sequencing. **Briefings in Bioinformatics**, 2010.
- LI, H.; RUAN, J.; DURBIN R. “Mapping short DNA sequencing reads and calling variants using mapping quality scores”. **Genome Res.** n.18, pp.1851–8. 2008.
- LI, D.; ZHANG, J. Diet Shapes the Evolution of the Vertebrate Bitter Taste Receptor Gene Repertoire. **Mol Biol Evol.**, v. 31, p.303-309, 2014.
- LINDROOS, K.; SIGURDSSON, S.; JOHANSSON, K. et al. Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. **Nucleic Acids Research**, v.30, n.70, 2002.
- LIU,G. E. Applications and Case Studies of the Next-Generation Sequencing Technologies in Food, Nutrition and Agriculture. **Recent Patents on Food, Nutrition & Agriculture**, v.1, p. 75-79, 2009.
- LUETKEMEIER, E.S.; SODHI, M.; SCHOOK, L. B.; MALHI, R.S. Multiple Asian pig origins revealed through genomic analyses. **Molecular phylogenetics and evolution.** v. 54,680-6, 2010.
- MACKAY, S., OLIVER, L. D.; LAIPIS, P. J. Template - directed arrest of mammalian mitochondrial DNA synthesis. **Mol. Cell.** n.6, p.1261-126. 1986.
- MANEL, S.,; SCHWARTZ,M.K, LUIKART, G. TABERLET, P. Landscape genetics: combining landscape ecology and population genetics. **Trends Ecol. Evol.** n18. pp.189 -197. 2003.
- MANTEL, N. The detection of disease clustering and a generalized regression approach. **Cancer Research**, v.27, n.2, p. 209-220, 1967.
- MANUNZA, A.; CASELLAS, J.; QUINTANILLA, R. A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants. **BMC Genomics**, v. 15, n. 758, p. 1-12, 2014.
- MARDIS, E. R. Next-generation DNA sequencing methods. **Annu Rev Genomics HumGenet.** n.9, pp.387–402. 2008.
- MARIANTE, A. da S.; EGITO, A.A.; ALBUQUERQUE M. do S.M. et al. Brazilian program on conservation of animal genetics resources. In: IX World congress on animal production. **Proceedings of the ix world conference on Animal Production**, Porto Alegre, RS, World Association of Animal Production, V.CD-Room, 12p, 2003.
- MARIANTE, A. DA. S.; CAVALCANTE, N. Animais do descobrimento: raças domésticas da história do Brasil. Brasília: Embrapa Sede / Embrapa Recursos Genéticos e Biotecnologia. 2006. 274p.

- MAUDET, C.; LUIKART, G.; TABERLET, P.. Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. **Journal Animal Science**, v.80, p.942-950, 2002.
- McMANUS, C.; PAIVA, S. R.; SILVA, A. V. R. Phenotypic Characterization of Naturalized Swine Breeds in Brazil, Uruguay and Colombia. **Brazilian Arch. of Biol. and tech.**, v.53, n. 3, pp.583-591, 2010.
- MEUWISSEN, T. H. E.; HAYES B. J., GODDARD, M. E., Prediction of total genetic value using genome – wide dense marker maps. **Genetics**, v. 157, p.1819 – 1829, 2001.
- MEYERHOF, M.; BORN, S.; BROCKHOFF, A. et al. Molecular biology of mammalian bitter taste receptors. A review. *Flavour Fragr J*, v.26, p 260–268, 2011.
- MIYAKI, C. Y. Filogeografia e a descrição da diversidade genética da fauna brasileira. **Megadiversidade**, v. 5, n.1-2 , p. 96-100, 2009.
- MILLER, J. M.; HALLAGER, S.; MONFORT, S. L. et al. Phylogeographic analysis of nuclear and mtDNA supports subspecies designations in the ostrich (*Struthio camelus*). **Conserv Genet** , v.12, p.423–431, 2011.
- NAGALAKSHMI, U.; WANG, Z.; WAERN, K. et al. The transcriptional landscape of the yeast whole genome defined by RNA sequencing. **Science**, v. 320, p.1344-1349, 2008.
- NICKERSON, D.A.; TAYLOR, S.L.; WEISS, K.M. et al. DNA sequence diversity in a 9.7 kb region of the human lipoprotein lipase gene. **Nat Genet.**, v.19, p.233–240, 1998.
- NIELSEN, R.; WILLIAMSON, S., KIM, Y.; et al. Genomic scans for selective sweeps using SNP data. **Genome Res.** v.15, pp.1566-1575. 2005.
- NONNEMAN, D. J.; ROHRER, G. A.; REMPEL, L. A. et al. Genome-Wide associations for age at puberty in a Duroc-Landrace-Yorkshire swine population. **Plant and Animal Genome Conference Proceedings**. 2011.
- OLIVEIRA, M.C.S.; ALENCAR, M.M.; CHAGAS, A.C.S. et al. Gastrointestinal nematode infection in beef cattle of different genetic groups in Brazil. **Vet. Parasitol.** v.166, p.249-254, 2009.
- PESOLE, G.; GISSI, C., CHIRICO, A.D.E. Nucleotide substitution rate of mammalian mitochondrial genomes. **J. Mol. Evol.**, v. 48, p.427–434, 1999.
- PINTO, A. P. G.; LOPES, P. S., LOPES, M. S. et al. Detecção de locos de características quantitativas nos cromossomos 9, 10 e 11 de suínos. **Rev. Bras. Zootec.**, v.39, n.10, pp.2174-2181, 2010.
- PIRES, A. V.; LOPES, P. S.; GUIMARÃES, S. E. F. Mapeamento de locos de características quantitativas no cromossomo 6, associados às características de carcaça e de órgãos internos de suínos. **Rev. Bras. Zootec.**, v.35, n.4, pp.1660-1668, 2006.
- POWEEL, W.; MORGANTE, M.; CHAZ, A. et al. The comparison of RFLP, RAPD,



- AFLP and (microsatellite) markers for germplasm analysis. **Molecular Breeding**, v. 2, p. 225-238, 1996.
- RANDI, E.; LUCCHINI, V.; Evolutionary genetics of the suiforms as reconstructed using mtDNA sequencing. **Journal of Mammalian Evolution**. n.3. pp. 163-194. 1996.
- RAMIREZ, O.; OJEDA, A., TOMAS, A. et al. Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. **Mol Biol Evol.**, 26:2061–2072, 2009.
- RAMAYO-CALDAS, Y.; CASTELLÓ, A., PENA, R. N. et al. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. **BMC Genomics**, v. 11, n. 1, p. 593, 2010.
- RAMOS, A.M; CROOIJMANS, R.P.M.A.; AFFARA, N.A. 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**, v4, p.6524, 2009,
- ROSA, G. O. BUZATO, IAB, BARROS D R et al. Avaliação do Javonteiro produto resultante do cruzamento do javali com o Porco-Monteiro. In: **Simpósio sobre recursos naturais e socioeconômico do Pantanal**, n3., 2000, Corumbá. Resumos. Embrapa Pantanal, p. 294, 2000.
- RUBIN, C-J.; MEGENS, H-J.; BARRIO, A. M. et al. Strong signatures of selection in the domestic pig genome. **PNAS Early Edition**. 2012.[www.pnas.org/cgi/doi/10.1073/pnas.1217149109](http://www.pnas.org/cgi/doi/10.1073/pnas.1217149109).
- SACHIDANANDAN, R.; WEISSMAN, D.; SCHMIDT, S.C. et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. **Nature**. n. 409. pp. 928-933, 2001.
- SBONER, A.; MU, X. J; GREENBAUM, D. et al. The real cost of sequencing: higher than you think! **Genome Biology**, 2011, 12:125 doi:10.1186/gb-2011-12-8-125.
- SCANDURA, M., IACOLINA, L., APOLLONIO . Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. **Mammal Rev**. 2011, v. 41, n. 2, pp.125–137.
- SCHLOTTERER, C.; TAUTZ, D. Slippage synthesis of simple sequence DNA. **Nucleic Acids Research**. n. 20. pp.211-215. 1992.
- SHEN, Y.; WAN, Z.; COARFA, C. et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. **Genome Res** v. 20, p. 273-280, 2009.
- SILVA, E. C. da. **Diversidade e estrutura genética de populações suínas locais de populações suínas locais no estado de Pernambuco, Brasil**. Universidade Federal Rural de Pernambuco (Dissertação de Mestrado, 110f). 2010.
- SILVA, E. C.; DUTRA JUNIOR, W.M.; IANELLA, P. Patterns of genetic diversity of local pig populations in the State of Pernambuco, Brazil, **R. Bras. Zootec.**, v.40, n.8, p.1691-1699, 2011.

- SILVA FILHA, O. L.; ALVES, D. N. M.; SOUZA, J. F. et al. Caracterização da criação de suínos locais em sistema de utilização tradicional no Estado da Paraíba, Brasil. **Archivos de Zootecnia**, v. 54, pp. 523-528, 2005.
- SILVA, M. P.; MAURO, R. Utilización de pasturas nativas por mamíferos e herbívoros en el pantanal. **Archivos de Zootecnia**, Córdoba, v. 51, n. 193, pp.161-173, 2002.
- SILVA, E. C.; MCMANUS, C. M.; GUIMARÃES, M. P. S. L. M. P. et al. Validation of a microsatellite panel for parentage testing of locally adapted and commercial goats in Brazil. **Genetics and Molecular Biology** (Impresso), v.37, p.54 - 60, 2014.
- SOLLERO, B. P.; PAIVA, S. R.; FARIA, D.A. et al. Genetic diversity of Brazilian pig breeds evidenced by microsatellite markers. **Livestock Science**, v. 13, p.1-8, 2009.
- SOMA, Y.; UEMOTO, Y.; SATO, S. et al. Genome-wide mapping and identification of new quantitative trait loci affecting meat production, meat quality, and carcass traits within a Duroc. **Journal of Animal Science**, v. 89, p. 601-608, 2011.
- SOUZA, C.A. **Padrões de diversidade mitocondrial e nuclear em raças de suínos naturalizadas do Brasil**. Universidade Católica de Brasília (Tese de Doutorado, 141f).2011.
- SOUZA, C. A.; PAIVA, S. R.; PEREIRA, R. W. et al. Iberian origin of Brazilian local pig breeds based on Cytochrome b (MT-CYB) sequence. **Animal Genetics, Oxford**, v. 40, n. 5, p. 759-762, 2009.
- SPRITZE, A.; EGITO, A. A.; MARIANTE, A.S. et al. Caracterização genética da raça bovina Crioulo Lageano por marcadores moleculares RAPD. *Pesquisa Agropecuária Brasileira*, v.38, p.1157-1164, 2003.
- TABERLET, P.; BOUVET, J. Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the bear *Ursus arctos* in Europe. **Proc. R. Soc. Lond. B.** n.255. pp. 195-200, 1994.
- TADANO, R.; SEKINO, M.; NISHIBORI, M. et al. Microsatellite Marker Analysis for the Genetic Relationships Among Japanese Long-Tailed Chicken Breeds. In: **Poultry Science**. v. 1, 2007.
- TAUTZ, D.. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acid Research*, v.17, p.6463-6471, 1989.
- TORGGLER, M.G.F.; CONTEL, E. P. B.; TORGGLER, S. P. Isoenzimas: Variabilidade genética em plantas. Ribeirão Preto. SBG, 1995. 186p.
- TREANGEN, T.; SALZBERG, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. **Rev. Nature**. v. 13, 2012.
- UNITED STATES DEPARTMENT OF AGRICULTURE - USDA. Livestock and Poultry: Word Markets and trade. Miami, 2013. Disponível em: <[http://apps.fas.usda.gov/psdonline/circulars/livestock\\_poultry.pdf](http://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf) >. Acesso em: 30

abril. 2014.

VIANNA, A.T. Os suínos. 14<sup>o</sup> ed. Nobel, 384p. 1985.

\_\_\_\_\_ Os suínos: criação prática e econômica. 12<sup>a</sup> ed. Editora, São Paulo, Brasil, 1986. 384p.

VICENTE, A. A.; CAROLINO, M. I.; SOUSA, M. C. O. et al. Genetic diversity in native and commercial breeds of pigs in Portugal assessed by microsatellites. **Journal Animal Science**, v. 20, 2008.

WANG, M.; SHENG, Z.; WU, X. Mitochondrial DNA- Based Genetic Structure Analysis of Pudong White Pigs. **Biochemical Genetics**. V.48, p.924–37, 2010.

WASSON, J.; SKOLNICK, G.; LOVE-GREGORY, L. et al. Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by pyrosequencing technology. **Biotechniques**. n.32, p.1144–1146. 2002.

WENINK, P.W.; BAKER, A.J.; TILANUS, M.G.J. Global mitochondrial DNA phylogeography of holarctic breeding dunlins (*Calidris alpina*). **Evolution**, v.50, p318-330, 1996.

WERNER, F.A.O.; DURSTEWITZ, G.; HABERMANN, F.A. et al. Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. **Animal Genetics**, n. 35. p. 44-49, 2004.

WIEDMANN, R. T.; SMITH, T.P.L.; NONNEMAN, D. J. SNP discovery in swine by reduced representation and high throughput pyrosequencing. **BMC Genetics**. v.81, n.9, 2008.

WONG, G.T.; GANNON, K.S.; MARGOLSKEE, R.F. Transduction of bitter and sweet taste by gustducin. *Nature*, v. 381, n 6585, p.796-800, 1996.

WU, G.S.; YAO, Y.G.; QU, K.X. et al. Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. **Genome Biol**. v.8, p245, 2007.

YANG, S. L.; WANG, Z. G.; LIU, B. et al. Genetic variation and relationships of eighteen Chinese indigenous pig breeds. **Gen. Sel. Evol**. v. 35, p.657–671, 2003.

YU, G.; XIANG, H. et al. The phylogenetic status of typical Chinese native pigs: analyzed by Asian and European pig mitochondrial genome sequences. **J Anim Sci Biotechnol**, v.4, n.1, p. 9, 2013.

ZHANG, Y.; JELTSCH, A.. The Application of Next Generation Sequencing in DNA Methylation Analysis. **Genes**, v.1., 2010. doi:10.3390/genes1010085.

ZINK. R.M.; REMSEN, J.V. Jr. Evolutionary processes and patterns of geographic variation in birds. **Current Ornithology** 4: 1–69, 1986.

## CAPÍTULO 2

### FILOGEOGRAFIA DO SUÍNO MONTEIRO NO PANTANAL DE NHECOLÂNDIA, MATO GROSSO DO SUL, BRASIL

#### RESUMO

O suíno Monteiro é o principal alvo da caça praticada pelos homens pantaneiros no Brasil. Desde sua introdução tornou-se uma forma alternativa de alimentação protéica para a população da região sazonal e inundável. Como os animais se mantêm livres no campo é esperado que o fluxo gênico e as distâncias físicas causem modificações na variabilidade genética dos grupos populacionais. Diante disso, foi realizado um estudo de filogeografia com 181 amostras de monteiros de 10 pontos geográficos do Pantanal-MS, Brasil, agregando informações moleculares de 19 microssatélites e de um fragmento da região controle do DNA mitocondrial (mtDNA). Nas análises com mtDNA foram incluídas mais 217 amostras de suínos de raças comerciais e locais de outras regiões brasileiras. A partir dos resultados obtidos com microssatélites, constatou-se que existe pouca diferença genética independente da distribuição geográfica desses animais como demonstrado pelos valores de  $F_{ST}$  que variaram entre 0,009 a 0,063. O teste de Mantel indicou que existe pouca correlação entre distância genética e distância geográfica variando de 23,09% ( $P=0,06$ ), usando a distância genética de Nei, a 24,66% usando o índice de diferenciação  $F_{ST}$  ( $P=0,055$ ). Apesar de essas populações apresentarem uma diversidade genética moderada, a inferência Bayesiana revelou clusters que separam os monteiros do Pantanal em, pelo menos, três populações. Nas análises realizadas com a região controle do mtDNA verificou-se que, apesar de os Monteiros pertencerem a um único haplótipo europeu, esses animais não possuem um haplótipo exclusivo por compartilharem seu principal haplótipo com outras populações suínas (comerciais e outras raças locais). Os resultados demonstraram que os monteiros de pontos geográficos diferentes não estão isolados geneticamente a partir de informações obtidas com microssatélites e DNA mitocondrial devido à presença de fluxo gênico entre as populações ocasionado pela capacidade de dispersão dos animais na região a procura de alimentos e água, ou em resposta às enchentes.

**PALAVRAS CHAVES:** distância geográfica, diversidade genética, microssatélites, mtDNA

**CHAPTER 2**  
**PHYLOGEOGRAPHY OF THE MONTEIRO PIG IN THE PANTANAL OF**  
**NHECOLÂNDIA, MATO GROSSO DO SUL, BRAZIL**

**ABSTRACT**

Monteiro is a pig quite exploited by men Pantanal of Brazil for a long time and has become an alternative form of protein food for them. As are animals bred freely in the field is expected that gene flow and the physical distances cause changes in genetic variability which is very important for the survival of a species and/or breed. Therefore, a study of phylogeography with 181 samples of Monteiro pigs from 10 geographical points of Pantanal-MS, Brazil was conducted, using molecular information of 19 microsatellites and of a fragment of the mitochondrial DNA control region. In the analyses with mtDNA also 217 samples of pigs from commercial and local breeds around the country were included. From the results obtained with microsatellites, it was found that there is little genetic difference regardless of the geographic distribution of these animals as demonstrated by  $F_{ST}$  values that varied from 0.009 to 0.063. The Mantel test indicated low correlation with geographic distance ranging from 23.09% ( $P = 0.06$ ) using a genetic distance of Nei, to 24.66% using the differentiation index  $F_{ST}$  ( $P = 0.055$ ). Although these populations present a moderate genetic diversity, Bayesian inference revealed clusters that separate Monteiro pigs in at least three populations. In the analyzes performed with the mtDNA control region was found that, although the Monteiros belong to a single European haplotype, these animals do not possess a unique haplotype for sharing your main haplotype with other (commercial and other local breeds) pig populations as the gene flow between them. The results showed that monteiros from different geographical points are not genetically isolated from the information obtained from microsatellite and mitochondrial due to the presence of gene flow among populations caused by the dispersion ability of the animals in the region looking for food and water.

**KEY-WORDS:** gentic diferentiation, genetic diversity, microsatellites, mtDNA

## 1. INTRODUÇÃO

O suíno Monteiro encontrado no Pantanal brasileiro é uma raça asselvajada e se originou de suínos trazidos pelos colonizadores para a cidade da região de Corumbá, Mato Grosso do Sul na segunda metade do século XVIII. Os Monteiros possuem características muito semelhantes as do javali (*Sus scrofa scrofa*), apresentando conformação do corpo e cabeça em formas de cunha, coloração preto acinzentado ou marrom, orelhas pequenas e eretas, membros fortes e ágeis (Herrera et al., 1996).

Os suínos Monteiros além de contribuírem para a preservação de outras espécies nativas contribuem na alimentação da população local no Pantanal brasileiro, porém sofrem ameaças causadas por fatores antropológicos e climáticos, os quais podem interferir na diversidade genética entre as populações. O homem interfere na reprodução destes animais, uma vez que os machos são castrados para engorda e posterior abate, enquanto as fêmeas, na maioria das vezes, são preservadas dependendo da fase reprodutiva. A variação sazonal no Pantanal também pode ser um fator que causa redução do tamanho populacional de Monteiros, uma vez que a caça é mais praticada no período de seca (Desbiez et al., 2010).

Uma alternativa para estudar as populações animais livres na natureza é a filogeografia. A filogeografia surgiu como uma disciplina com o objetivo de estudar a variação genética em populações de uma espécie ou entre espécies próximas utilizando ferramentas moleculares, como enzimas, RAPDs, RFLPs, cromossomo Y, DNA mitocondrial (mtDNA), SNPs e os microssatélites (Avise et al., 1987; Avise, 1998). As informações obtidas com os marcadores genéticos são úteis nos trabalhos de estrutura geográfica e genética das espécies que podem ser definidas como o resultado dos processos microevolutivos e demográficos que atuam dentro e entre populações (Wenink et al., 1996).

Várias espécies de plantas e animais domésticos foram alvos de estudos de filogeografia (Eizirik et al., 2001; Joshi et al., 2004; Miller et al., 2011; Klütsch et al., 2012). Para suínos, um estudo dessa natureza demonstrou a existência de múltiplos centros de

domesticação da espécie em toda a Eurásia e que os javalis originaram os porcos domésticos europeus (Larson et al., 2005). García et al. (2011) identificaram (com marcadores do mtDNA), elevados níveis de diferenciação e baixa taxa de migração entre populações de javalis colombianos de diferentes localidades, sugerindo que cada população do estudo poderia ser considerada uma única unidade de manejo. O objetivo deste trabalho foi caracterizar e comparar os padrões de distribuição geográfica da variabilidade genética em populações de suínos monteiros do Pantanal brasileiro utilizando polimorfismos de microssatélites e da região controle do mtDNA.

## **2. MATERIAL E MÉTODOS**

### **2.1 Amostragem**

Um total de 406 amostras de suínos foi utilizado neste estudo. Desse total, 189 amostras foram de suínos monteiros, das quais 181 provieram de dez pontos geográficos (PG) identificados pelas coordenadas geográficas em Nhecolândia, sub-região do Pantanal, Mato Grosso do sul, Brasil (Tabela 1, Figura 1) e oito amostras foram obtidas de propriedades particulares de Brasília, DF (MOB). Todas as amostras provenientes do Pantanal foram georreferenciadas utilizando o GPS Garmin 60 (Datum-WGS 84). As 189 amostras foram genotipada com microssatélites. As demais amostras (n=217) de raças suínas localmente adaptadas (n = 168), comerciais (n = 29) e suínos mestiços (n = 06) foram obtidas dos estudos de Cavalcante Neto (2010) e Souza (2011) e foram incluídas nas análises realizadas com sequências da região controle do DNA mitocondrial de 122 suínos Monteiros do Pantanal e seis suínos MOB.

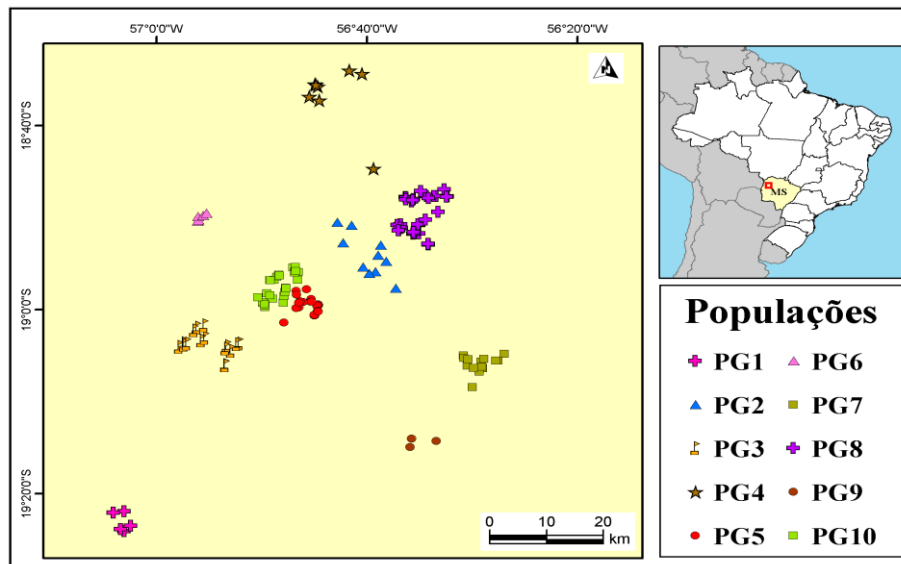
### **2.2 Genotipagem e análises estatísticas dos microssatélites**

As reações em cadeia da polimerase (PCR) de 19 microssatélites (Tabela 2) foram feitas em termociclador (Tabela 1A do ANEXO A). Nessas reações foi utilizado o Kit para PCR multiplex da Qiagen (contendo HotstartTaq DNA Polimerase, Tampão para PCR



Tabela 1 Populações, número de suínos monteiros amostrados de cada ponto geográfico e as coordenadas geográficas.

Populações	Nº de animais amostrados	Coordenadas geográfica	
		Latitude Sul	Longitude Oeste
MOB	08	-	-
PG1	05	19° 23' 06"	57° 03' 14"
PG2	16	18° 53' 49"	56° 39' 38"
PG3	16	19° 03' 45"	56° 55' 03"
PG4	18	18° 36' 04"	56° 43' 19"
PG5	26	18° 59' 28"	56° 45' 49"
PG6	10	18° 49' 58"	56° 55' 48"
PG7	15	19° 05' 52"	56° 29' 24"
PG8	36	18° 49' 52"	56° 35' 09"
PG9	04	19° 14' 24"	56° 34' 39"
PG10	35	18° 57' 45"	56° 48' 35"
Total	189	-	-



**Figura 1** Distribuição das populações amostradas de suínos monteiros do Pantanal-MS, Brasil.

multiplex com MgCl<sub>2</sub>, mix de dNTP, água RNase-free e QSolution). As condições para amplificação foram: desnaturação a 95°C por 15 minutos, 30 ciclos para desnaturação a 94°C por 1 minuto, anelamento a 57°C por 1 minuto, extensão a 72°C por 1 minuto e 30 segundos, e uma etapa de extensão final a 72°C por 10 minutos.

Os produtos das reações de PCR dos microssatélites foram desnaturados com um mix contendo um padrão interno ROX e formamida HIDI a 95°C por cinco minutos e imediatamente deixados no gelo por cinco minutos e submetidos à eletroforese capilar no sequenciador automático de DNA do modelo ABI PRISM 3730 (*Applied Biosystems*). Em seguida, os alelos dos 19 locos de microssatélites (Tabela 2) foram identificados com o software GeneMapper v. (*Applied Biosystems*) e o software FlexiBin v. 2.0 (Amos et al., 2006) foi utilizado para determinação das classes alélicas.

A distância genética entre as populações de suínos Monteiros foi calculada com o programa Arlequin 2000 v. 3.5 (Schneider et al., 2000). Nessa análise, as populações foram submetidas à comparação par a par, de acordo com suas frequências alélicas para verificação de diferenciação entre elas pelo  $F_{ST}$  (Weir & Cockerham, 1984). Além do índice  $F_{ST}$  foram calculadas as distâncias genéticas de Nei (1978) por ser uma das mais utilizadas nos estudos populacionais com o programa GenAlex. Essas distâncias genéticas foram utilizadas para construir árvores filogenéticas baseadas no algoritmo neighbour-joining (Saitou & Nei, 1987) com *software* SplitsTree (Huson & Bryant, 2006).

O teste de Mantel (1967) foi realizado para verificar a existência de correlação significativa entre as matrizes de valores de distâncias genéticas e geográficas (em km). A matriz de distâncias geográficas e o teste de Mantel foram realizados com GenAlex. Os níveis de significância para cada prova foram obtidos mediante 10.000 permutações.

Para verificar se os indivíduos amostrados pertenciam realmente à população pré-definida foi realizado um teste de designação individual. Esse tipo de análise é muito importante e instrutivo para fins de conservação já que cada indivíduo analisado é alocado a uma população genética determinada que possa ou não coincidir com a sua população de origem. Essa análise foi desempenhada com o programa GenAlex utilizando as informações dos indivíduos alocados nas populações a partir dos pontos geográficos como já descritos anteriormente.

As estruturas genética e geográfica foram analisadas usando a Análise de Variância Molecular (AMOVA, Excoffier et al., 1992) com o Arlequin 3.1. Essa análise foi desempenhada de três formas: 1) um grupo com as 11 populações de Monteiros; 2) um grupo com as 10 populações de Monteiros do Pantanal para ser contrastado com a população de

MOB e 3) um grupo apenas com as 10 populações de Monteiros do Pantanal. O método Bayesiano também foi utilizado para analisar a estrutura genética e geográfica dos suínos Monteiros distribuídos no Pantanal usando *software* Structure (Prichard et al., 2000). Nesse programa foram estimados o número e grupos genéticos reais (K). A análise foi realizada utilizando os modelos de frequências alélicas correlacionadas e de ancestralidade por mistura com o número de K variando de 1 a 20, período de *burn-in* de 100.000 e cadeia de Markov e Monte Carlo (MCMC) de 400.000; para cada k utilizaram-se cinco iterações. A determinação do número de k foi estimado com uma estatística DeltaK, a qual se baseia na taxa de mudança do logaritmo da probabilidade dos dados entre sucessivos valores de K (Evanno et al., 2005). O programa DISTRUCT (Rosenberg, 2004) possibilitou a visualização gráfica dos resultados obtidos com o *Structure*, buscando mostrar por meio de diferentes cores a composição das amostras populacionais.

### 2.3 Sequenciamento da região controle do mtDNA e análises estatísticas

Na PCR da região controle do DNA mitocondrial, flanqueada pelos *primers* direto 5' -CGCCATCAGCACCCAAAGCT - 3' e reverso 5' - TGGGCGATTTTGGTGAGATGGT -3' (Alves et al., 2003) foram utilizados reagentes da Promega (Taq Polimerase e Buffer) e as condições para amplificação foram: desnaturação a 95°C por 3 minutos, 30 ciclos para desnaturação a 95°C por 1 minuto, anelamento a 55°C por 1 minuto, extensão a 72°C por 1 minuto e uma etapa de extensão final a 72°C por 30 minutos. Os produtos de PCR foram purificados mediante os reagentes ExoSAP-IT® (USB® Product Affymetrix) a 37°C por 30 minutos e 80°C por 20 minutos para eliminar os dNTPs e *primers* restantes da PCR, assim como produtos incompletos. Posteriormente, as reações de sequenciamento foram realizadas pelo método de incorporação de dideoxynucleotídeos (ddNTP) com o kit BigDye® v.3 (*Applied Biosystems*) para realizar a PCR de sequenciamento seguindo as recomendações do fabricante (Tabela 1A do ANEXO A).

Tabela 2 Nome dos loci, cromossomos (Cr.) seqüências dos primers e tamanho esperado (Te) obtido neste trabalho, tipo de marcação fluorescente (Marc.) e multiplex (M).

Locos	Cr.	Seqüências <i>primers</i> 5' ->3'	Te	**Mar.	Mult.
<sup>3</sup> S0026*	16	AACCTTCCCTTCCCAATCAC	84-102	F	A
		CACAGACTGCTTTTTACTCC			
<sup>4</sup> S0155*	1	AACCTTCCCTTCCCAATCAC	142-162	H	A
		CACAGACTGCTTTTTACTCC			
<sup>1</sup> SW830*	1	AAGTACCATGGAGAGGGAAATG	174-204	H	A
		ACATGGTTCCAAAGACCTGTG			
<sup>3</sup> S0097*	4	GACCTATCTAATGTCATTATAGT	209-250	F	A
		TTCCTCCTAGAGTTGACAAACTT			
<sup>7</sup> S0228*	6	GGCATAGGCTGGCAGCAACA	220-246	H	A
		AGCCCACCTCATCTTATCTACACT			
<sup>5</sup> SW1517	2	CAAATGATTTTATCCATCCTTGC	119-154	F	A
		TCTTAGTGATGCATTCTTAAGCTG			
<sup>2</sup> S0002*	3	GAAGCCAAAGAGACAACCTGC	191-219	H	B
		GTTCTTTACCCACTGAGCCA			
<sup>7</sup> S0355*	15	CTCAGTTCTTTGGGACTGAACC	244-271	H	B
		CATCTGTGGAAAAAAAAAGCC			
<sup>6</sup> SW936*	15	TCTGGAGCTAGCATAAGTGCC	94-112	H	B
		GTGCAAGTACACATGCAGGG			
<sup>6</sup> SW2406*	6	AATGTCACCTTTAAGACGTGGG	222-262	F	B
		AATGCGAAACTCCTGAATTAGC			
<sup>1</sup> SW24*	17	CTTTGGGTGGAGTGTGTGC	93-121	F	C
		ATCCAAATGCTGCAAGCG			
<sup>1</sup> OPN	8	CCAATCCTATTCACGAAAAAGC	142-164	H	C
		CAACCCACTTGCTCCCAC			
<sup>1</sup> S0005*	5	TCCTTCCCTCCTGGTAACTA	203-241	H	C
		GCACTTCCTGATTCTGGGTA			
<sup>3</sup> S0101*	7	GAATGCAAAGAGTTCAGTGTAGG	197-221	F	C
		GTCTCCCTCACACTTACCGCAG			
<sup>1</sup> S0068*	3	AGTGGTCTCTCTCCCTCTTGCT	211-260	H	D
		CCTTCAACCTTTGAGCAAGAAC			
<sup>4</sup> S0090*	11	CCAAGACTGCCTTGTAGGTGAATA	227-249	F	D
		GCTATCAAGTATTGTACCATTAGG			
<sup>1</sup> SW72*	3	ATCAGAACAGTCGCCGT	103-113	H	D
		TTTGAAAATGGGGTGTTTCC			
<sup>1</sup> SW455	4	CCTCCCTGGCACTCATTG	184-206	F	D
		CACACACACAAGCAGGTGC			
<sup>1</sup> SW911*	9	CTCAGTTCTTTGGGACTGAACC	149-173	F	D
		CATCTGTGGAAAAAAAAAGCC			

\*Locos recomendados pela FAO/ISAG (2004), \*\* F = 6<sup>3</sup>FAM (azul); H= HEX (verde)

<sup>1</sup> Roher et al. (1994), <sup>2</sup>Fredholm et al. (1993), <sup>3</sup>Roher et al. (1996), <sup>4</sup>Ellegren et al. (1994), <sup>5</sup>Alexander et al. (1996a), <sup>6</sup>Alexander et al. (1996b). <sup>7</sup>Robic et al. (1994).

Os produtos de sequenciamento foram purificados em duas etapas: 1) etapa – com 3µL de EDTA a 125mM e 25µL de etanol a 100% e posterior centrifugação a 4°C por 30 minutos a 3.000 rpm. 2) etapa - foi feita adicionando 30µL de etanol 70% e posterior centrifugação – 4°C por 15 minutos a 3.000 rpm. Em seguida, o sobrenadante foi descartado e após secagem dos produtos de sequenciamento à temperatura ambiente, foi feita a ressuspensão com 10µL de formamida Hi-Di (*Applied Biosystems*) e desnaturação a 95°C por cinco minutos. Os produtos gerados na reação de sequenciamento foram lidos mediante a eletroforese capilar no sequenciador automático ABI Prism 3100 Genetic Analyzer® (*Applied Biosystems*), utilizando o polímero Pop-6TM (*Applied Biosystems*).

Um fragmento de 467 pb da região controle, entre as posições 15.543 e 15.840pb do mtDNA completo da espécie *Sus scrofa domesticus*, foi sequenciado e editado. A edição das sequências de DNA se realizou mediante o programa SeqScape v. 2.5 (*Applied Biosystems*), no qual as sequências foram inicialmente avaliadas quanto à qualidade e editadas manualmente para identificação de possíveis erros. O alinhamento das sequências foi realizado no Mega v. 5 (Tamura et al., 2005). Após a edição e o alinhamento das sequências da região controle do mtDNA foram obtidas 331 sequências: 122 sequências de suínos Monteiros do Pantanal, seis de Monteiros de Brasília (grupo externo = MOB) e 203 de raças suínas localmente adaptadas (n = 168), comerciais (n = 29) e suínos mestiços (n = 06) que foram analisadas pelos autores Cavalcante Neto (2010) e Souza (2011).

Os níveis de variabilidade genética foram quantificados pela diversidade de nucleotídeos ( $\pi$ ) e haplótipos (h) com o programa DnaSP v 5 (Librado & Rozas et al. 2009). O programa Arlequin 3.11 (Schneider et al., 2000) foi utilizado para verificar a possível existência de populações estruturadas de Monteiros entre os diferentes pontos geográficos do Pantanal e avaliar o grau de variabilidade genética dentro e entre populações por meio de uma Análise de Variância Molecular (AMOVA). Distâncias genéticas foram utilizadas para construir dendogramas baseados no algoritmo neighbour-joining (Saitou & Nei, 1987) com *software* SplitsTree (Huson & Bryant, 2006).

### 3. RESULTADOS E DISCUSSÃO

#### 3.1 Resultados

##### 3.1.1 Diversidade genética intrapopulacional analisada com microsatélite

O conjunto de 19 microsatélites mostrou variabilidade alélica de 4 a 11 alelos por loco com uma média de 6,789. O número mínimo de alelos diferentes por loco foi igual a 4, como sugerido por Barker (1994) para estudo de diversidade genética (Tabela 2A do ANEXO A). Esses resultados foram congruentes com os valores obtidos por Sollero et al. (2009) e por Silva et al. (2011) que analisaram microsatélites em populações suínas brasileiras.

O número de alelos diagnósticos é uma estimativa indireta do fluxo gênico e foram detectados 21 nas populações MOB, Pop3, Pop8 e Pop10 (Tabela 3A do ANEXO A). Alelos desse tipo surgem por mutação e podem ser fixados por evento de deriva genética em uma população, servindo para caracterizar as populações mais antigas, e a sua completa ausência caracteriza populações originadas recentemente (Slatkin, 1985). Caso seja detectada ausência de fluxo gênico, esses alelos poderão ser perdidos, principalmente, aqueles raros com frequência inferior a 5% detectados nas populações de Monteiros do Pantanal.

Os valores médios dos parâmetros  $N_m$ ,  $N_{ae}$ ,  $N_{ar}$ ,  $H_o$  e  $H_e$ , e  $F_{IS}$  (Tabela 3) para as populações do Pantanal (4,6791, 3,0784, 3,4037, 0,6303, 0,6331 e 0,0427) foram levemente inferiores aos obtidos para a MOB (5,421, 3,896, 4,165, 0,611, 0,717 e 0,192). A população MOB e Pop1 apresentaram os maiores valores para  $F_{IS}$  (0,192, 0,118) e foram

Tabela 3 Parâmetros de diversidade intrapopulacional para o conjunto de 19 locos de microssatélites.

Populações	Nam	Nae	Nar (8)	Ho	He	F <sub>IS</sub>	EHW
MOB	5,421	3,896	4,165	0,611	0,717	0,192	0,005*
Pop1	3,684	2,729	3,412	0,600	0,604	0,118	0,128
Pop2	5,316	3,360	3,557	0,620	0,649	0,079	0,034*
Pop3	4,632	2,985	3,328	0,599	0,620	0,058	0,070
Pop4	4,684	3,074	3,342	0,632	0,646	0,045	0,160
Pop5	5,053	3,205	3,374	0,670	0,652	-0,014	0,658
Pop6	4,263	3,139	3,437	0,658	0,655	0,048	0,158
Pop7	5,211	3,148	3,304	0,629	0,638	0,012	0,307
Pop8	3,474	2,734	3,474	0,671	0,594	0,013	0,460
Pop9	4,421	3,007	3,275	0,637	0,614	-0,031	0,755
Pop10	6,053	3,403	3,534	0,587	0,659	0,099	0,000*
Total (n=11)	4,746	3,152	3,472	0,628	0,640	0,056	
Mont. Pantanal	4,679	3,078	3,403	0,630	0,633	0,042	

Nam = número médio de alelos; Nae = número efetivo de alelos; Número de alelos corrigidos pelo método de rarefação; Ho = heterozigosidade observada; He = heterozigosidade esperada segundo o EHW; F<sub>IS</sub> = índice de consanguinidade intra-populacional; EWE = Equilíbrio de Hardy-Weinberg. \*P<0,05.

significativos (P<0,05), provavelmente, pelo excesso de homozigoto como demonstrado pelo valor da He esperada (0,717), que foi maior do que a Ho (0,611). A média de alelos por população variou de 3,47 (Pop8) a 6,053 (Pop10), o número efetivo de alelos variou de 2,72 (Pop1) a 3,89 (MOB), a qual também apresentou maior número de alelos corrigido pelo método de rarefação (Nar =4,165). Duas populações do Pantanal (Pop5 e Pop 9) demonstraram excesso de heterozigoto (F<sub>IS</sub> = -0,014, -0,031), porém não foi significativo (P>0,05).

Os valores de F<sub>IS</sub> positivos levantam a hipótese de que há acasalamentos entre indivíduos aparentados nas populações. Diante disso, realizou-se um teste de parentesco (*relatedness*) pelo método de Blouin (2003). Os valores dos índices de parentesco (*r*) igual ou maiores que 0,25, indicam que há acasalamentos entre animais parentes próximos dentro das populações (Figura 1A do ANEXO A). A Pop1 do Pantanal apresentou mais variações entre os indivíduos nas frações de alelos idênticos por descendência (-0,25 a 0,18) com uma média

aproximada de 0,05. Os indivíduos das populações Pop2 (-0,012) e Pop8 (-0,045) não compartilham alelos por descendência, e para as populações Pop3, Pop4, Pop5, Pop7, Pop9 e Pop10 os valores de  $r$  foram significativamente próximo de 0.

### **3.1.2 Teste de alocação dos indivíduos a população de origem**

A coleta de animais criados a campo pode afetar os estudos filogeográficos ou genéticos de uma população por não se ter informação exata da origem dos animais. Como foi observado, maior percentual dos animais (73%) não foi atribuído às populações de origem que foram amostradas, o que pode ser considerado um entrave no momento de elaborar estratégias de conservação (Tabela 4A do ANEXO A). Por outro lado, foi verificado um ponto positivo a partir do teste de parentesco, que mesmo a coleta sendo feita a campo, sem o conhecimento da genealogia, o nível de parentesco apresentado entre os indivíduos de cada população do Pantanal foi baixo, significando que as populações foram originadas por indivíduos pouco relacionados. Essas informações foram importantes e podem ser consideradas como uma ferramenta para elaboração de estratégias de conservação visando evitar a consanguinidade nas populações por meio de seleção de matrizes e reprodutores menos relacionados para impedir a perda de variabilidade genética.

### **3.1.3 Análises filogeográficas**

Para definir as medidas de manejo e conservação que devem ser direcionadas a uma dada população, faz-se necessário ter o conhecimento da composição genética da espécie em estudo e de como ela está distribuída nas populações para verificar se há estruturação. A estruturação genética nas populações pode ser uma causa natural ou uma consequência da ação do homem (Frankham et al., 2002).

No estudo filogeográfico dos suínos monteiros no Pantanal foi gerada uma matriz de distância geográfica (Tabela 5A do ANEXO A) entre as populações do Pantanal. A distância média, em km, entre os pontos geográficos variaram de 28,42km a 63,04km. As populações amostradas de pontos geográficos mais próximos e mais distantes foram Pop5 e



Pop10 (5,77km) e Pop1 e Pop4 (93,513km), respectivamente. Essa matriz foi utilizada na análise de correlação genética para verificação de isolamento genético causado pela distância geográfica.

O estimador de subdivisão populacional ( $F_{ST}$ ), que também pode ser uma boa medida de distância genética, foi utilizado para analisar a diferenciação entre os pares de população (Tabela 6A do ANEXO A). Constatou-se que nas amostras populacionais do Pantanal existe pouca diferença genética independente da distribuição geográfica como demonstrado pelos baixos valores de  $F_{ST}$  que variou de 0,009 a 0,063.

A população MOB foi mais diferenciada das populações de suínos Monteiros do Pantanal, especialmente, das populações 1 e 8 ( $F_{ST}=0,080$  e  $0,088$ , respectivamente). Entre as populações do Pantanal, verificou-se que a Pop1 foi a mais divergente das populações Pop3 e Pop8 ( $F_{ST}=0,063$  e  $0,062$ , respectivamente). A Pop1 encontra-se geograficamente mais distante da Pop4 e Pop8 (93,5 e 78,6 km, respectivamente), levando a inferir que o fluxo de genes pode estar sendo impedido entre a Pop1 e Pop3 por alguma barreira geográfica, uma vez que estavam geograficamente mais próximas (38,9 km). A presença de uma estrada entre os dois pontos geográficos (Pop1 e Pop3) pode estar impedindo o fluxo de genes entre elas.

O menor valor de  $F_{ST}$  foi entre as Pop5 e Pop7, porém não foi significativo ( $P>0,05$ ). Resultados semelhantes aos obtidos com o  $F_{ST}$  foram verificados com a distância genética  $D$  de Nei (Tabela 7A do ANEXO A). A árvore filogenética gerada (Figura 2) mostrou que as populações tenderam a se agrupar quando foram fisicamente próximas e pouco diferenciadas (grupo I), fisicamente distantes e mais diferenciadas (grupo II) e fisicamente distantes e pouco diferenciadas (grupo III) independente da distância utilizada.

O teste de Mantel indicou que independente do tipo de distância genética empregada entre as populações de suínos Monteiros do Pantanal existiu pouca correlação com a distância geográfica variando de 23,09% ( $P=0,06$ ), usando a distância genética  $D$  (Figura 3A), a 24,66% quando foi utilizada a  $F_{ST}$  ( $P=0,055$ ; Figura 3B). Uma possível homogeneidade genética pode ser notada para os valores obtidos para as distâncias por permutações, sugerindo mais uma vez a presença de fluxo gênico.

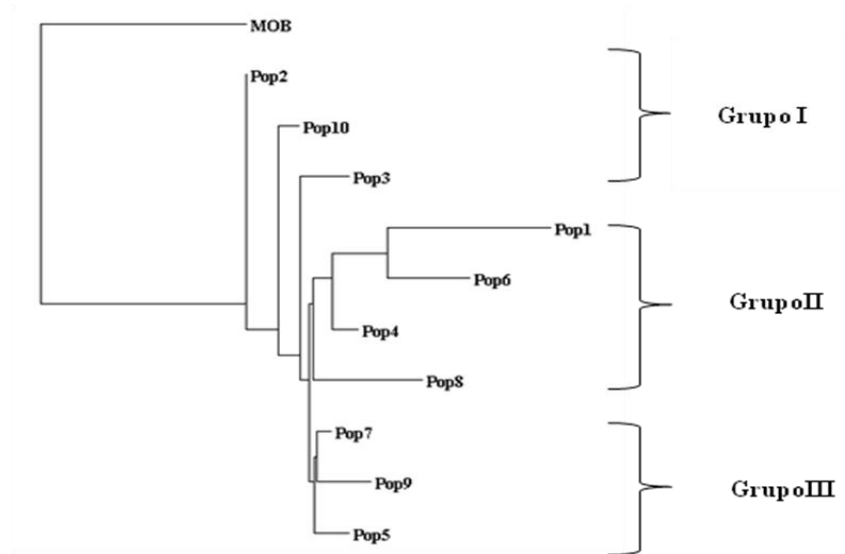


Figura 2 Árvore filogenética construída pelo método Neighbor-Joining baseando-se no índice  $F_{ST}$ . Grupos: I (fisicamente próximos e pouco diferenciadas); II (fisicamente distantes e mais diferenciadas) e III (geograficamente distantes e pouco diferenciadas).

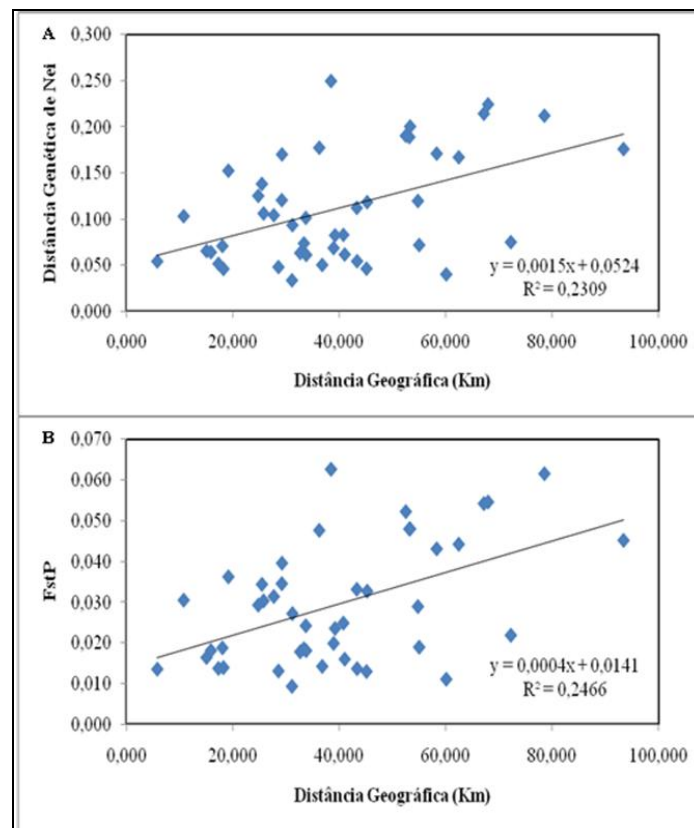


Figura 3 Teste de Mantel entre distâncias genéticas de Nei e geográficas (A) e entre  $F_{ST}$  e distâncias geográficas (B).

### 3.1.4 Definição da estrutura genética e geográfica

A maior parte da variação foi retida dentro das populações (91 a 98%) de acordo com os resultados da AMOVA. Uma fraca estruturação genética ( $F_{ST} < 0,02$ ,  $P=0,000$ ) foi verificada entre as populações de Monteiros com ou sem a inclusão da MOB (Tabela 4), provavelmente, causada pelo intenso fluxo de genes. Uma estruturação moderada ( $F_{ST}=0,085$ ;  $P=0,000$ ) foi verificada quando a análise foi realizada entre a MOB e as populações do Pantanal. A presença de alelos diagnósticos identificados nos Monteiros de Brasília pode justificar esta diferença entre as duas populações.

A análise Bayesiana mostrou que o número provável de populações, representado pelo número de K segundo Evanno et al. (2005), foi 4, porém houve alta probabilidade para  $K=2$  (Figura 2A do ANEXO A). Essa análise de agrupamento de populações diferenciou claramente as populações MOB e Pantanal para  $K=2$  (Figura 4) e seus alelos foram agrupados, em maior parte, no cluster 1(0,7588, Tabela 8 A do ANEXO A), enquanto os monteiros do Pantanal foram distribuídos mais no cluster 2 com altas proporções, variando entre 0,8102 (Pop2 ) e 0,9918 (Pop6).

### 1.1.5 Variação da região controle do DNA mitocondrial

Após a edição e o alinhamento das sequências da região controle do mtDNA foram obtidas 331 sequências: 122 sequências de suínos Monteiros do Pantanal, seis de Monteiros de Brasília (grupo externo = MOB) e 217 sequências da região controle analisadas pelos autores Cavalcante Neto (2010) e Souza (2011) em raças suínas localmente adaptadas ( $n = 168$ ), comerciais ( $n = 29$ ) e suínos mestiços ( $n = 06$ ).

O fragmento da região controle do mtDNA foi de 363 pb e apresentou 23 sítios polimórficos caracterizados por 20 transições, uma transversão ( $A \leftrightarrow T$ ) e uma deleção em muitas sequências bem como a presença de uma deleção em todos os indivíduos na posição 15707 pb do mtDNA quando comparado com a sequência referência do GenBank (acesso AJ002189, Ursing e Arnason 1998).

Tabela 4 Estrutura genética analisada pela Análise de Variância Molecular (AMOVA).

Estrutura	Fontes de Variação	G.L.	SQ	Componentes de variância	% de variação	F <sub>ST</sub>
Análise I: 11 populações	Entre populações	10	99,111	0,11467	1,84	0,018*
	Dentro das populações	367	2245,347	6,11811	98,16	
Análise II: 10 populações	Entre populações	9	76,320	0,06875	1,12	0,012*
	Dentro das populações	352	2142,222	6,08586	98,88	
Análise III: Geografia	Entre grupos	1	22,791	0,50560	7,56	0,085 <sup>ns</sup>
	Entre populações dentro dos grupos	9	76,320	0,06783	1,01	
	Dentro das populações	367	2245,347	6,11811	91,43	

\*P&lt;0,05; ns= P&gt;0,05

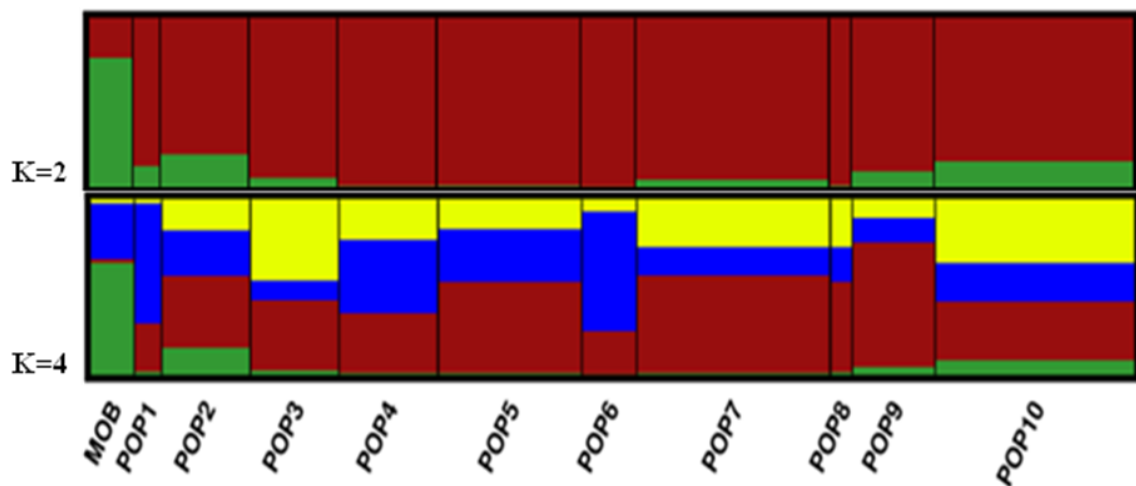


Figura 4 Proporção de cada população designada a cada cluster para K2 e K4. As quatro cores são referentes às populações definidas (*clusters*) com os resultados do programa STRUCTURE (Pritchard et al., 2000) e visualizadas com o programa DISTRUCT (Rosemberg, 2004): verde=cluster1; vermelho=cluster 2; amarelo=cluster 3 e azul=cluster 4.

Na população total composta por 331 animais foi detectada baixa diversidade nucleotídica ( $0,00797 \pm 0,00093$ ) e moderada diversidade haplotípica ( $0,483 \pm 0,041$ ). A população de suínos Monteiros do Pantanal teve os menores valores para diversidade nucleotídica ( $0,00045 \pm 0,00044$ ) e haplotípica ( $0,016 \pm 0,016$ ), e os maiores valores foram para a população de suínos comerciais ( $0,01874 \pm 0,00186$  e  $0,884 \pm 0,030$ ), cuja diversidade haplotípica pode ser considerada alta (Tabela 9A do ANEXO A). Em todas as regiões fisiográficas observou-se pouca diversidade nucleotídica que variou entre  $0,01736 \pm 0,0041$  para a região Sul e  $0,02018 \pm 0,0029$  para a Centro-Oeste, enquanto isso, os valores de diversidade haplotípica foram considerados altos em todas as regiões, variando entre  $0,964 \pm 0,051$  a 1 (Tabela 10A do ANEXO A).

### **3.1.6 Relações filogeográficas**

Os 23 sítios polimórficos geraram 21 haplótipos. Os haplótipos H1, H13, H14 a H21 e H5 a H8 foram exclusivos das raças localmente adaptadas e comerciais, respectivamente. Os demais haplótipos foram compartilhados entre as diferentes populações (H2, H3, H4 e H9 a H12). O H2 foi o de maior frequência e distribuição (234 indivíduos das cinco populações e das quatro regiões), seguido pelos H7, H4 e H13. Esse último foi exclusivo dos suínos locais. O H2 também foi identificado em cinco dos seis suínos Monteiros de Brasília e foi o único compartilhado entre a população do presente estudo e as populações de Cavalcante Neto (2010) e Souza (2011) (Tabela 5).

As regiões nordeste e sul, com uma amostragem de 70 e 42 sequências, concentraram maior número de haplótipos ( $n=10$ ) e as regiões sul e sudeste concentram maior número de haplótipos exclusivos 5 (H5 e H8 a H11) e 3 (H16 a H18), respectivamente. Nas regiões sul e sudeste foram amostrados suínos das raças comerciais Landrace e Large White e suínos do fenótipo “casco de burro”, respectivamente. A presença de haplótipos exclusivos pode ser explicada pela falta de fluxo gênico entre essas populações (Figura 3A do ANEXO A).

Tabela 5 Distribuição dos 22 haplótipos por raça ou grupo genético de suíno e por geografia para o fragmento de 363 pb da região D-loop do mtDNA. Os hífen indicam deleções e os pontos indicam ausência de polimorfismo.

Haplótipos (Hap)	*Região	*Raças suínas	Posição do sítio polimórfico (pb)																						
			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				
AJ002189			T	A	G	C	C	C	A	T	C	T	T	T	G	C	T	A	A	C	A	C	T	C	T
Hap_1 (n=3)	CO1; NE2	SBA1; SCT1; SPI1	C	.	A	-	T	T	G	C	.	.	.	.	.	-	.	.	G	.	G	T	.	.	.
Hap_2 (n=168)	NE18;CO146; SE1; S3	SBA3; SCB3; SCR4; SCTA1; SLD1; SME3; SMO2; SMT121_P; SMT5_DF; SNI9; SPI9; SPT3; SRP2; STA2	.	.	.	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	T	C	.	.
Hap_3 (n=6)	NE1; S5	SBA1; SLW 2; SMS3	C	.	A	-	T	T	G	C	.	.	.	C	.	-	.	.	G	.	.	T	.	T	.
Hap_4 (n=10)	CO1; S9	SCB1; SMO8; SMS1	.	.	.	.	.	.	.	.	T	.	.	.	.	-	C	.	.	.	.	.	.	.	.
Hap_5 (n=4)	S4	SDC3; SMS1	.	T	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	C	.	.
Hap_6 (n=3)	S2; NE1	SDC1; SMS1; SCB1	.	.	.	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	.	.	.	.
Hap_7 (n=23)	S9; CO1; NE13	SLD1; SMS6; SMT1_DF; SPN2; SCB13	.	T	.	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	.	C	.	.
Hap_8 (n=3)	S3	SLD3	.	T	.	.	.	.	C	.	C	.	.	A	-	C	.	.	.	.	.	.	.	.	.
Hap_9 (n=1)	S1	SLD1	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.
Hap_10 (n=5)	S5	SLD2; SMO1; SMS2	C	.	A	-	T	T	G	C	.	.	.	C	.	-	C	.	G	.	.	.	T	C	.
Hap_11 (n=1)	S1	SLW1	.	T	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	.	C	.	C	.
Hap_12 (n=4)	NE3; CO1	SME3; SPI1	C	.	A	-	T	T	G	C	.	.	.	.	-	C	.	G	T	.	T	.	T	.	.
Hap_13 (n=8)	CO3; NE5	SPI3; SCB5	.	T	.	.	.	.	C	.	.	.	.	.	-	C	.	.	.	.	.	C	.	.	.
Hap_14 (n=1)	CO1	SMT1_P	.	.	A	-	T	T	G	C	.	.	.	C	.	-	.	.	G	.	.	T	.	T	.
Hap_15 (n=70)	SE24; CO29; NE17	SCB70	C	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	T	C	.	.
Hap_16 (n=2)	SE2	SCB2	.	T	.	.	.	.	C	.	.	.	.	.	-	.	.	G	.	.	.	.	C	.	.
Hap_17 (n=1)	SE1	SCB1	C	.	A	.	T	T	G	C	.	.	.	.	-	.	.	G	.	.	T	.	.	.	.
Hap_18 (n=5)	SE5	SCB5	.	.	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	T	.	T	.	.
Hap_19 (n=3)	SE3	SCB3	C	.	A	.	T	T	G	C	.	.	C	.	-	.	.	G	.	.	T	.	T	.	.
Hap_20 (n=2)	NE2	SCB2	C	.	A	.	T	T	G	.	.	.	.	.	-	.	.	G	.	G	T	.	.	.	.
Hap_21 (n=5)	NE5	SCB5	C	.	A	.	T	T	G	C	.	.	.	.	-	.	.	G	.	G	T	.	.	.	.
Hap_22 (n=3)	NE3	SCB3	C	.	A	.	T	T	G	C	.	.	.	.	-	C	.	G	T	.	T	.	T	.	.

Raças localmente adaptadas: SBA=Baé; SCB=casco de burro; SRC=Caruncho; SCT=Canastra ; SME=mestiços; SMO=Moura; SMT\_P=Monteiro do Pantanal; SMT\_DF= Monteiro do DF; SNI=Nilo; SRP=Rabo de Peixe; SPI=Piau; STA=Tatu, Comerciais: SLD=Landrace; SDC=Duroc; SLW=Large White; SMS=MS60; SPN=Pietrain. Regiões: NE = nordeste; SE = sudeste; S = sul e CO = centro-oeste.\* número após siglas são referentes ao número de sequências analisadas para cada raça ou região.

A árvore construída demonstrou que existe alta relação entre os haplótipos pelos ramos curtos e valores de bootstrap observados e também agrupou os haplótipos pela origem asiática e europeia (Figura 3A do ANEXO A). Um aspecto importante notado nesta árvore foi a formação de um grupo composto por haplótipos (H4, H5, H6, H7, H8, H11) detectados na região sul (S), que é berço das raças comerciais amostradas nesse estudo. Observa-se ainda que os haplótipos exclusivos das raças locais de origem asiática (H1, H19, H20 e H16) foram mais relacionados geneticamente.

### **3.1.7 Estrutura genética e geográfica com mtDNA**

Considerando que existe certa diferenciação entre as populações, foi realizada uma estimativa indireta de valores de  $F_{ST}$  (Tabela 11A do ANEXO) para verificar se tal diferenciação pode ser devido ao isolamento genético causado pela ausência de fluxo gênico. Os resultados desta análise revelaram que a população de suínos Monteiros do Pantanal apresenta diferenças genéticas que a torna mais diferenciada das demais populações com  $F_{ST}$  variando entre 0,16288 (SMT\_DF) e 0,85847 (mestiços), embora não significativo  $P > 0,05$ , e possivelmente a distribuição geográfica destas populações contribuíram para maior e menor proximidade genética entre elas. Verifica-se que houve maior isolamento genético entre os Monteiros do Pantanal (SMT\_P) localizados na região Centro-Oeste e raças comerciais, cuja maioria foi proveniente da região sul ( $F_{ST} = 0,58517$ ;  $P < 0,05$ ). Quanto às regiões fisiográficas, observa-se que existe maior fluxo gênico entre as regiões nordeste (NE) e sudeste (SE) e entre NE e sul (S) com valores de  $F_{ST}$  muito baixos 0,08714 e 0,07985, respectivamente (Tabela 12A do ANEXO A).

## **3.2 Discussão**

As populações apresentaram um baixo índice de estruturação ( $F_{ST} < 0,03$ ) e comum para as populações suínas de localização geográfica próxima, como reportado por Silva et al. (2011). No entanto, o  $F_{ST}$  foi moderado (0,08) ao considerar os grupos pela disposição geográfica, indicando que existe diferença entre os suínos do Pantanal e Brasília

promovida pelo isolamento geográfico ou, possivelmente, pelo efeito de deriva genética. O  $F_{ST}$  tende a ser maior quando se trata de populações mais distantes fisicamente, por exemplo, as populações suínas brasileiras amostradas de várias regiões do Brasil apresentaram um  $F_{ST}$  igual a 14% (Sollero et al., 2008) e em suínos de diferentes continentes o  $F_{ST}$  foi de aproximadamente 27% (Laval et al., 2000; Kim et al., 2005), este último pode ser considerado o mais alto índice de diferenciação disponível na literatura para suínos. De acordo com Knackfuss et al. (2013), os porcos monteiros apresentam baixa diferenciação genética e a pequena distância genética em relação aos espécimes domésticos a partir de informações do citocromo b do DNA mitocondrial. Isso pode ser resultado do recente processo de asselvajamento dos porcos monteiros no Pantanal.

Os baixos valores obtidos para o índice de diferenciação tanto pela análise de diferenciação par a par quanto pela AMOVA pode ter sido resultante do alto fluxo gênico entre as populações do Pantanal que não foi impedido pelas distâncias geográficas e, ao mesmo tempo, demonstra a capacidade migratória destes animais entre os diferentes pontos geográficos. A maior diferenciação foi entre as populações localizadas próximas, inferindo que existe alguma barreira impedindo o fluxo de genes entre essas populações e que, talvez, essa barreira seja causada pelo isolamento dessa população em relação às demais pelos proprietários das localidades onde foram amostradas.

Com respeito aos resultados da região controle do mtDNA pode-se inferir que a maior diferenciação observada nesse estudo foi causada pela variabilidade genética das diferentes raças ou grupos genéticos analisados que compuseram o grupo de suínos locais e comerciais. Além disso, a distribuição geográfica dos indivíduos em cada grupo impossibilita o fluxo de genes entre eles e, embora, o valor de  $F_{ST}$  entre as regiões tenha sido o menor, deve-se considerar que na região do Centro-Oeste os Monteiros do Pantanal foram agrupados com outras raças diminuindo a diferença genética entre eles, não sendo, portanto, a melhor forma de compreender a estrutura genética e geográfica das populações deste estudo.

Essas informações condizem com os resultados do teste de Mantel que revelou que 24% da diferenciação genética, observada nas populações, podem ser explicadas pela hipótese de isolamento por distância, indicando que a diversidade genética nas populações não parece ser afetada pela distância geográfica. Isso implica dizer que uma população distante fisicamente não necessariamente será distante geneticamente, mas a divergência poderá surgir sob um processo de deriva genética equilibrada pelo fluxo gênico (Wright, 1943). Como é conhecido, o fluxo gênico desempenha um papel



fundamental na estrutura de populações, quando existem altos níveis de fluxo gênico entre as populações significa que elas podem evoluir juntas, caso contrário, a tendência é que cada população evolua quase independente (Slatkin, 1994).

Em todos os contrastes realizados pela AMOVA (Tabela 4) com ou sem inclusão do grupo externo (MOB) a maior variação ocorreu entre indivíduos dentro das populações (acima de 92%) como em geral ocorre para as populações de animais domésticos (Arranz et al., 2001; Jordana et al., 2003; Tapio et al., 2010).

A designação das populações em cada cluster condiz com as posições geográficas dessas populações (Brasília e Pantanal). No entanto, quando a estrutura geográfica foi analisada para o número de populações reais ( $K=4$ ), a população MOB permanece diferenciada e subestruturada com uma proporção de 51% dos genótipos compondo o cluster 1, mas 49% foram compartilhados com as demais populações do Pantanal. Esse resultado já era esperado uma vez que os indivíduos amostrados em Brasília são descendentes de animais provenientes do Pantanal. Embora a distribuição das populações em cada cluster (para  $K=4$ ) não tenha corroborado com as distâncias geográficas das populações do Pantanal, esteve de acordo com os valores das distâncias genéticas, por exemplo, as Pop1 e Pop4 que estão distantes por 93 km (Tabela 5A do ANEXO A), mas geneticamente próximas, apresentaram maior parte de seus genótipos no cluster 4, entretanto, também compartilharam esse cluster com a Pop6, a qual apresentou altos valores de distâncias genéticas quando comparadas com as Pop1. No geral, observa-se uma grande heterogeneidade nestas populações, exceto para a MOB.

#### 4. CONCLUSÕES

As populações de suínos Monteiros não estão isoladas geneticamente de acordo com informações de microssatélites e DNA mitocondrial devido à presença de fluxo gênico, que pode estar relacionado com a capacidade de dispersão dos animais na região em busca de alimentos e água. Ademais, os microssatélites detectaram fraca estruturação genética nas populações.

As análises de estrutura genética e geográfica realizadas com as sequências da região controle do mtDNA de suínos Monteiros e de outras populações suínas comerciais e locais demonstraram que essas populações encontram-se estruturadas, inferindo que numa análise intraespecífica o mtDNA pode ser mais eficiente para separar populações distantes geograficamente.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- ALEXANDER, L.J.; ROHRER, G.A.; BEATTIE, C.W. et al. Cloning and characterization of 414 polymorphic porcine microsatellites. **Animal Genetics**, n 27, v 3, p. 137-148, 1996a.
- ALEXANDER, L.J., TROYER, D.L., ROHRER, G.A. et al. Physical assignments of 68 porcine cosmid and lambda clones containing polymorphic microsatellites. **Mammalian Genome**, n 7, v 5, p. 368-372, 1996b.
- AMOS, W. The hidden value of missing genotypes. **Mol. Biol. Evol.** 23: 1995–1996, 2006.
- ARRANZ, J. J.; BAYON, Y.; SAN, P. F. Differentiation among Spanish sheep breeds using microsatellites. **Genetics Selection Evolution**. v. 33, p.529-542, 2001.
- AVISE, J. C. ARNOLD, J.; BALL, R. M. et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. **Annual Review of Ecology, Evolution and Systematics**, Palo Alto, v. 18, p. 489-522, 1987.
- AVISE, J. C.; BALL, R. M.; ARNOLD, J. Current versus historical population sizes in vertebrate species with high gene flow: A comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. **Mol. Biol. Evol.**, n. 5. pp.331- 344, 1988.
- BARKER, J.S.F.. Global protocol for determining genetics distance among domestic livestock breeds. **Applied Livest. Prod.** v.21, p. 501-508, 1994.
- BLOUIN, M.S. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. **Trends in Ecology & Evolution**, n.18, p.503–511, 2003.
- CAVALCANTE NETO, A. Origem do suíno casco-de-burro e sua relação genética com populações ibéricas e americanas. Jaboticabal, 2010. **Tese (doutorado)** – Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2010.
- EIZIRIK, E.; KIM, J.; MENNOTI-RAYMOND, M. et al. Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, mammalia, Felidae). **Molecular Ecology**. v. 10, p. 65-79. 2001.
- ELLEGREN, H. B.; CHOWDHARY, M.; JOHANSSON, L. et al. A primary linkage map of the porcine genome reveals a low rate of genetic recombination. **Genetics**, v. 137, p. 1089-1100, 1994.

- EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology**, v.14, p. 2611-2620, 2005.
- EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin (Versão 3.1). An Integrated Software Package for Population Genetics Data Analysis. Computational and Molecular Population Genetics Lab (CMPG). Institute of Zoology, University of Berne, Switzerland. 2006. 145p.
- FAO, Food and Agriculture Organization [2004]. Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans Measurement of Domestic Animal Diversity: Recommended Microsatellite Markers New Microsatellite Marker Sets Recommendations of Joint ISAG/ FAO Standing Committee. In: Bradley, D.G., Fries, R., Bumstead, N., Nicholas, F.W., Cothran, Ollivier E.G., Grawford, A.M. (Eds.), Food and Agriculture Organization of the United Nations. Online. Disponível em: <<http://dad.fao.org>>
- FRANKHAM, R.; BALLOU, J.D.; BRISCOE, D.A. **Introduction to Conservation Genetics**. University of Cambridge, Cambridge. 2002.
- FREDHOLM, M.; WINTWRO, A. K.; CHRISTENDEN, K. et al. Characterization of 24 porcine (DA-DC) N microsatellites: Genotyping of unrelated animals from breeds and linkage studies. **Mammalian Genome**, v. 4, pp. 187, 1993.
- GARCÍA, G.; VERGARA, J.; LOMBARDI, R. Genetic characterization and phylogeography of the wild boar *Sus scrofa* introduced into Uruguay. **Genetics and Molecular Biology**, v.34, n.2, pp329-337, 2011.
- GUTIÉRREZ, J.P., ROYO, L.J., ÁLVAREZ, I., GOYACHE, F. MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. **Journal of Heredity**, n.96, pp. 718-721, 2005.
- HERRERA, R. C. S. P., SOUZA, R.; HERRERA, H. M.; MAURO, R. de A. Hábitos alimentares do Porco Monteiro (*Sus scrofa*) no Pantanal da Nhecolândia, Mato Grosso do Sul. In: Simpósio sobre recursos naturais e sócio econômicos do Pantanal, 2; 1996, Corumbá. Manejo e conservação. **anais...**Brasília: Embrapa SPI, p.67, 1996.
- HUSON, D. H.; BRYANT, D. Application of Phylogenetic Networks in Evolutionary Studies, **Mol. Biol. Evol.**, v. 23, n.2, p.254-267, 2006.
- JORDANA, J.; ALEXANDRINO, P.; BEJA-PEREIRA, A. et al. Genetic structure of eighteen local south European beef cattle breeds by comparative F-statistics analysis. **Journal of Animal Breeding and Genetics**, v. 120, n. 2, p. 73-87, 2003.
- JOSHI, M.B.; ROUT, P.K.; MANDAL, A.K. et al. Phylogeography and origin of Indian domestic goats. **Mol. Biol. Evol.** 21, 454-462, 2004.
- KIM, T. H.; KIM, K. S.; CHOI, B. H. et al. Genetic structure of pig breeds from Korea and China using microsatellite loci. **Journal Animal Science**, v.83, 2255-2263, 2005.
- KLÜTSCH, C.F.C.; MANSEAU, M.; WILSON, P.J. Phylogeographical Analysis of mtDNA Data Indicates Postglacial Expansion from Multiple Glacial Refugia in Woodland Caribou (*Rangifer tarandus caribou*). **PLoS ONE**, v. 7, n.12, p.52661, 2012.

- KNACKFUSS, F. B.; PAES, R. C. S.; MENEZES, A. N. et al. Análise Filogenética entre Suínos Asselvajados no Pantanal e Comerciais (*Sus scrofa*, Artiodactyla: Suidae). *Bol. Soc. Bras. Mastozool.* v. 67, p 39-45, 2013.
- LARSON, G.; DOBNEY, K., ALBARELLA, U. et al. Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication. 2005, v. 307, n. 5715, p. 1618-1621, 2005.
- LAVAL, G.; IANNUCELLIA, N.; LEGAULT, C. et al. Genetic diversity of eleven European pig breeds. **Genetic Selection Evolution**, v.32, p. 187–203, 2000.
- LIBRADO, P.; ROZAS, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25, p.1451-1452, 2009.
- MANTEL, N. The detection of disease clustering and a generalized regression approach. **Cancer Research**, v.27, n.2, p. 209-220, 1967.
- MILLER, J. M.; HALLAGER, S.; MONFORT, S. L. et al. Phylogeographic analysis of nuclear and mtDNA supports subspecies designations in the ostrich (*Struthio camelus*). **Conserv Genet**, v.12, p.423–431, 2011.
- NEI, M. Genetic distance between populations. *Amer. Natur.*, n. 6, p. 283-292. 1972.
- PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v.155, pp. 945–959, 2000.
- RAYMOND M.; ROUSSET, F. Genepop (version 1.2), population genetics software for exact tests and ecumenicism. **Journal of Heredity**. v.86, p.248-249, 1995.
- ROBIC, A.; DALENS, M.; WOLOSZYN, N. et al. Isolation of 28 new porcine microsatellites revealing polymorphism. **Mammalian Genome**, n 5, v 9, pp. 580-583, 1994.
- ROHRER, G. A.; ALEXANDER, L. J.; KEELER, J. W. et al. A microsatellite linkage map of the porcine genome. **Genetics**, v. 136, pp. 231, 1994.
- ROHRER, G. A.; ALEXANDER, L. J.; HU, Z. et al. A comprehensive map of the porcine genome. **Genome Research**, v. 6, pp. 371, 1996.
- ROSENBERG, N.A. Distruct: a program for the graphical display of population structure. **Molecular Ecology Notes**. v. 4, p.137-138, 2004.
- SAITOU, N.; NEI, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, n.4, p.406-425, 1987.
- SCHNEIDER, S.; ROESSLI, D.; EXCOFFIER, L. Arlequin: a software for population genetics data analysis User manual ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva; Geneva: 2000.
- SILVA, E. C.; DUTRA JUNIOR, W.M.; IANELLA, P. Patterns of genetic diversity of local pig populations in the State of Pernambuco, Brazil, **R. Bras. Zootec.**, v.40, n.8, p.1691-1699, 2011.
- SLATKIN, M. Gene flow in natural populations. En: **Annual Review of Ecology and Systematics** v.16, p. 393-430, 1985.

- SLATKIN, M. Gene flow and population structure. In *Ecological Genetics* (L.A. Real, ed.), p. 3–17. **Princeton**, NJ: Princeton University Press.1994.
- SOLLERO, B. P.; PAIVA, S. R.; FARIA, D.A. et al. Genetic diversity of Brazilian pig breeds evidenced by microsatellite markers. **Livestock Science**, v. 13, p.1-8, 2009.
- SOUZA, C.A. Padrões de diversidade mitocondrial e nuclear em raças de suínos naturalizadas do Brasil. Universidade Católica de Brasília (**Tese de Doutorado, 141f**). 2011.
- TAPIO, M.; OZEROV, M., TAPIO, I. Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia. **BMC Genetics**, p.11-76, 2010.
- WEIR, B.S.; COCKERHAM, C.C. Estimating F-statistics for the analysis of the population structure. **Evolution**, v. 38, p. 1358–1370, 1984.
- WENINK, P.W.; BAKER, A.J.; TILANUS, M.G.J. Global mitochondrial DNA phylogeography of holarctic breeding dunlins (*Calidris alpina*). **Evolution**, v.50, p318-330, 1996.
- WRIGHT, S. Isolation by Distance. **Genetics** , v.28, n2, 1943.

### CAPÍTULO 3

## VARIAÇÃO GENÉTICA DO GENOMA MITOCONDRIAL EM SUÍNOS E JAVALIS

### RESUMO

A proposta deste estudo foi utilizar sequências de DNA completas (16.472 pb; n=108) e parciais do genoma mitocondrial (575 pb;n=342) em 25 raças suínas de vários países e 59 amostras de javalis europeus e asiáticos para analisar assinatura de seleção em genes codificadores de proteínas, além de comparar os padrões de diversidade genética e relação filogenética entre suínos domésticos e javalis. A diversidade nucleotídica ( $\% \pi$ ) para a população global, obtida com sequências da região controle ( $1,20 \pm 0,07$ ) foi aproximadamente duas vezes maior que a obtida com as sequências completas do mtDNA ( $0,76 \pm 0,06$ ). Para as sequências completas do mtDNA, os javalis asiáticos (ASWB) apresentaram maior valor de diversidade nucleotídica em relação as demais populações ( $1,02\% \pm 0,31$ ) e o mesmo resultado foi observado usando a região controle ( $1,75\% \pm 0,13$ ), enquanto nos javalis europeus (EUWB) este valor foi baixo usando mtDNA completo ( $0,08 \pm 0,01$ ) e muito alto usando a região controle ( $1,33 \pm 0,18$ ). Os suínos americanos (AME) apresentaram alta frequência de haplótipos europeus (78,6%), com alguns haplótipos asiáticos relacionados a essa população. A percentagem de introgressão de asiáticos nas raças europeias (EUD), como Pietrain, Large White, Hampshire e Duroc foi de 11,6%. Essa introgressão foi a responsável por um aumento considerável na variabilidade, que foi 0,63 e 0,11 com e sem haplótipos asiáticos, respectivamente. A diversidade nucleotídica para os 13 genes codificadores do mtDNA variou de 0,003 a 0,008 e foi mais alta para javalis asiáticos ASWB (0,011) e mais baixa para AME (0,006). Os genes ATPase6, NADH4L, NADH4 e NADH6 foram mais polimórficos em sítios não-sinônimos nos grupos ASWB, EUD e AME, enquanto os genes ATPase6 e NADH4 mostraram mais diferenças não-sinônimas nos cinco grupos, sugerindo a presença de seleção. O teste de McDonald–Kreitman realizado por gene e por grupo confirmou o padrão de não neutralidade em pelo menos seis genes (NADH1, NADH2, COX3, NADH3, NADH4 e NADH5). Os

resultados demonstraram que tanto os suínos domésticos quanto os javalis apresentaram alta variabilidade nas sequências completas e parciais do mtDNA. A análise filogenética também revelou que algumas raças/grupos genéticos AME, por exemplo, a raça brasileira Moura e a crioula do Peru, apresentam mais de um haplótipo de origem asiática e européia. Apesar de alguns genes apresentarem muitas mutações não sinônimas, a taxa de fixação dessas mutações foi reduzida possivelmente pela ausência de seleção positiva. Portanto, as estatísticas baseadas nas mutações não sinônimas versus sinônimas indicaram que a seleção direcional foi, como esperado, predominante nos genes do mtDNA de suínos.

**PALAVRAS-CHAVE:** genes codificadores, mutações não-sinônimas, suínos americanos, NADH5, McDonald–Kreitman



### CHAPTER 3

## GENETIC VARIATION OF MITOCHONDRIAL GENOME IN PIGS AND WILD BOARS

### ABSTRACT

Variability analyses were performed in 25 pig breeds from several countries and 59 samples of European and Asian wild boars. These analyzes were conducted with 108 sequences of the complete mitochondrial genome sequences. Furthermore, in order to generate information about the pattern of variability distribution in the coding and noncoding regions of proteins genetic analyzes of 342 samples of a fragment of the mtDNA control region were carried out. The nucleotide diversity was high for all populations using the complete mtDNA or control region. American pigs showed a high frequency of European haplotypes (78.6%), with some Asian haplotypes related to this population. The percentage of Asian introgression into European breeds such as Pietrain, Large White, Duroc and Hampshire was 11.6%. This introgression was responsible for a considerable increase in variability, which was 0.63 and 0.11 with and without Asian haplotypes, respectively. The estimated nucleotide diversity in 13 protein-coding genes of mtDNA varied among populations, and showed the highest value in European wild boars, followed by European and American pigs. The nucleotide diversity obtained for each gene ranged from 0.003 to COX1 and COX3 genes, and 0.008 to NADH4L. Furthermore, ATPase6 and NADH4 genes showed higher number of non-synonymous differences in these populations. The McDonald-Kreitman test performed by gene and group confirmed a non-neutrality pattern at least six genes (NADH1, NADH2, COX3, NADH3, and NADH4 NADH5). The results showed that both domestic pigs as the boars showed high variability in complete and partial sequences of mtDNA. Phylogenetic analysis also revealed that some AME breeds /genetic groups, for example, the Brazilian breed Moura and Creole from Peru presented more of one haplotype (Asian and European origin). Although some genes presented many non-synonymous mutations, the fixation rate of these mutations was possibly reduced by the absence of positive selection. Therefore, the statistics based on

non-synonymous versus synonymous mutations indicated that directional selection was, as expected, the predominant in the mtDNA genes from pigs.

**KEY-WORDS:** coding genes, nonsynonymous mutations, American pigs, NADH5, McDonald–Kreitman

## 1. INTRODUÇÃO

O DNA mitocondrial (mtDNA) é largamente utilizado nos estudos de diversidade e filogeografia por apresentar características únicas, tais como herança materna, uma taxa de evolução relativamente rápida e falta de recombinação (Brown et al., 1979; Berlin & Ellegren 2001; Savolainen et al., 2002). Além disso, o genoma mitocondrial é pequeno (de 16 a 18 kb em tamanho nos mamíferos) e contém apenas treze genes que codificam proteínas (NADH desidrogenase (ND1, 2, 3, 4, 4L, 5, e 6), citocromo b (CytB), citocromo oxidase c (COX1, 2, 3), e ATP6 e 8 sintase). O COX1 foi selecionado recentemente como uma ferramenta padrão para taxonomia molecular e identificação individual (Ratnasingham & Hebert, 2007), assim como o gene citocromo b pela sua moderada taxa evolucionária e claro padrão evolutivo (Astorga, 1998; Barraclough et al., 2001).

As mitocôndrias estão envolvidas na respiração celular e produção de ATP, qualquer variante funcional nos genes dessa organela pode comprometer o *fitness* dos indivíduos (Rand, 2001; Garcia-Matinez et al., 2011) e ainda pode afetar o processo evolutivo de populações naturais (Excoffier, 1990). Evidência de seleção no mtDNA pode ser identificada usando testes clássicos, como o teste de McDonald–Kreitman (MK, McDonald & Kreitman, 1991) que é realizado utilizando a proporção de substituições não-sinônimas fixadas.

Em suínos e em outras espécies, análises de mtDNA estão sendo realizadas baseadas, principalmente, na região não codificadora (região controle ou D-loop) ou do citocromo b do mtDNA (Souza et al., 2009; Luetkemeier et al., 2010; Wang et al., 2010; Ramayo et al., 2010; Yu et al., 2013). Poucos estudos analisaram o mtDNA completo (Yang et al., 2003; Wu et al., 2007; Fernandez et al., 2008). As análises de genomas

completos podem identificar forças evolutivas no mtDNA que podem ser diferentes entre javalis e animais domésticos, ou variarem nas raças e/ou populações.

Os suínos do mundo inteiro têm origem Asiática e Européia de acordo com os estudos realizados com mtDNA (Giuffra et al., 2000; Kijas & Andersson, 2001). Diferente das raças Européias, também conhecidas como raças internacionais e comerciais, as raças americanas possuem, em sua maioria, a origem materna desconhecida e, por essas raças e/ou ecótipos serem originadas de várias raças. Isso contribui para gerar um alto nível de diversidade não apenas fenotípica, mas também genética. Diversos trabalhos reportaram um alto nível de diversidade genética das raças suínas espalhadas pelo mundo, em termos de DNA mitocondrial e nuclear (Yang et al., 2003; Souza et al., 2009; Jin et al., 2012; Burgos et al., 2013; Li et al., 2013), porém existem poucas informações na literatura a respeito da diversidade genética existente nos genes codificadores do mtDNA em suínos (Yang et al., 2003), os quais são muito conhecidos por estarem envolvidos na capacidade de adaptação dos indivíduos aos diferentes habitats (Mishmar et al., 2003).

A proposta deste estudo foi utilizar sequências completas do mtDNA de suínos domésticos de vários países, incluindo três raças suínas brasileiras, e javalis para comparar o padrão de diversidade genética e para detectar assinatura de seleção na região codificadora do mtDNA. Além disso, a fim de gerar mais informações sobre o padrão de distribuição da variabilidade nas regiões codificadoras e não codificadoras de proteínas foram realizadas análises genéticas de um fragmento da região controle do mtDNA.

## 2. MATERIAL E MÉTODOS

### 2.1 Amostragem e sequenciamento

Os 342 suínos domésticos e javalis usados no presente estudo foram escolhidos como representantes de cinco grupos: domésticos europeus (EUD (n=91): Landrace (n=22), Large White (n=23), Duroc (n=20), Hampshire (n=4), Berkshire (n=02), Pietrain (n=06), Turopolje (n=02), Mangalica (n=03) e Ibéricas (n=09)), americanos (AME (n=153): Piau (n=06), Monteiro (n=08) e Moura (n=12) do Brasil; crioulos de Cuba (n=20), Peru (n=25), Argentina (n=28), Costa Rica (n=12), Colombia (n=11), Guatemala (n=07), Bolivia (n=10) e Mexico (n=09), Yucatan (n=04) e Ossabaw (n=1), domésticos asiáticos (ASD (n=38): Meishan (n=04), Xiang (n=2), Jiangquhai (n=1), Jeju (n=05), Dahe (n=1), Taiwanensis (n=05), Jinhua (n=04), Lantang (n=1), Banna Mini (n=01), Neijiang (n=02), Penzhou (n=03), Wujin (n=03), Tibetan (n=03), Ya'nan (n=02) e Wuzhishan (n=01), javalis asiáticos (ASWB, n=36) e javalis europeus (EUWB, n=23), mais um outgroup (*Sus cebifrons*-ERX149172). Das 342 amostras, 28 sequências do mtDNA completo e 125 da região controle desse genoma foram geradas nesse estudo e o restante foram obtidas dos bancos de dados públicos RSA (Sequence Read Archive) e GenBank. Além disso, foram sequenciadas a região controle de todos os animais previamente estudados por Burgos-Paz et al (2013). Os números de acesso do GenBank estão mostradas na Figura 1 para sequência de mtDNA completo e na Figura 4B do ANEXO B para as sequências da região controle. As sequências de mtDNA completo foram usadas nas análises da região controle.

O DNA genômico foi extraído de sangue total pelo método padrão fenol/clorofórmio. O mtDNA completo foi sequenciado no Centro Nacional de Analisis

Genômico (CNAG, [www.cnag.cat](http://www.cnag.cat)) usando a plataforma HiSeq2000 da Illumina e a preparação das bibliotecas foi de acordo com o protocolo de sequenciamento *paired-end* da Illumina. A região controle foi sequenciada usando o método de Sanger. Em resumo, PCR para região controle (aproximadamente 600 bp) foi conduzida usando 0.5 µM de cada primer (5'-CGCCATCAGCACCCAAAGCT-3' (Fernández et al., 2008) e 5'-GTAACCATTTGACTGAATAGCACCT-3' (Ottoni et al., 2012)). Os produtos de PCR foram sequenciados com o kit de sequenciamento Big Dye Terminator v.1.1 (*Applied Biosystems*) e a plataforma ABI 3730 Analyser (*Applied Biosystems*). As sequências geradas nesse estudo serão depositadas no GenBank.

Para obter as sequências consensos dos dados de Sequenciamento de Nova Geração desse estudo, os *reads* foram alinhados com a ferramenta *Burrows Wheeler Alignment* (BWA; Li & Durbin, 2009) seguido de 7 *mismatches*. A sequência *consenso* no formato FASTA do mtDNA foi obtida usando arquivos bam combinados com SAMtools v.0.1.18-sl61 (Li et al., 2009) mais *vcfutils.pl vcf2fq*. Todas as sequências, de mtDNA completo (n=108) e da região controle (n=342), foram alinhadas em arquivos separados, usando o software MUSCLE (Edgar, 2004). Após o alinhamento, um fragmento de 575 pb da região controle (de 58 to 630 do genoma referência) e 16.472 pb do mtDNA completo, sem a região de repetição em tandem 'TACACGTGCG', foram usados para análises genéticas.

Os SNPs de todos os 13 genes codificadores do mtDNA foram anotados com *Variant Effect Predictor* (VEP) a partir de um *script* em PERL disponível no Ensembl database v. 72. (Flicek et al., 2013) usando parâmetros padrão incluindo a opção SIFT (Sorting Intolerant From Tolerant) para predizer o efeito da substituição do aminoácido na função da proteína (Ng & Henikoff, 2003) dos SNPs não-sinônimos (nsSNPs).

## **2.2 Análises de genética de populações e teste para seleção**

O Software MEGA v6 (Tamura et al., 2013) foi utilizado para encontrar o melhor modelo para reconstruir as árvores filogenéticas usando análises Bayesianas. Para o mtDNA completo e a região controle, foi o modelo KHY + I + G. Este modelo e outros parâmetros foram incluídos no software Beast 1.7.5 (Drummond et al, 2012) para construir árvores consensos aplicando o método de Monte Carlo via Cadeias de Markov. As árvores

foram desenhadas usando o software FigureTree v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). O método Median-Joining foi usado para conduzir análises de network de 341 sequências da região controle (Bandelt et al., 1999) com o software Network v4.6.1.1 (Bandelt & Dress, 1992).

O programa DnaSP v5.10 (Librado & Rozas, 2009) foi usado para estimar a diversidade nucleotídica ( $\pi$ ) e haplotípica ( $hd$ ), assim como o teste de neutralidade D de Tajima (Tajima, 1989) para o mtDNA completo e a região controle. Análises básicas de polimorfismo e divergência em sítios sinônimos ( $\pi_s$  e  $K_s$ ) e não-sinônimos ( $\pi_a$  e  $K_a$ ), assim como suas razões ( $\omega=\pi_a/\pi_s$  e  $\omega_1=K_a/K_s$ ) foram conduzidas em cada população e em cada um dos 13 genes codificadores. Os testes D de Tajima (Tajima, 1989), McDonald–Kreitman-MKT (1991) e Índice de Neutralidade (NI, Rand, 1996) foram usados para detectar sinais de seleção. Distâncias genéticas com respectivos erros padrão foram calculadas utilizando o modelo 3-parâmetro de Tamura (Tamura, 1992) disponível no MEGA. Para garantir que as análises da população européia não fossem enviesadas pela introgressão asiática, essas análises também foram realizadas sem incluir aquelas amostras suínas européias com haplótipo asiático. Considerando que para algumas raças existem mais sequências disponíveis, nas análises de diversidade foram usadas apenas 94 sequências de mtDNA completo e 192 sequências da região controle para balancear o tamanho da amostragem das populações. Essas sequências foram selecionadas aleatoriamente das 342 sequências amostradas nesse estudo.

### 3. RESULTADOS E DISCUSSÃO

#### 3.1 Resultados

##### 3.1.1 Relação filogenética usando mtDNA completo e região controle

As distâncias genéticas ( $d \pm$  erro padrão) entre os pares de grupos obtidas para o mtDNA completo e a região controle (Tabela 13B do ANEXO B) foram maiores entre javalis asiáticos (ASWB) e os suínos domésticos europeus sem haplótipos asiáticos (EUD1) ( $0,015 \pm 0,001$ , mtDNA completo;  $0,026 \pm 0,005$ , região controle) e entre ASWB e javalis europeus (EUWB,  $0,026 \pm 0,005$ , ambos para mtDNA completo e região controle). Os suínos americanos (AME) foram mais próximos dos EUWB ( $0,003$  para mtDNA completo;  $0,007 \pm 0,002$  para região controle) e dos EUD1 ( $0,004$  para mtDNA completo;  $0,006 \pm 0,002$  para região controle) que os domésticos asiáticos (ASD) e ASWB. Isso confirma a origem maternal dos haplótipos presentes na constituição genética de suínos americanos e corrobora com os resultados de estudos anteriores (Souza et al., 2009; Burgos-Paz et al., 2013).

A árvore Bayesiana construída a partir de 108 sequências de mtDNA completo (Figura 1) gerou dois clusters: um com haplótipos asiáticos e outro com haplótipos europeus. Dos 14 suínos americanos, 11 foram agrupados com EUD e EUWB, e as amostras restantes agruparam-se com as populações de suínos ASD e ASWB. Oito suínos domésticos europeus apresentaram haplótipos asiáticos: Large White ( $n = 3$ ), Hampshire ( $n = 1$ ), Pietrain ( $n=2$ ), Berkshire ( $n=1$ ) e Duroc ( $n=1$ ) e a linhagem Ibérica ‘Manchado de Jabugo’ que é conhecida ter sido cruzada, por volta de um século atrás, com



suínos britânicos (García et al., 1990). Sequências de suínos Yucatan foram muito similares às sequências de Large-White e um crioulo de Cuba (CRCU-1148) que se agruparam em um subgrupo com suínos Mangalica e Hampshire. Os outros nove suínos americanos do clado europeu agruparam-se em um subgrupo separado com suínos ibéricos e alguns EUWB, tais como um javali (WB) da França (WBF\_ERXI49180) e outro da Tunísia (WBTN\_0965).

No clado asiático, três WB do Norte da china (WBCN), um suíno doméstico asiático, um WB da Coreia (WBK), e um Berkshire foram geneticamente mais distantes dos outros suínos. Três suínos americanos com haplótipos asiáticos foram agrupados juntos e permanecendo próximos dos suínos Meishan, Lantang e Xiang.

A análise filogenética com 342 sequências da região controle mostraram que os suínos AME permaneceram no Clado Europeu, separando haplótipos asiáticos e o *Sus cebifrons*. A variabilidade observada na população asiática da região controle foi suficiente para separar os suínos domésticos asiáticos e ASWB com dois crioulos de Cuba entre eles (Figura 5B do ANEXO B).

A Análise de Network de 341 sequências da região controle forneceu informações adicionais sobre a relação entre as populações suínas (Figure 6B do ANEXO B). Três haplótipos mais frequentes relacionados com os suínos AME foram encontrados : H1 (36 indivíduos), H3 (43 indivíduos) e H7 (92 indivíduos). Foi observada uma única cópia de 32 haplótipos. Muitos destes haplótipos, em baixa frequência, estiveram localizados próximos a um haplótipo de maior frequência (H7) e, por isso, apresentaram um formato *star-like* que é consistente com uma expansão que pode ter ocorrido recentemente.

### 3.1.2 Diversidade genética usando mtDNA complete e região controle

Os 575-pb da região controle geraram 55 sítios polimórficos e 54 haplótipos, enquanto os 16.472-pb do mtDNA completo (excluindo a região de repetição em tandem), sem *gaps* ou dados faltantes, revelaram 856 sítios polimórficos e 78 haplótipos (H).

A diversidade haplotípica foi considerada alta para mtDNA completo, variando de  $0,96 \pm 0,05$  (em ASWB) para  $0,99 \pm 0,01$  (na população europeia com e sem

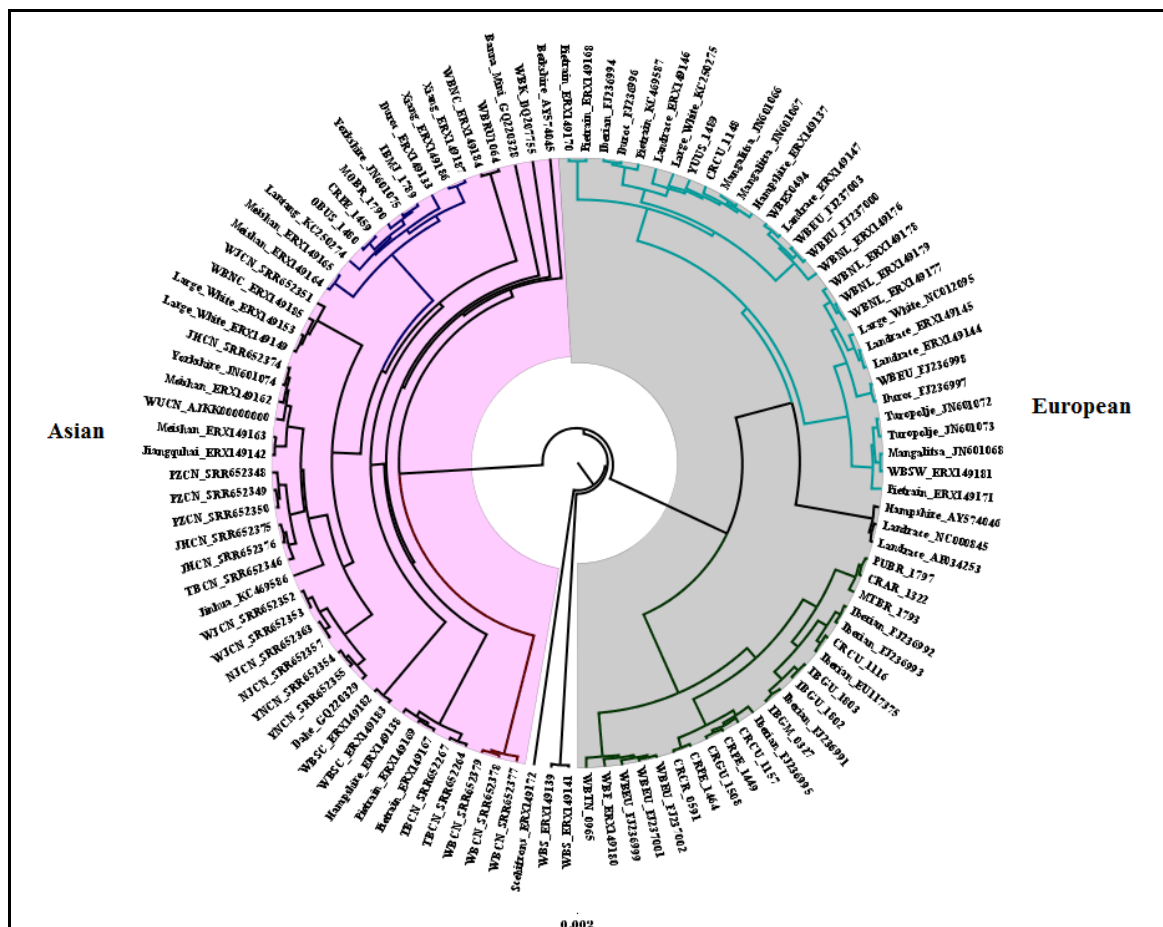


Figura 1 Árvore consenso obtida pela análise Bayesiana a partir de 108 seqüências de mtDNA completo. As cores rosa e cinza indicam os suínos de origem asiática e europeia, respectivamente. Dentro do clado europeu, os ramos de cor azul claro indicam amostras europeias, os ramos de cor verde as amostras americanas. No clado asiático os suínos americanos com haplótipos asiáticos estão destacados pela cor azul e a cor vermelha destaca os javalis do norte da China.

haplótipos asiáticos), e para a região controle esse parâmetro variou de  $0,64 \pm 0,01$  (europeus sem haplótipos asiáticos-EUD1) a  $0,98 \pm 0,01$  (em ASWB).

A diversidade nucleotídica ( $\% \pi$ ) para a população global, obtida com seqüências da região controle ( $1,20\% \pm 0,07$ ), foi aproximadamente duas vezes maior que a obtida com as seqüências do mtDNA completo ( $0,76\% \pm 0,06$ ). Para as seqüências de mtDNA completo, ASWB apresentaram maior valor de diversidade nucleotídica em relação as demais populações ( $1,02\% \pm 0,31$ ) e o mesmo resultado foi observado usando a região controle ( $1,75\% \pm 0,13$ ), enquanto na EUWB este valor foi baixo usando mtDNA

completo ( $0,08\% \pm 0,01$ ) e muito alto usando a região controle ( $1,33\% \pm 0,18$ ) (Tabela 1). De acordo com a literatura a região controle é mais variável que outras regiões do genoma mitocondrial completo (Cann et al., 1984). A população AME mostrou variabilidade nucleotídica mais alta para o mtDNA completo ( $0,49\% \pm 0,1$ ) e para a região controle ( $0,52\% \pm 0,10$ ) em relação à população europeia sem haplótipos asiáticos (EUD1). A introgressão de haplótipos asiáticos observados na população EUD aumentou a quantidade de diversidade nucleotídica em quase 5 vezes quando foram utilizadas as sequências de mtDNA completo ( $\% \pi = 0,63 \pm 0,08$  em EUD diminuiu para  $0,11\% \pm 0,02$  em EUD1) e a região controle ( $\% \pi = 0,83 \pm 0,16$  em EUD diminuiu para  $0,36\% \pm 0,16$  em EUD1).

A estatística D de Tajima foi negativa e significativa na população doméstica europeia sem haplótipos asiáticos ( $-2,019$  e  $-2,009$  respectivamente,  $P < 0,05$ ), como observado em outros estudos comumente proveniente de expansão populacional recente. Um valor significativo também foi observado em suínos domésticos asiáticos ( $-1,93$ ,  $P < 0,05$ ) usando mtDNA completo (Tabela 1). Para os suínos americanos, D de Tajima foi positivo proveniente da mistura natural que ocorre nessas populações (Burgos-Paz et al., 2013).

### 3.1.3 Variação genética em genes codificadores e sinais de seleção

A diversidade nucleotídica ( $\pm$ SD) para todos os 13 genes variam entre populações (Figura 2) e foi alta em ASWB ( $0,011 \pm 0,003$ ), EUD ( $0,008 \pm 0,002$ ) e AME ( $0,006 \pm 0,001$ ). Dois genes apresentaram valores superiores a média em EUD (ATPase8 e NADH4L), um em ASD (COX3), um em EUWB (ATPase6) e dois em ASWB (NADH4L e COX3) (Figura 2A). Os genes COX3 ( $0,003 \pm 0,002$ ) e COX1 ( $0,003 \pm 0,003$ ) mostraram menor variabilidade na população global que os genes NADH4L ( $0,008 \pm 0,006$ ) e ATPase8 ( $0,006 \pm 0,005$ ) (Figura 2B). A ferramenta *Variant Effect Predictor* (VEP) foi utilizado para anotação de cada variante de cada gene codificador do mtDNA (Tabela 14B do ANEXO B). A proporção de substituições sinônimas e não-sinônimas foi aproximadamente 3 para 1 para todos os 13 genes codificadores (número de polimorfismos sinônimos e não-sinônimos foi 497 e 183, respectivamente). ATPase8 mostrou mais mutações não-sinônimas, enquanto no NADH6 foram observadas poucas mutações desse tipo. Sessenta e dois dos polimorfismos não-sinônimos foram preditos como deletérios (SIFT scores

Tabela 1 Parâmetros de diversidade nucleotídica e teste de neutralidade (D de Tajima) por população suína usando sequências de mtDNA completo e região controle.

<sup>a</sup> População	mtDNA					Região controle				
	N	H	% $\pi$ ( $\pm$ SD)	Hd ( $\pm$ SD)	Tajima's D	N	H	% $\pi$ ( $\pm$ SD)	Hd ( $\pm$ SD)	D de Tajima
AME	14	11	0,49 $\pm$ 0,1	0,96 $\pm$ 0,04	0,344	70	13	0,52 $\pm$ 0,10	0,84 $\pm$ 0,03	-1,326
EUD	28	25	0,63 $\pm$ 0,08	0,99 $\pm$ 0,01	0,097	34	11	0,83 $\pm$ 0,16	0,68 $\pm$ 0,09	-0,127
EUD1	20	18	0,11 $\pm$ 0,02	0,99 $\pm$ 0,02	-2,019*	27	8	0,36 $\pm$ 0,16	0,64 $\pm$ 0,01	-2,009*
ASD	27	21	0,10 $\pm$ 0,01	0,98 $\pm$ 0,01	-1,933*	29	14	1,35 $\pm$ 0,08	0,91 $\pm$ 0,03	0,567
EUWB	14	12	0,08 $\pm$ 0,01	0,97 $\pm$ 0,04	-1,125	34	17	1,33 $\pm$ 0,18	0,93 $\pm$ 0,03	0,001
ASWB	11	9	1,02 $\pm$ 0,31	0,96 $\pm$ 0,05	-0,023	32	23	1,75 $\pm$ 0,13	0,98 $\pm$ 0,01	-0,624
Todos	94	78	0,76 $\pm$ 0,06	0,99 $\pm$ 0,00	-	192	54	1,20 $\pm$ 0,07	0,91 $\pm$ 0,01	-

N=número de sequências; H = número de haplótipos;  $\pi$  = diversidade nucleotídica; Hd= diversidade haplotípica, EUD1 = europeus sem haplótipos asiáticos, e significância estatística: \* P < 0,05.

<sup>a</sup>Americanos (AME), Domésticos europeus (EUD), Domésticos europeus sem haplótipos asiáticos (EUD1) e javalis europeus e asiáticos (EUWB e ASWB).

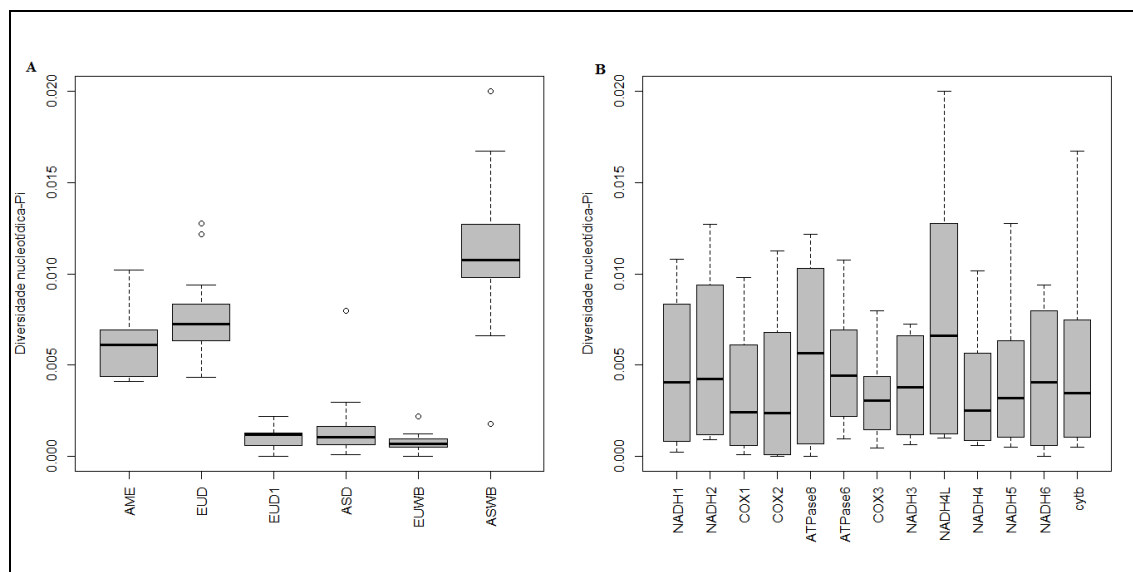


Figura 2 Diversidade nucleotídica por população (A) para todos os genes codificadores do mtDNA por genes e em todas as cinco populações suína (B).

<0.05); o gene COX1 apresentou 9 dessas mutações nos haplótipos europeus ou europeus com haplótipos asiáticos. No NADH6 nenhuma dessas mutações deletérias foi encontrada (Figura 6B do ANEXO B).

Nove genes codificadores mostraram maior diversidade nucleotídica em sítios sinônimos ( $\pi_s$ ) que em não-sinônimos ( $\pi_a$ ) (Figura 3A e B). No entanto, os genes ATPase6, NADH4L, NADH4 e NADH6 mostraram variabilidade mais alta em não-sinônimos que os genes referidos anteriormente, principalmente nas populações ASWB, EUD e AME, resultando em uma razão ( $\omega = \pi_a/\pi_s$ ) maior que 1 (Figura 3C). Também foram estimadas divergências nucleotídica sinônimas ( $K_s$ ) e não-sinônimas ( $K_a$ ) para cada gene (Figuras 3D e E). ATPase6 e NADH4 mostraram um excesso de diferenças não-sinônimas em todas as seis populações, com razões  $K_a/K_s$  maiores que 1 ( $\omega > 1$ ; Figura 3F).

O teste de McDonald–Kreitman (MKT) detectou seleção positiva ou negativa em seis genes codificadores (NADH1, NADH2, COX3, NADH3, NADH4 e NADH5) nas cinco populações (AME, EUD, EUD1, EUWB e ASWB). Cinco e três dos genes mostraram valores significativos para MKT ( $P < 0.01$  ou  $P < 0.001$ , teste exato de Fisher, bicaudal) nas populações EUD e europeus sem haplótipos asiáticos, respectivamente (Tabela 2). Os resultados significativos de MKT para os genes NADH2, NADH3 e NADH5 foram devido ao excesso de polimorfismo intraespecífico como

mostrados pelos valores do Índice de Neutralidade (NI) que foram maiores que 1 (Tabela 2). Ao contrário, o valor de NI para NADH4 foi menor que 1 (variando de 0,04 a 0,07), sugerindo que vários polimorfismos intraespecíficos foram fixados durante o processo de divergência.

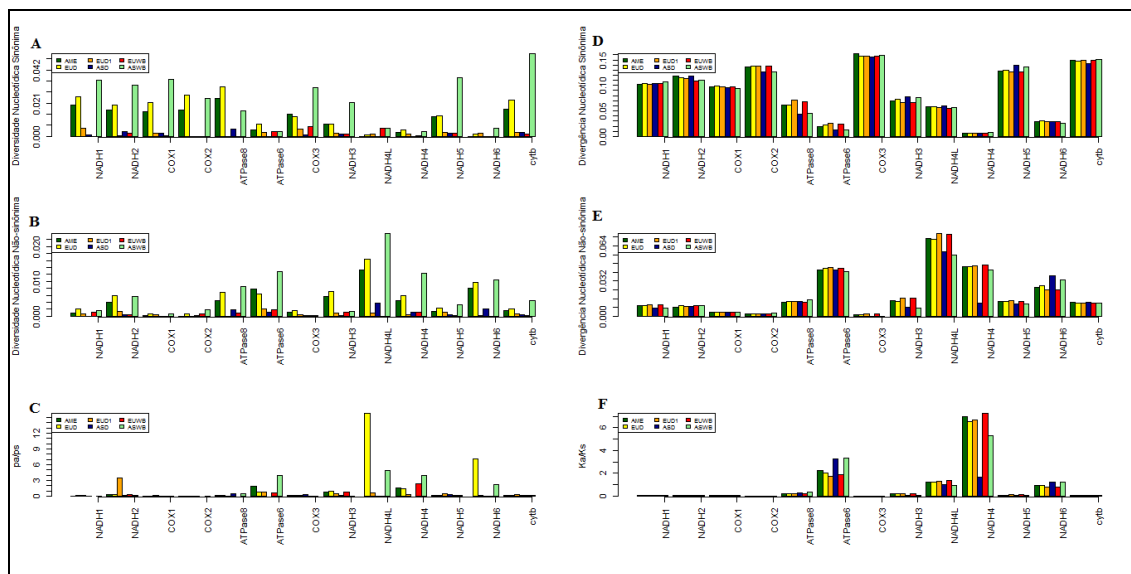


Figura 3 Diversidade nucleotídica sinônima e não-sinônima (A e D), divergência (B e E) e razões (C e F) para cada população suína e cada gene do mtDNA.

### 1.3 Discussão

O DNA mitocondrial é um marcador largamente empregado nos estudos populacionais e foi escolhido nesse estudo por desempenhar uma função importante na etapa de desvendar a origem materna dos indivíduos e analisar a diversidade genética materna nas populações (Giles et al., 1980; Luikart et al., 2001; Fang & Andersson 2006; de Jong et al., 2011).

A divergência materna entre as populações europeias e as asiáticas é um fato esperado, considerando que essas populações divergiram há centenas de anos atrás (Groenen et al., 2012). Ademais, os suínos domésticos divergiram maternalmente dos javalis e, essa divergência foi mais pronunciada entre as raças domésticas e javalis de origem asiática, como um indicativo de isolamento genético ou falta de fluxo gênico entre elas.

**Tabela 2** Teste de McDonald–Kreitman (MKT) e Índice de Neutralidade (NI) por população suína e por cada gene do mtDNA. Os P-values foram obtidos pelo teste exato de Fisher.

Genes	AME		EUD		EUD1		ASD		EUWB		ASWB	
	P-value	NI	P-value	NI	P-value	NI	P-value	NI	P-value	NI	P-value	NI
NADH1	0,679	0,507	0,532	1,800	0,143	2,976	1,000	0,000	0,015*	n.a	n.a	1,206
NADH2	0,018*	6,000	0,002**	8,462	0,0003***	37,800	0,217	2,417	0,244	3,714	0,037*	8,500
COX1	0,471	0,456	0,148	2,750	0,010	4,625	0,940	1,094	1,000	0,000	0,174	0,238
COX2	1,000	0,000	0,323	3,923	n.a	n.a	0,083	n.a	0,077	n.a	1,000	1,650
ATPase8	1,000	0,667	1,000	1,333	n.a	n.a	1,000	1,000	1,000	1,333	1,000	2,500
ATPase6	0,511	1,625	0,327	2,321	0,239	3,333	0,352	2,667	1,000	0,556	1,000	0,800
COX3	0,103	n.a	0,050*	n.a	n.a	11,600	0,491	n.a	0,559	7,250	1,000	n.a
NADH3	0,070	12,000	0,001***	14,000	0,086	n.a	n.a	3,500	0,269	3,000	0,442	0,667
NADH4L	0,140	n.a	0,604	2,727	0,338	0,176	0,545	n.a	0,190	0,000	0,072	10,200
NADH4	0,075	0,120	0,008**	0,073	0,002**	0,042	1,000	n.a	0,293	0,208	1,000	0,889
NADH5	0,773	1,310	0,004**	3,619	0,003**	4,654	0,065	3,796	0,362	2,417	0,647	1,307
NADH6	0,214	n.a	0,317	4,125	0,476	0,300	0,558	n.a	n.a	n.a	1,000	0,929
Cytb	0,735	1,267	n.a	n.a	n.a	n.a	0,193	3,265	1,000	1,630	1,000	1,037
Total	0,028*	1,559	<0,000***	2,296	<0,000***	2,988	0,0003**	2,459	0,0203*	2,250	0,378	1,186

n.a=não analisado; \* 0,01<P<0,05; \*\* 0,001<P<0,01 e \*\*\* P<0,001.

A população suína americana apresentou haplótipos europeus usando o genoma mitocondrial completo ou mesmo um pequeno fragmento da região controle desse genoma. Estudos anteriores já reportaram que as raças suínas da América do Sul pertencem à haplótipos europeus, principalmente, por terem origem materna ibérica (Ramírez et al., 2009; Souza et al., 2009). A raça Moura, uma raça naturalizada brasileira, que tem origem obscura, apresentou origem asiática empregando mtDNA completo e a região controle, divergindo das raças Piau e Monteiro de origem Européia. Essa raça também foi analisada com o citocromo b por Souza et al. (2009) e demonstrou um perfil haplotípico diferente ao observado em outras raças suínas brasileiras e muito próximo da variedade ibérica Black Hair e da raça Mangalica da Hungria por compartilharem haplótipos em alta frequência.

A árvore genética gerada pela análise Bayesiana nesse estudo distribuiu a raça Moura em um clado asiático composto por raças européias, como Duroc e Large White e raças chinesas, como Meishan e Xiang e muito próxima das raças Ossabaw e uma crioula do Peru, enquanto, os dois suínos amostrados da raça Mangalica foram agrupados no clado de haplótipos europeus.

A diversidade nucleotídica foi alta para todas as populações empregando a região controle ou mtDNA completo. Entretanto, observou-se que a diversidade nucleotídica foi menor que a diversidade haplotípica em quase todas as populações, exceto ASWB, quando se empregou o mtDNA completo, e em três populações (AME, EUD e EUD1) usando informações da região controle. Esses resultados sugerem que essas populações, possivelmente, possuíam um tamanho efetivo pequeno e que passaram por um processo de expansão demográfica (Avice, 2000). Além disso, o teste de neutralidade D de Tajima (Tajima, 1989) (Tabela 15B do ANEXO B) indicou uma expansão demográfica pelos valores negativos e significativos observados, assim como a forma da rede network apresentou o grupo europeu com muitos haplótipos em baixa frequência também retratando esse tipo evento.

As raças chinesas, que nesse trabalho representaram a população asiática, são conhecidas por apresentarem pouca variação na região controle do mtDNA quando comparada à população suína Européia (Okumura et al., 2001). No entanto, essa informação não corrobora com o nível de diversidade nucleotídica encontrado nas regiões codificadora e não codificadora dos suínos asiáticos deste estudo. Contudo, deve-se considerar que as raças suínas da China estão classificadas em diferentes grupos de acordo com a distribuição geográfica, características morfológicas e fisiológicas dessas raças



podendo resultar em diferentes variações genéticas dependendo do grupo amostrado (Zang et al., 1986).

As raças suínas parecem sofrer com a pressão de seleção como foi observado no nível de diversidade nucleotídica nas regiões codificadoras do mtDNA, que foi menor na população de raças selecionadas EUD1, enquanto, na população AME que representa as raças não selecionadas, comparada com as demais populações, o nível de diversidade nucleotídica foi bastante elevado. De fato, alguns genes do mtDNA, que desempenham importantes funções na cadeia respiratória, apresentaram um padrão não neutro de acordo com os resultados do teste de McDonald–Kreitman. Mutações que mudam a sequência de aminoácido de uma proteína geralmente sofrem fraca pressão de seleção positiva, que reduz a taxa de fixação dessas mutações (Kimura, 1983). Isso pode justificar aqueles genes ATPase6, NADH4, NADH4L e NADH6 que apresentaram alta variabilidade nos sítios não-sinônimos, porém não foi significativo pelo teste MK. Em comparações entre populações européias e chinesas o NADH6 não apresentou polimorfismos não sinônimos (Kijas & Anderson, 2001), enquanto, Yang et al. (2003) verificaram em populações chinesas um excesso de substituições não-sinônimas nos genes ATPase6, ATPase8 e NADH3.

#### 4. CONCLUSÕES

Neste estudo foi possível observar que tanto os suínos domésticos quanto os javalis apresentaram alta variabilidade nas sequências completas e parciais do mtDNA. A análise filogenética, além de detectar sinais de introgressão genética de haplótipos asiáticos em suínos europeus, que aumenta a diversidade genética desses últimos, também revelou que algumas raças/grupos genéticos AME, por exemplo, a raça brasileira Moura e a raça crioula do Peru, apresentam mais de um haplótipo de origem asiática e europeia. Além disso, apesar de alguns genes do mtDNA apresentarem muitas mutações não sinônimas, a taxa de fixação dessas mutações foi reduzida, possivelmente pela ausência de seleção positiva e, por isso, não foram detectados pelo teste de McDonald–Kreitman. Portanto, as estatísticas baseadas nas mutações não sinônimas versus sinônimas indicaram que a seleção direcional foi, como esperado, predominante nos genes (NADH1, NADH2, COX3, NADH3, NADH4 e NADH5) do mtDNA de suínos.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- ASTORGA, M.; GALLEGUILLOS, R. Genetic divergence of jack mackerel of genus *Trachurus* from northwestern and southeastern Pacific. **Review Biology Marine Oceanography**. n. 33, p.155-61, 1998.
- BANDELT, H-J, DRESS, A.W.M. Split Decomposition: A new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution*, n. 1, p.242-252, 1992.
- BANDELT, H-J.; FORSTER, P.; RÖHL, A. Median-joining networks for inferring intraspecific phylogenies. **Mol. Biol. Evol.** n.16, p.37-48, 1999.
- BARRACLOUGH, T.G.; NEE, S. Phylogenetic and speciation. **Trends in Ecology and Evolution**. n.16, p.391-99, 2001.
- BERLIN, S., ELLEGREN, H. Evolutionary genetics. Clonal inheritance of avian mitochondrial DNA. **Nature**. n. 413, p.37-8, 2001.
- BROWN, W.M.; GEORGE, M. Jr.; WILSON A.C. Rapid evolution of animal mitochondrial DNA. **Proceedings of the National Academy of Sciences USA**. n. 76, p.1967-71, 1979.
- BURGOS-PAZ, W.; SOUZA, C. A.; MEGENS, H. J. Porcine colonization of the Americas: a 60k SNP story. **Heredity**. n. 110, p. 321-30, 2013.
- DE JONG, M.A.; WAHLBERG, N.; VAN EIJK, M. et al. Mitochondrial DNA Signature for Range-Wide Populations of *Bicyclus anynana* Suggests a Rapid Expansion from Recent Refugia. **PLoS ONE**. v. 6, n.6, p.21385, 2011.
- DRUMMOND, A.J.; SUCHARD, M.A.; XIE, D.; RAMBAUT, A. Bayesian phylogenetics with beauti and the beast 1.7. **Molecular Biology Evolution**. n. 29, p.1969-73, 2012.
- EDGAR, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**. n. 32, p. 1792-97, 2004.
- EXCOFFIER, L. Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. **J. Mol. Evol.** n. 30, p. 125-139, 1990.

- FANG, M.; ANDERSSON, L. Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. **Proc. R. Soc. Lond. Biol.** n. 273, p.1803-1810, 2006.
- FERNÁNDEZ, A. I.; ALVES, E.; FERNÁNDEZ, A. et al. Mitochondrial genome polymorphisms associated with longissimus muscle composition in Iberian pigs. **Journal of Animal Science.** n. 86, p.1283-1290, 2008.
- FLICEK, P.; AMODE, M.R; BARRELL, D. ENSEMBL 2013. **Nucleic Acids Research.** Jan;41(Database issue):D48-55, n.30, 2013.
- GARCÍA, D. M; MARTÍNEZ, S. O. F.: **Guía de campo de las razas autóctonas españolas.** Madrid: Alianza Editorial; 1990.
- CHRISTIE, J. S.; PICORNELL, A.; MOYA, A. et al. Mitochondrial DNA effects on fitness in *Drosophila subobscura*. **Heredity**, n.107, p.239–245, 2011.
- GILES, R. E.; BLANC, H.; CANN, H. M.; WALLACE, D. C. Maternal inheritance of human mitochondrial DNA. **Proc. Natl. Acad. Sci. USA.** n. 77, p. 6715-6719, 1980.
- GIUFFRA, E.; KIJAS, J.M.H.; AMARGER, V. et al. The origin of the domestic pig: independent domestication and subsequent introgression. **Genetics.** n.154, p.1785–1791, 2000.
- GROENEN, M. A.; ARCHIBALD, A. L. et al. Analyses of pig genomes provide insight into porcine demography and evolution. **Nature**, v.491, n.7424, p. 393-398, 2012.
- JIN, L., ZHANG, M., MA, J. et al. Mitochondrial DNA Evidence Indicates the Local Origin of Domestic Pigs in the Upstream Region of the Yangtze River. **PLoS ONE.** v. 7, n.12, p.51649, 2012.
- KIJAS, J.M.H.; ANDERSON, L. A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. **J Mol Evol**, n.52, p.302-308, 2001.
- KIMURA, M. The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University Press; 1983.
- LI, H.; DURBIN, R. Fast and accurate short read alignment with BurrowsWheeler Transform. **Bioinformatics.** v.25, p.175460, 2009.
- LI, H.; HANDSAKER, B.; WYSOKER, A.; DURBIN, R.; 1000 Genome Project Data Processing Subgroup The Sequence alignment/map (SAM) format and SAMtools. **Bioinformatics.** n. 25, p.207-89, 2009.
- LI, R.; ZOU, H.; JIA, Y.; ZHAO, R. Glucocorticoid receptor is involved in the breed-dependent transcriptional regulation of mtDNA- and nuclear-encoded mitochondria genes in the liver of newborn piglets. **BMC Veterinary Research**, n.9, p.87, 2013.
- LIBRADO, P.; ROZAS, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. **Bioinformatics.** n.25, p. 1451–52, 2009.

- LUETKEMEIER, E.S.; SODHI, M.; SCHOOK, L. B.; MALHI, R.S. Multiple Asian pig origins revealed through genomic analyses. **Molecular phylogenetics and evolution**. 54,680-6, 2010.
- LUIKART, G.; GIELLY, L.; Excoffier, L. et al. Multiple maternal origins and weak phylogeographic structure in domestic goats. **Proc Natl Acad Sci USA** 98: 5927–5932, 2001.
- McDONALD, J.H.; KREITMAN, M. Adaptive protein evolution at the Adh locus in *Drosophila*. **Nature**. n. 351, p.652–54, 1991.
- MISHMAR, D.; RUIZ-PESINI, E.; GOLIK, P. et al. Natural selection shaped regional mtDNA variation in humans. **Proc. Natl. Acad. Sci.** n. 100, p.171–176. 2003.
- NG, P.C.; HENIKOFF, S. SIFT: predicting amino acid changes that affect protein function. **Nucleic Acids Research**. n. 31, p.3812, 2003.
- OKUMURA, N.; KUROSAWA, Y.; KOBAYASHI, E. et al. Genetic relationship amongst the major non-coding regions of mitochondrial DNAs in wild boars and several breeds of domesticated pigs. **Anim. Genet.** n. 32, 139–147, 2001.
- OTTONI, C.; FLINK, L. G.; EVIN, A. et al. Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. **Mol Biol Evol** 30(4):824–832, 2013.
- RAMAYO, Y.; SHEMERET'EVA, I. N.; PEREZ-ENCISO, M. Mitochondrial DNA diversity in wild boar from the Primorsky Krai Region (East Russia). **Animal Genetics**. 42, 96–9, 2010.
- RAND, D. M. neutrality tests of molecular markers and connection biology. **Conserv. Biol.** n. 10, p.665-671, 1996.
- RAND, D.M. The units of selection on mitochondrial DNA. **Annu. Rev. Ecol. Syst.** n. 32, p. 415–448, 2001.
- RATNASINGHAM, S.; HEBERT, P.D. bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). **Molecular Ecology Notes**. n.7, v. 355–364, 2007.
- RAMIREZ, O.; OJEDA, A., TOMAS, A. et al. Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. **Mol Biol Evol.**, n.26, p.2061–2072, 2009.
- SAVOLAINEN, P.; ZHANG, Y.P.; LUO, J. et al. Genetic evidence for an East Asian origin of domestic dogs. **Science**. n. 298, p.1610–13, 2002.
- SOUZA, C. A.; PAIVA, S. R.; PEREIRA, R. W. et al. Iberian origin of Brazilian local pig breeds based on Cytochrome b (MT-CYB) sequence Genetic diversity and structure in Asian native goat analyzed by newly developed SNP markers. **Animal Genetics**. 40, 759-62, 2009.

- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis. **Genetics**. n.123, p. 585–595, 1989.
- TAMURA, K. Estimation of the number of nucleotide substitutions when there are strong transition transversion and G + C-content biases. **Molecular Biology and Evolution**. n. 9, p. 678-87, 1992.
- TAMURA, K.; STECHER, G.; PETERSON, D. et al. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. **Molecular Biology and Evolution**, n.30, p. 2725-29, 2013.
- WANG, M.; SHENG, Z.; W.U, X. Mitochondrial DNA- Based Genetic Structure Analysis of Pudong White Pigs. **Biochemical Genetics**. n.48, p.924–37, 2010.
- WU, G.S.; YAO, Y.G.; QU, K.X. et al. Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. **Genome Biol.** n. 8, p.245, 2007.
- YANG, S. L.; WANG, Z. G.; LIU, B. et al. Genetic variation and relationships of eighteen Chinese indigenous pig breeds. **Gen. Sel. Evol.** n.35, p.657–671, 2003.
- YU, G.; XIANG, H. et al. The phylogenetic status of typical Chinese native pigs: analyzed by Asian and European pig mitochondrial genome sequences. **J Anim Sci Biotechnol.** v. 4, n.1, p. 9, 2013.
- ZHANG, Z.G.; LI, B.T.; CHEN, X.H. Pig Breeds in China. Shanghai: Shanghai Scientific and Technical Publishers. p. 1-24, 1986.

## CAPÍTULO 4

### VARIABILIDADE EM GENES DE RECEPTORES DE NUTRIENTES E DO GOSTO EM SUÍNOS

#### RESUMO

Acredita-se que os receptores do gosto desempenham um papel fundamental na sobrevivência de uma espécie e/ou raça. As variações observadas na percepção do gosto estão sendo relacionadas aos polimorfismos de genes de receptores em populações humanas que podem estar relacionados a adaptações geográficas/ambientais. O repertório de genes de receptores do gosto em suínos foi caracterizado parcialmente e, por isso, as informações são limitadas sobre associações com variação genética, raças e localização geográfica. Foram realizadas análises de sequências de 21 genes de receptores de nutrientes e do gosto em 79 genomas que representam 14 raças/populações suínas, incluindo três raças brasileiras: Piau, Monteiro e Moura. O subconjunto de genes de receptores do gosto amargo (Tas2R) apresentaram mais variabilidade que o subconjunto de não-Tas2R nas diferentes raças de suínos baseando-se nos polimorfismos não-sinônimos de nucleotídeo único (nsSNP). Foram encontrados 12.235 SNPs, dos quais 110 foram não sinônimos. A população ibérica foi a menos variável ( $\pi=0,0011$ ), enquanto as populações, crioula americana e brasileira foram mais variáveis ( $\pi=0,0030$  e  $0,0032$ , respectivamente). Em particular, o gene Tas2R39 mostrou maior número (19) de SNPs não-sinônimos, enquanto no gene GPR120 (sensor de PUFA) esses polimorfismos foram escassos. A população genômica desse estudo mostrou que os genes de receptores do gosto amargo (Tas2Rs) apresentaram uma diversidade nucleotídica mais alta que os genes não-Tas2Rs (aminoácidos e ácidos graxos), sugerindo que o gosto amargo é uma característica plástica, possivelmente associada com a habilidade dos suínos adaptarem-se a vários ecossistemas e que pode ser passível de seleção em programas de melhoramento e conservação.

**PALAVRAS-CHAVE:** Tas2R, mutações não-sinônimas, SNPs, variabilidade genética, genoma suíno

## CHAPTER 4

### VARIABILITY OF THE NUTRIENT AND TASTE RECEPTOR GENES IN PIGS

#### ABSTRACT

Observed variations in the perception of taste are related to gene polymorphisms receptors in human populations that may be related to geographical/ environmental adaptations. In pigs, such information is scarce because the repertoire of taste receptor genes in pigs was partially characterized and therefore, the information is limited on associations with genetic variation, breeds and geographic location. Therefore, sequence analysis of the 21 taste and nutrient receptor genes on 79 genomes representing 14 breeds /pig populations, including three Brazilian breeds, such as Piau, Monteiro Moura were performed. The analyzes revealed that the subset of genes of bitter taste receptors (Tas2R) showed more variability than the subset of non-Tas2R in different pig breeds based on polymorphisms in non-synonymous single nucleotide (nsSNP). A total of 12,235 SNPs were found in the 21 genes, of which 110 were non-synonymous SNPs. The Iberian population was less variable ( $\pi = 0.0011$ ), while populations, American and Brazilian Creole were more variable ( $\pi = 0.0030$  and  $0.0032$ , respectively). In particular, the Tas2R39 gene showed highest number (19) of non-synonymous SNPs in GPR120 while sensor (PUFA) these gene polymorphisms were scarce. The genomic population of this study montrou that bitter taste receptors genes (Tas2Rs) showed a higher nucleotide diversity than non-Tas2Rs (amino acids and fatty acids) suggesting that the bitter taste is a plastic characteristic, possibly associated with ability of pigs adapt to various ecosystems and may be subject to selection in breeding and conservation programs.

**KEY-WORDS:** Tas2R, nonsynonymous mutations, SNPs, genetic variability, pig genome



## 1. INTRODUÇÃO

A dieta pode desempenhar um papel como uma pressão de seleção, uma vez que vários genes envolvidos na percepção sensorial do gosto estão entre os genes mais conservados em seis genomas de mamíferos (Kosiol et al., 2008). Existem cinco tipos de sabores presentes na cavidade oral que foram mais estudados até o momento: o sabor amargo que detecta a presença de componentes tóxicos nos alimentos e os que estão relacionados a carboidratos (doce), aminoácidos (umami), sódio (sal) e ácidos (azedo), além dos que detectam nutrientes, como ácidos graxos e uma variedade de aminoácidos/proteínas (Kinnamon & Cummings, 1992; Bachmanov & Beauchamp, 2007). Receptores para todas essas categorias de receptores de sabor e nutrientes foram caracterizados em humanos e roedores (Conte et al., 2002; Shi et al., 2003; Bachmanov & Beauchamp, 2007; Wellendorph et al., 2010).

Os suínos são mais tolerantes ao gosto amargo que os humanos. Estudos comparativos entre as duas espécies revelaram que isso é devido à presença de maior número de receptores para o gosto amargo encontrados na cavidade oral dos humanos (pelo menos 25 funcionais segundo Conte et al., 2002), que na cavidade oral de suínos (Groenen et al., 2012).

O surgimento das tecnologias de sequenciamento de nova geração permitiu grandes avanços nos estudos genômicos de diversas espécies de animais domésticos de interesse econômico (International Chicken Genome Sequencing Consortium, 2004; Bovine Genome Sequencing and Analysis Consortium et al., 2009), incluindo a suína que teve o genoma completo de um animal da raça Duroc sequenciado pelo Consórcio Internacional de Sequenciamento do Genoma Suíno (SGSC) (Archibald, Bolund et al., 2010). Na versão atual do ensembl v75 existem 25.322 genes codificadores anotados no *S. scrofa assembly* 10.2 ([http://www.ensembl.org/Sus\\_scrofa/Info/](http://www.ensembl.org/Sus_scrofa/Info/)). No entanto, ainda são

excassas as informações sobre todos os genes de receptores de sabores e de nutrientes presentes na cavidade oral de suínos (Groenen et al. 2012).

Os genes de receptores do gosto e de nutrientes são de grande importância para a suinocultura, uma vez que estão envolvidos na escolha do tipo de alimentos pelos animais, além disso, alguns desses genes de receptores desempenham um importante papel na detecção de toxinas que podem comprometer a saúde e até mesmo a vida dos animais (Wong et al., 1996; Dong et al., 2009; Li & Zhang, 2014). Li & Zhang (2014) estudaram os genes do gosto amargo em pelo menos 54 espécies de vertebrados e verificaram que a diversidade genética, observada nesses genes, pode está relacionada com a adaptação das espécies para detectar a presença de toxinas nas dietas.

Diante do exposto, o objetivo deste estudo foi quantificar e comparar a variabilidade dos genes de receptores do gosto (chamados de TASR) em diferentes raças suínas domésticas, incluindo três raças brasileiras: Piau, Monteiro e Moura, e javalis espalhados pelo mundo.

## 2. MATERIAL E MÉTODOS

### 2.1 Amostragem e sequenciamento

Neste estudo foram analisadas sequências do genoma completo (WGS), que foram obtidas por sequenciamento *shotgun*, de 77 suínos internacionais, crioulos americanos, europeus e raças domésticas asiáticas, javalis europeus e asiáticos. Dois javalis de Sumatra foram incluídos nas análises como grupo externo (*outgroup*) (Tabela 1). Das 77 sequências, 51 foram obtidas do Sequence Read Archive-SRA (Rubin et al., 2012; Fang et al., 2012; Esteve-Codina et al., 2013) e 28 são sequências não publicadas (na Tabela 16C do ANEXO C encontram-se os números de acesso de cada amostra). As novas sequências foram obtidas com a tecnologia HiSeq da Illumina, “paired end reads” de 100 pb de comprimento. Arquivos alinhados no formato bam foram obtidos para as 51 sequências (Rubin et al., 2012) e para o restante, os *reads* foram alinhados usando a ferramenta Burrows Wheeler Alignment (BWA) (Li et al., 2009) seguido de 7 mismatches.

### 2.2 SNP calling e anotação

Os genes candidatos de suínos analisados neste estudo incluíram todos os 21 receptores do gosto disponíveis no *assembly Sscrofa* 10.2 para sabor amargo (Tas2R7C, Tas2R9, Tas2R10, Tas2R16, Tas2R20, Tas2R38, Tas2R39, Tas2R41, Tas2R42 e Tas2R60), receptor para aminoácido (GPRC6A, mGLUR1, mGLUR4, Tas2R3, Tas1R1 e CaSR) e receptores de ácidos graxos (GPR40, GPRA43, GPR41, GPR120 e GPR84).

Os polimorfismos de nucleotídeo único (SNPs) foram obtidos para cada amostra individualmente usando SAMtools v.0.0.18 com a função mpileup (Li et al., 2009), filtrando por qualidade de base e mapeamento de pelo menos 20. O “*depth*” mínimo e máximo foram inseridos para cinco e duas vezes o *depth* médio por amostra, respectivamente. Os arquivos Variant Call Format versão 4.0 (VCF) resultando da “SNP calling” foram então unidos em um múltiplo VCF individual usando scripts customizados em PERL. Para posições perdidas, os arquivos bam foram inspecionados para checar se o alelo ancestral estava presente (sempre filtrando pelo mesmo critério de qualidade como descrito acima) e o arquivo vcf foi completado se possível.

Após obter os arquivos vcf unidos, a região de interesse dos 21 genes candidatos distribuídos entre os cromossomos suínos 1, 5, 6, 7, 13, 14 e 18 foram obtidos para análise dessas janelas menores, adicionando-se, quando possível, 10 kb flanqueando regiões de acordo com as coordenadas dos genes de referência (Sscrofa 10.2). Se dois genes foram mais próximos do que 20kb, a região intergênica foi diminuída e designada a cada gene correspondente.

Cada SNP foi anotado com a ferramenta *Variant Effect Predictor* (VEP) usando script em PERL disponível no Ensembl: <http://www.ensembl.org/tools.html> (Flicek, 2013), usando o banco de dados Ensembl v. 72. Isso foi feito apenas para aqueles genes (Tas2R9, Tas2R39, Tas2R41, Tas2R60, GPRC6A, mGLUR1, mGLUR4, Tas1R1, Tas2R3, CaSR, GPR40, GPR43, GPR41, GPR120 e GPR84) onde a anotação oficial coincidiu com nossa anotação obtida manualmente e com MSTATSPOP.

Durante a etapa de anotação foram incluídos parâmetros padrões com a opção *Sorting Intolerant From Tolerant* (SIFT), para prever o efeito da substituição de aminoácido na função da proteína (Ng & Henikoff, 2003) para os SNPs não-sinônimos (nsSNPs). Para ter acesso aos números de referência (rsID) para cada SNPs no banco de dados de dbSNP foi utilizada a opção para identificar variantes colocadas que retorna o número de ID do SNP. Para os demais genes, a classificação de cada SNP quanto à classe (em intergênica, exônica, intrônica e nas regiões não traduzidas (UTRs) como também consequências de em transcritos foi realizada manualmente ou com o programa MSTATSPOP v.0.998978b: <http://bioinformatics.cragenomica.es/numgenomics/people/sebas/software/software.html> (Ramos-Onsins, unpublished). Arquivos no formato GFF3 v3 (disponível no ANEXO D) e FASTA de cada gene foram gerados usando scripts customizados em PERL. Os arquivos

de formato FASTA, no qual as posições perdidas são substituídas por N, foram usados como entrada para o programa MSTATSPOP.

### 2.3 Análises genéticas

A frequência alélica para cada SNP e por população foi calculada com o programa VCFtools versão 0.1.11 (Danecek et al., 2011) e com o programa mstatpop (Ramos-Onsins não publicado, disponível em <http://bioinformatics.cragenomica.es/numgenomics//people/sebas/software/software.html>) foram estimadas as percentagens de dados perdidos e parâmetros de diversidade tais como diversidade nucleotídica total ( $\pi_t$ ), por exemplo, considerando a região total que incluiu região total do gene e região intergênica ( $\pi_g$ ), apenas para região intergênica ( $\pi_{int}$ ), íntrons ( $\pi_i$ ), éxons ( $\pi_e$ ) para genes com mais de um éxon (para aqueles genes com apenas um éxon a diversidade nucleotídica é a mesma encontrada na região do gene), e em UTRs ( $\pi_{utr}$ ).

A taxa de variabilidade sinônima ( $\pi_s$ ) e não-sinônima ( $\pi_a$ ) foi calculada para identificar pressão de seleção sobre os genes de receptores do gosto ( $\pi_a/\pi_s$ ). Uma razão ( $\omega = \pi_a/\pi_s$ )  $> 1$  é um indicativo de seleção positiva e menor que um sugere seleção purificadora, e uma razão de um pode indicar neutralidade (Zhang et al., 2006). Para cada gene foi calculado o índice de fixação ( $F_{ST} = 1 - (\pi_w/\pi_t)$ ), onde  $\pi_w$  é o número médio de diferenças nucleotídicas entre duas sequências nas populações e  $\pi_t$  é o número de diferenças nucleotídicas na população total. Todos esses parâmetros foram calculados com o programa mstatpop. A significância foi computada com 1000 permutações. Erro padrão aproximado (SE) de diversidade nucleotídica para cada gene foi obtido usando o simulador de coalescência neutra no DnaSP v5 (Librado & Rozas, 2009) pela geração de 95% de intervalo de confiança (IC) incluindo 1000 amostras aleatórias e usando como padrão um modelo de taxa de recombinação intermediário ( $R=10$ ). Para esta simulação, também foram inseridas estimativas de variabilidade nucleotídica, diversidade e número de sítios corrigidos para dados perdidos computado com mstatpop.

Análises de componentes principais (PCA) foram conduzidas em R (R Development Core Team, 2008) (<http://cran.rproject.org>) usando os arquivos no formato de entrada para o programa PLINK (Purcell et al., 2007) que foram extraídos de arquivos VCF usando um script customizado em linguagem PERL. Essa análise foi realizado para o

painel completo de SNPs, para o painel de SNPs não-sinônimos e para os grupos de genes (receptores para o gosto amargo, de aminoácido e de ácidos graxos), visando estudar a estrutura genética da população. A relação genética foi analisada com o STRUCTURE versão 2.3.4 (Pritchard, 2000).

A análise de estrutura foi realizada de duas formas: (1) incluindo apenas nsSNP e (2) incluindo um conjunto de SNPs da região codificadora e não codificadora. Nessa análise foram inseridas cinco permutações para cada número de populações ( $K$ ), que variou de 1 a 15, 100.000 MCMC (Markov chain Monte Carlo) e um período de burning de 10.000 passos. Foram selecionados os parâmetros 'admixture' e frequências alélicas correlacionadas. O número significativo de  $K$  para os diferentes clusters foi obtido pela estatística Delta  $K$  (Evanno et al., 2005) que foi calculado usando STRUCTURE HARVESTER versão 0.9.93 (Earl, 2011).

### 3. RESULTADOS E DISCUSSÃO

#### 3.1 Resultados

##### 3.1.1 Distribuição dos SNPs nas regiões analisadas

Nesse estudo foi realizada uma análise dos genes 21 dos 28 genes que presentes na atual montagem do genoma suíno (*Sus scrofa* 10.2). Sete genes (Tas1r2, Tas2r1, Tas2r134, Tas2r3, Tas2r40, Tas2r4 and GPR92) foram excluídos das análises porque foram isolados em *contigs* e não em um dos 18 cromossomos autossômicos ou dos cromossomos sexuais. Foram analisados dados de sequenciamento de DNA de nova geração (NGS) por *shotgun* de 79 genomas completos pertencentes a 14 populações ou raças distribuídas pelo mundo; 28 dos genomas não foram publicados (Tabela 1).

Os novos genomas, inicialmente de suínos ibéricos e americanos (Crioulo), foram um subconjunto daqueles descritos previamente por Burgos-Paz et al. (2013). As amostras foram agrupadas em internacional (incluindo raças altamente selecionadas e bem conhecidas, como Large White, Landrace, Duroc, Pietrain e Hampshire), suínos crioulos de vários países americanos, raças locais do Brasil (Moura, Monteiro e Piau), raças chinesas (Meishan, Xiang, Jiangquhai e Wuzhishan) e javalis da Europa e Ásia (Tabela 1).

Um total de 12.235 SNPs foi encontrado em todos os 21 genes e regiões flanqueantes (Tabela 2). A taxa média de transições vs. transversões foi  $ti/tv = 2.35$  e foi similar a taxa obtida para o genoma completo de suínos (Amaral et al., 2011) e a que foi encontrada em outras espécies de mamíferos (Lindblad-Toh et al., 2000; Rosenberg et al., 2003).

Tabela 1 Raças, códigos e distribuição dos 79 suínos analisados.

Grupo	Raças	País	Código	N (n)
Internacional (INT)	Large White	NA	LW	15 (1)
	Landrace	Vários países	LR	5
	Duroc	Vários países	DU	4
	Pietrain	Vários países	PI	5
	Hampshire	Vários países	HS	2
Domésticos europeus	Ibéricos	Espanha	IB	4(4)
Javalis europeus	javalis	Holanda, Suécia, França e Espanha	EUWB	9(3)
Domésticos asiáticos	Meishan, Xiang, Jiangquhai e Wuzhishan	China	ASD	8
Javalis asiáticos	javalis	Japão, Sul e Norte da china, Este da Russia	ASWB	6(1)
Crioulos	Crioulos, Ossabaw, Yucatan	Peru, Cuba, Argentina e Estados Unidos	CR	14(14)
Brasileiros	Brasileiros <sup>a</sup>	Brasil	BR	3(3)
Sem grupo	Tamworth	Reino Unido	TW	1(1)
Sem grupo	‘Manchado Jabugo’	Espanha	MJ	1(1)
Outgroup	Javalis de Sumatra	Sumatra	SWB	2

<sup>a</sup> Raças Piau, Monteiro e Moura, (n)=amostras novas sequenciadas neste estudo.

Um total de 8.962 SNPs (73%) foi previamente identificado com número de referência (refSNP) do dbSNP v. 138. Dos 8.259 SNPs posicionados entre 5' e 3' UTRs, 7.963 estiveram presentes em íntrons, 296 em éxons nas regiões codificadoras de proteínas, e apenas 17 em UTRs (12 em 5'UTR e 5 em 3'UTR). Entre os SNPs funcionais, presentes nas regiões codificadoras, foram detectados um códon de parada perdido, um códon de parada ganho, 167 sinônimos e 110 não-sinônimos. Os 3.274 SNPs novos serão submetidos ao dbSNP.



Um aspecto preocupante dos dados de NGS por *shotgun* é o fato que a cobertura é um processo aleatório e isso é, portanto, menos comum que todas as amostras tenham suficiente *depth* e qualidade para serem analisadas. Nos dados desse estudo foi encontrada uma taxa média de 20% de dados perdidos ou faltantes (Tabela 2). Esse percentual, no entanto, foi acima de 50% para dois genes (CaSR e Tas1R3) e, por isso, foram removidos das análises de variabilidade que foram baseadas em 19 *loci*.

### 3.1.2 Padrão de variação nucleotídica

A diversidade nucleotídica e o índice de fixação global ( $F_{ST}$ ) foram calculados usando mstatpop (Ramos-Onsins não publicado). Este programa fornece estimativas não viesadas de estatísticas genéticas básicas até em dados com altas taxas de informações perdidas ou faltantes e aplica a teoria desenvolvida por Ferretti et al. (2013) (Tabela 3). A variabilidade média foi de  $2,1 \times 10^{-3}$  nas populações analisadas, que foi comparável àquelas encontradas nas regiões flanqueantes (média de  $1,8 \times 10^{-3}$ ). A taxa de variabilidade sinônima ( $\pi_s$ ) foi, em média,  $3,8 \times 10^{-3}$ , e a taxa de não sinônima ( $\pi_a$ ) foi quatro vezes menos (média de  $1,2 \times 10^{-3}$ ), concordando com muitos resultados encontrados na literatura (Makalowski & Boguski 1998) e também consistente com o modelo de prevalência de seleção purificadora. Nas regiões do gene, a diversidade nucleotídica ( $\pi_g$ ) variou de  $0,5 \times 10^{-3}$  em GPRC6A a  $4,7 \times 10^{-3}$  em Tas2R42.

Os genes de receptores do gosto amargo exibiram mais alta diversidade nucleotídica quando foram analisados sem a região intergênica (média de  $\pi_g = 2,6 \times 10^{-3}$ ) que com as regiões intergênicas (média de  $\pi_i = 1,9 \times 10^{-3}$ ). O oposto foi observado para os genes dos demais grupos de genes que mostraram maior diversidade na região completa ( $\pi_t = 2,1 \times 10^{-3}$ ). Os grupos de genes de receptores para aminoácidos e ácidos graxos mostraram menor diversidade nucleotídica que o observado para os genes de receptores para o gosto amargo (Tabelas 3 e 4).

Em geral, a variabilidade desses genes, especialmente aqueles genes para receptores do gosto amargo, são mais altas que àquelas normalmente descritas para genoma completo de suínos que estão na ordem de  $1,2 \times 10^{-3}$  para raças suínas internacionais e  $0,7 \times 10^{-3}$  para suínos ibéricos (Amaral et al., 2011; Groenen et al., 2012; Esteve-Codina et al., 2013).

Estimativas de diversidade nucleotídica na região dos genes e na região total variaram grandemente entre genes (Tabela 3) e entre populações (Tabela 17C do ANEXO C). Os suínos domésticos asiáticos (ASD), javalis asiáticos (ASWB) e javalis europeus (EUWB) exibiram alta variabilidade com valor médio de  $\pi_g=2,3 \times 10^{-3}$ . A população ibérica foi a menos variável ( $\pi_g = 1,2 \times 10^{-3}$ ) enquanto os suínos americanos e brasileiros apresentaram o mais alto nível de diversidade ( $\pi_g=2,6 \times 10^{-3}$  e  $\pi_g=2,9 \times 10^{-3}$ , respectivamente) devido à ancestralidade miscigenada (Burgos-Paz et al., 2013).

A razão de substituições sinônimas e não sinônimas ( $\omega=\pi_a/\pi_s$ ) foi menor que 1 em todos os genes (Tabela 3), como esperado para seleção purificadora ou negativa prevalente. Alguns casos extremos foram observados nos genes; não foram encontrados polimorfismos não-sinônimos em GPR120 nem sinônimos em Tas2R9. Cinco genes (Tas2R10, Tas2R39, Tas2R41, Tas2R7C e GPRC6A) apresentaram valores de  $\omega$  maiores que 0,5 e menor que 1, comumente proveniente de uma seleção purificadora mais fraca. Portanto, a seleção purificadora foi prevalente e pode ser considerada uma força relevante e responsável pelas variações genéticas observadas entre os genes TasR em suínos.

A diversidade nucleotídica também foi analisada por grupo de gene e por população (Tabela 4). A população brasileira, crioula e EUWB mostraram a mais alta variabilidade para Tas2R, principalmente, quando foram analisadas apenas regiões dos

Tabela 2 Informações gerais sobre genes de receptores do gosto em suínos: cromossomo, posição inicial e final, SNPs totais e distribuição por região genômica, consequências funcionais e percentuais de dados perdidos (%DP). As análises foram realizadas em 77 suínos (os dois javalis de Sumatra não foram incluídos).

Grupo de genes	Genes	exon*	Cr.	Região total		Região do gene			SNP distribution by class and functional consequences								
				Coordenadas (pb)	SNPs	Coordenadas (pb)	%DP	SNPs	Intergênica	Exon	Intron	5'UTR	3'UTR	Perda stop	Ganho stop	Si	Ns
Amargo	Tas2R20	1	5	63894140-63915054	310	63904140-63905054	10.98	13	297	13	-	-	-	-	-	5	8
	Tas2R9	1	5	63971739-63982674	229	63976739-63977674	9.10	6	223	6	-	-	-	-	-	0	6
	Tas2R10	1	5	63958146-63971725	266	63965446-63966375	12.12	9	257	9	-	-	-	-	-	2	7
	Tas2R42	1	5	63857091-63878041	257	63867091-63868041	10.06	17	240	17	-	-	-	-	-	9	8
	Tas2R16	1	18	25873452-25894354	474	25883452-25884354	10.25	16	458	16	-	-	-	-	-	8	8
	Tas2R38	1	18	8347518-8368525	79	8357518-8358525	16.13	21	58	21	-	-	-	-	-	11	10
	Tas2R39	1	18	7348848-7369855	310	7358848-7359855	10.08	28	282	28	-	-	-	-	1	8	19
	Tas2R41	1	18	7008806-7029729	88	7018806-7019729	20.90	18	70	18	-	-	-	-	-	6	12
	Tas2R60	1	18	7035247-7056597	110	7045247-7046597	11.59	24	86	14	10	-	-	-	-	10	4
Tas2R7C	1	5	63982692-63996080	254	63985142-63986080	8.82	10	244	10	-	-	-	-	-	5	5	
Aminoácidos	TAS1R3 <sup>a</sup>	6	6	58109541-58116535	34	58111612-58115907	57.76	17	17	14	3	1	4	-	-	8	1
	GRM4	10	7	34829241-34938071	2344	34839241-34928071	40.00	1976	368	36	1940	8	-	-	-	26	2
	GPRC6A	3	1	50121244-50146085	253	50131244-50136085	11.74	41	212	6	35	-	-	-	-	3	3
	MGLUR1	7	1	21466379-21824964	5671	21476379-21814964	16.07	5421	250	20	5401	3	1	-	-	15	1
	CASR <sup>a</sup>	6	13	147897932-147940283	345	147907932-147935070	75.62	138	208	22	116	-	-	-	-	20	2
	TAS1R1	7	6	62349877-62363714	117	62350603-62363204	31.62	112	5	13	99	-	-	1	-	7	5
Ácidos graxos	GPR40	1	6	40335081-40343065	89	40339566-40340468	50.96	8	81	8	-	-	-	-	-	4	4
	GPR43	1	6	40272039-40293031	243	40282039-40283031	40.38	11	232	11	-	-	-	-	-	10	1
	GPR41	1	6	40323922-40335070	172	40333922-40334902	43.46	11	161	11	-	-	-	-	-	8	3
	GPR120	3	14	114733575-114769008	516	114743575-114765158	18.73	360	156	1	359	-	-	-	-	1	0
	GPR84	2	5	20404644-20416116	74	20410225-20411195	31.99	2	72	2	-	-	-	-	-	1	1
Total			778077†	12235	480139†		8259	3977	296	7963	12	5	1	1	167	110	

<sup>a</sup> Genes excluídos pela alta taxa de dados perdidos; Si=sítios sinônimos e Ns= sítios não sinônimos. \*número de exons por gene. † Comprimento total em pares de bases para todos os genes (pb).

A % de dados perdidos, para cada posição, foi calculado como o número de genótipos não chamáveis, e também por apresentarem baixa qualidade ou baixo *depth* (um *depth* mínimo de 5x foi requerido), dividido pelo total de número de amostras (77). A taxa total de dados perdidos foi a média em todas as posições.

genes (médias  $\pi_g = 3,6 \times 10^{-3} \pm 1,0 \times 10^{-3}$ ,  $\pi_g = 3,3 \times 10^{-3} \pm 0,8 \times 10^{-3}$  e  $\pi_g = 3,1 \times 10^{-3} \pm 0,8 \times 10^{-3}$ , respectivamente) e a população ibérica, em comparação as outras populações, mostraram duas vezes menos diversidade nucleotídica para esse mesmo grupo de genes (média  $\pi_g = 2,0 \times 10^{-3} \pm 0,4 \times 10^{-3}$ ). No entanto, as demais populações suínas analisadas exibiram uma alta variabilidade nesses genes, exceto nos suínos ibéricos que foram analisados juntamente com uma linhagem altamente endogâmica a ‘Guadyerbas’ (Toro et al., 2000).

Tabela 3 Diversidade nucleotídica por região genômica ( $\pi$ ), sinônima ( $\pi_s$ ) e não-sinônima ( $\pi_a$ ) e índice de fixação ( $F_{ST}$ ) para cada gene dos receptores do gosto e um pseudogene analisados na população suína total excluindo os dois javalis de Sumatra (n=77).

Grupo de genes	Genes	$\pi_i \pm SE \times 10^3$	$\pi_{int} \pm SE \times 10^3$	$\pi_g \pm SE \times 10^3$	$\pi_e \pm SE \times 10^3$	$\pi_i \pm SE \times 10^3$	$\pi_{utr} \pm SE \times 10^3$	$\pi_s$ x 10 <sup>3</sup>	$\pi_a / \pi_s$	$F_{ST}^a$
Amargo	Tas2R20	2,0±0,6	2,1±0,5	2,7±1,2	2,7±1,2	-	-	5,5	0,3359	0,28*
	Tas2R9	2,3±0,7	2,4±0,7	1,0±0,6	1,0±0,6	-	-	0,0	NA†	0,41*
	Tas2R10	1,9±0,6	2,0±0,6	0,8±0,6	0,8±0,6	-	-	1,0	0,6879	0,38*
	Tas2R42	2,0±0,6	1,9±0,9	4,7±1,6	4,7±1,6	-	-	10,0	0,3098	0,29*
	Tas2R16	3,1±0,8	3,2±0,5	2,2±0,9	2,2±0,9	-	-	4,9	0,2745	0,04
	Tas2R38	0,7±0,2	0,5±0,2	3,6±1,3	3,6±1,3	-	-	8,6	0,2371	0,29*
	Tas2R39	3,0±0,2	3,0±0,1	4,2±1,5	4,2±1,5	-	-	5,7	0,6438	0,12
	Tas2R41	0,5±0,2	0,5±0,2	2,3±1,0	2,3±1,0	-	-	3,3	0,6041	0,29*
	Tas2R60	0,8±0,2	0,7±0,2	2,4±0,9	1,6±1,0	4,3±1,6	-	4,8	0,0940	0,38*
Tas2R7C	2,9±0,2	3,0±0,2	1,8±0,8	1,8±0,8	-	-	2,8	0,5311	0,15*	
Amino-ácidos	GRM4	3,6±0,5	3,2±0,3	3,7±0,4	2,1±0,6	3,7±0,4	0,8±0,7	7,7	0,0188	0,28*
	GPRC6A	1,0±0,8	1,1±0,8	0,5±0,2	0,5±0,3	0,6±0,3	-	0,8	0,6738	0,29*
	MGLUR1	2,1±0,3	1,5±0,3	2,1±0,2	0,8±0,4	2,1±0,7	3,0±1,8	3,5	0,0007	0,18*
	TAS1R1	0,8±0,3	0,5±0,8	0,9±0,3	0,4±0,3	1,0±0,3	-	0,8	0,3351	0,36*
Ácidos graxos	GPR40	1,9±0,3	1,9±0,4	1,6±0,8	1,6±0,8	-	-	3,6	0,1940	0,23*
	GPR43	1,9±0,6	2,0±0,5	1,5±0,8	1,5±0,8	-	-	5,1	0,0378	0,09
	GPR41	2,9±0,6	3,1±0,5	1,1±0,7	1,1±0,7	-	-	3,4	0,0986	0,26*
	GPR120	1,9±0,9	1,4±0,6	2,2±0,5	0,3±0,2	2,3±0,7	-	1,3	0,0000	0,15
	GPR84	0,5±0,5	0,6±0,5	0,1±0,1	0,1±0,1	-	-	0,1	0,3891	0,26*
Média		1,9±0,6	1,8±0,6	2,1±0,9	1,7±0,9	2,3±0,7	1,9±1,2	3,8		

SE=erro padrão, a=computado usando a região total (intergenômica e gene), Diversidade nucleotídica para total ( $\pi_t$ ), intergenômica ( $\pi_{int}$ ), gene ( $\pi_g$ , incluiu: CDS, íntrons e UTRs), éxons, íntrons ( $\pi_i$ ) e em UTR ( $\pi_{utr}$ ). †Gene sem mutações sinônimas. \*P<0,05, Significância baseada em 1000 permutações.

Tabela 4 Diversidade nucleotídica na região total ( $\pi_t$ ) e na região do gene ( $\pi_g$ ) por população e por grupo de genes.

Populações*	N	Amargo		Aminoácidos		Ácido graxos		Todos TasR	
		$\pi_t \pm SE \times 10^3$	$\pi_g \pm SE \times 10^3$	$\pi_t \pm SE \times 10^3$	$\pi_g \pm SE \times 10^3$	$\pi_t \pm SE \times 10^3$	$\pi_g \pm SE \times 10^3$	$\pi_t \pm SE \times 10^3$	$\pi_g \pm SE \times 10^3$
Internacional	31	1,8±0,3	2,8±0,5	1,9±0,1	1,7±0,0	1,4±0,3	1,0±0,2	1.8±0.3	2.2±0.4
Ibéricos	04	1,1±0,1	2,0±0,4	0,5±0,0	0,4±0,0	0,4±0,0	0,2±0,1	0.7±0.1	1.2±0.2
Crioulos	14	2,7±0,6	3,3±0,85	1,8±0,3	1,7±0,1	2,5,0±0,3	1,8±0,3	2.5±0.5	2.6±0.5
Brasileiros	03	2,8±0,5	3,6±1,0	2,1±0,0	2,1±0,1	2,9±0,4	1,7±0,3	2.8±0.5	2.90±0.7
Asiáticos	08	2,0±0,3	2,6±0,7	2,0±0,1	1,7±0,0	2,1±0,4	1,9±0,5	2.1±0.4	2.3±0.6
WB asiáticos	06	2,5±0,6	2,5±0,60	2,2±0,3	2,1±0,1	1,9±0,2	1,7±0,4	2.4±0.5	2.3±0.6
WB europeus	09	2,1±0,3	3,1±0,8	1,0±0,1	0,9±0,0	0,4±0,0	0,5±0,1	1.5±0.2	1.9±0.5
Total	75	2,1±0,4	2,8±0,7	1,7±0,1	1,5±0,1	1,6±0,2	1,2±0,3	2.0±0.3	2.2±0.5

Amostras de Tamworth, 'Manchado de Jabugo', e dos dois javalis de Sumatra não foram incluídos.

### 3.1.3 Estrutura e filogeografia

Os suínos, como muitas espécies domésticas, estão organizados em raças com diferenças fenotípicas específicas que são geneticamente isoladas ou com limitadas trocas genéticas. Os altos índices de diferenciação ( $F_{ST}$ ) são, portanto, esperados em espécies estruturadas com ampla distribuição e muitas raças especializadas que impede o fluxo de genes entre elas. Os valores de  $F_{ST}$  para 15 genes foram significativamente diferentes de zero, exceto em quatro genes: Tas2R16, Tas2R39, GPR43 e GPR120 (Tabela 2) indicando que a diferenciação entre as populações é resultante de um fluxo de genes limitado entre as populações. Os genes Tas2R apresentaram mais alto grau de diferenciação. Os valores de  $F_{ST}$  significativos variaram de 0,15 a 0,41.

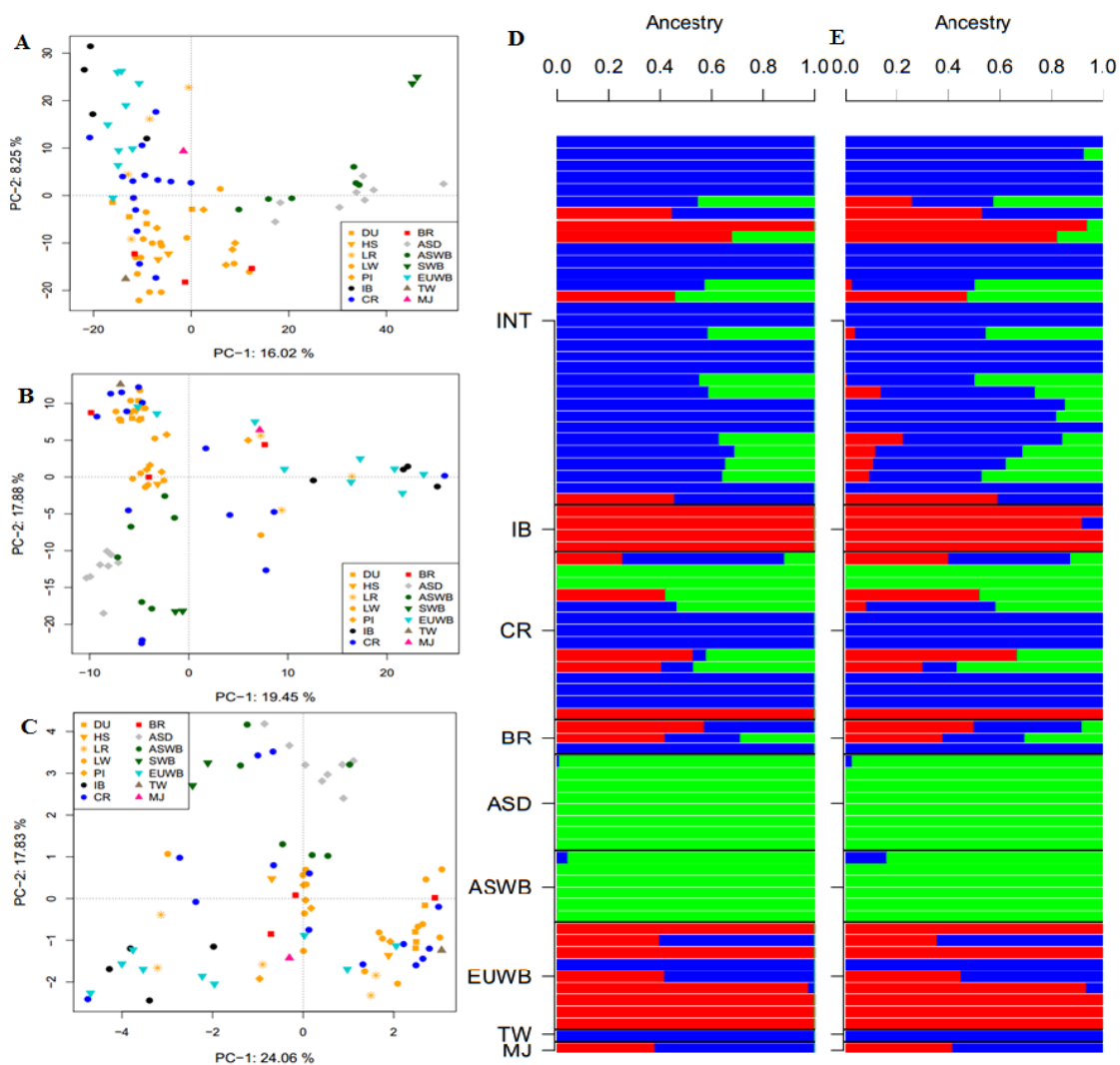
As análises de Componentes Principais e STRUCTURE (Figura 1) mostraram uma separação entre asiáticos e raças locais européias (Ibéricas) como é tipicamente observado no mtDNA e em termos autossomal (Yang et al., 2011; Yu et al., 2013), porém, isso é uma série contínua em vez de uma clara separação geográfica, o que foi observado entre Ásia e Europa. A Figura 1A mostra o gráfico com todos os SNPs dos 19 genes. O primeiro axis explicou 16,02% da variância e, isso discriminou a separação entre Ásia e Europa com as raças internacionais agrupando-se entre elas, mas existe uma série contínua ao invés de uma divisão abrupta (e.g. o leitor é convidado a comparar esse resultado da Figura 1A com a Figura 1 a partir de Burgos-Paz, Souza et al. 2013). O segundo axis na Figura 1A contabilizou 8,25% da variância e separou as raças altamente selecionadas dos javalis e das raças locais que não são selecionadas (ibéricas e chinesas), sugerindo que a seleção moderna exerceu uma consistente influência nas raças modificando o padrão de variabilidade observado nos suínos atuais. Suínos crioulos e brasileiros tendem a cair no cluster internacional (Figura 1A). Esse padrão é mais pronunciado quando apenas polimorfismos de genes de receptores do gosto amargo são considerados (Figure 1B), com algumas mudanças interessantes: o primeiro axis (19,45% variância) agora separa javalis europeus e ibéricos vs. as demais populações, e o segundo axis (17,88% variância) distingue Ásia do restante; algumas raças internacionais (Large White) foram fortemente agrupadas devido ao baixo nível de diversidade nucleotídica que foi mais baixo do que a média para esse grupo de genes ( $\pi_t = 1,9 \times 10^{-3} \pm 1,0 \times 10^{-3}$ ), sugerindo um forte impacto da pressão seletiva nos TasRs nas raças comerciais. Resultados

similares foram observados quando apenas SNPs não-sinônimos foram empregados (Figura 1C).

As análises com STRUCTURE utilizando dados de SNP dos genes de receptores do gosto amargo sugeriram que os valores ótimos de K foram 3 para nsSNP e 3 - 4 para o conjunto completo de SNPs (Figura 7C do ANEXO C) concordando com os resultados da PCA. Para K=3, as Figuras 1D, E mostram uma clara separação entre asiáticos (ASD e ASWB), europeus (IB e EUWB) e internacionais (INT). Houve uma grande heterogeneidade entre indivíduos em cada raça. Na população brasileira (BR), a raça Piau foi alocada ao cluster Internacional com 100% de probabilidade, enquanto Monteiro e Moura apresentam uma fração miscigenada de genoma de origem asiática e européia. Na EUWB, três indivíduos foram assignados como um importante componente de raça internacional. Para ambos os conjuntos de SNPs a população ibérica foi assignada ao próprio cluster e apenas um indivíduo apresentou 3% da composição do genoma assignado ao cluster asiático. Isso significa que a população ibérica foi altamente homogênia, como uma consequência de que os indivíduos sequenciados pertenciam a um rebanho altamente consanguíneo ou fechado.

### 3.1.4 Impacto funcional da substituição de aminoácidos

A ferramenta *Sorting Intolerant From Tolerant* (SIFT) implementada no Variant Effect Predictor do Ensembl foi usada para prever, *in silico*, as mudanças nos aminoácidos que afetam a função das proteínas de um conjunto de nsSNPs do dbSNP. Dos 110 nsSNPs, para os 19 TasR, foi possível prever o índice de tolerância para 59 nsSNPs, dos quais 11 (rs320709106, rs342189509, rs342228000, rs345262132, rs339482728, rs325274060, rs330666697, rs323728911, rs318787211 do dbSNP; e 5:63977077 e 1:21476805 dos novos SNPs) apresentaram uma pontuação para o índice de tolerância  $\leq 0,05$ , e podem, portanto, ser considerados potencialmente deletérios para a função da proteína (Tabela 5), enquanto os índices de tolerância médios  $\geq 0,05$ , a priori, podem causar um pequeno impacto funcional na substituição de aminoácidos (mais informações são mostrada na Tabela 18C do ANEXO).



**Figure 1** Análises de Componentes Principais (PCA) e estrutura. PCA foi realizada com todos os SNPs dos 19 genes, excluindo o pseudogene (A) e para diferentes subconjuntos: SNPs totais (B) e não-sinônimos para o sabor amargo (C). Análise de estrutura com todos os SNPs (D) e SNPs não-sinônimos para o gosto amargo (E). Cada indivíduo está representado por uma linha vertical com as proporções de atribuição a cada cluster mostradas no topo, e as diferentes cores são referentes aos diferentes grupos: internacional (INT) = azul; asiáticos (ASD e ASWB) = verde; e europeus (IB e EUWB) = vermelho.





Em geral, e de acordo com a natureza destas mutações potencialmente deletérias, esses alelos foram raros e foram frequentes em uma única população; eles são provavelmente mutações recentes que ainda não foram purgados. No entanto, destacam-se algumas mutações interessantes; dos 12 nsSNPs detectados no gene *Tas2R41*, quatro apresentaram um índice  $\leq 0.05$  e duas dessas mutações foram compartilhadas entre as populações: rs342228000 foi encontrado na internacional e ASD, e rs345262132 na brasileira, ASD e ASWB. Essas duas mutações, possivelmente, são alelos originados na Ásia que foram introduzidos nas raças estrangeiras (Europeias).

O gene *Tas2R9* apresentou um novo nsSNP (posição do SNP 5:63977077) que foi frequente apenas nos suínos ibéricos (25%). Porém, talvez, o caso mais interessante é que um dos nsSNPs rs330666697 (*Tas1R1*), que estava em frequência intermediária em domésticos Asiáticos ( $p=0,43$ ) também está presente nas populações internacional e americana, ou seja, essa mutação está sendo compartilhada entre populações e pode ainda estar sob pressão de seleção, provavelmente, seleção positiva. A alta frequência em domésticos asiáticos, mas ausente em javalis asiáticos sugere que esta mutação pode ter surgido após o processo de domesticação e que rapidamente aumentou a frequência posteriormente, pode ser porque suas consequências potencialmente deletérias foram compensadas por outras vantagens e foi positivamente selecionado. Muitos estudos funcionais adicionais são requeridos para confirmar esta hipótese.

### 3.2 Discussão

Observou-se que a alta diversidade nucleotídica nas regiões de gene (exon, intron e UTR) comparada às regiões intergênica para os receptores do gosto amargo esteve em contraste ao restante dos grupos de genes que mostraram o efeito oposto. Além disso, os genes não-*Tas2* apresentaram mais variação nas regiões flanqueantes que não são esperadas que sofressem pressão seletiva.

Dos 19 genes, o *GPR120* mostrou a diversidade nucleotídica mais baixa em região codificadora e não foram encontradas mutações não-sinônimas. Os javalis asiáticos e europeus divergiram há 1,2 milhões de anos (Groenen et al., 2012). Essa longa distância evolutiva resulta em dois grupos altamente diferenciados quando ambos suínos asiáticos e

européus são analisados usando, por exemplo, arrays de SNP de alta densidade ou filogenia com mtDNA (Groenen et al., 2012; Burgos-Paz et al., 2013). Vale salientar que a filogeografia dos TasRs afasta-se significativamente do padrão observado no genoma completo e, para esses receptores do gosto, a divergência autossomal Ásia-Europa é altamente atenuada (Figure 1A).

Assumindo que o padrão do genoma completo é principalmente o resultado de deriva, a menos que diferenciação esperada pode ser explicada por algum tipo de seleção balanceadora nos genes TR. No entanto, a seleção purificadora também parece ter desempenhado um papel na formação da diversidade de TASR, dada a prevalência de proporções entre diversidade nucleotídica não-sinônima e sinônima serem menores que 1 ( $\pi_a/\pi_s$ ) (Tabela 3). No entanto, Groenen et al. (2012) encontraram quatro genes de receptores do gosto (Tas1R2, Tas2R1, Tas240 e Tas2R39) sob seleção positiva (proporção dN/dS variou de 1.5 to 1.9). Vale destacar que algumas amostras usadas no presente estudo de suínos crioulos, brasileiros e ibéricos não foram amostradas no estudo desses autores, além disso, as análises deste estudo foram feitas baseando-se na proporção de polimorfismos não-sinônimos e sinônimos ( $\pi_a/\pi_s$ ) e não na proporção de divergência (dN/dS).

Análises de genomas completos (Bosse et al., 2012) mostraram uma diversidade nucleotídica mais alta em Ásia que em Europeus, como esperado devido ao “gargalo de garrafa” (bottleneck) sofrido pelos javalis europeus originários das raças suínas atuais quando migraram da Ásia. Nos asiáticos, uma diversidade nucleotídica reduzida nos domésticos versus javalis também foi observado por Bosse et al. 2012 e Groenen et al., 2012. Não foi observada essa redução na diversidade para os genes receptores do gosto nem quando foram comparados javalis asiáticos vs. europeus, nem entre suínos asiáticos domésticos e javalis.

A única população com uma redução marcada na diversidade foi a ibérica, e deve ser mencionado que a linhagem sequenciada nesse estudo pertence a uma população fechada (Guadyerbas) mantida geneticamente isolada desde 1945 (Toro et al., 2000). Como argumentado por Esteve-Codina et al. (2013), a consanguinidade devido ao confinamento explica muito da perda na variabilidade, enquanto que, as linhagens ibéricas possuem uma variabilidade comparável à encontrada nos javalis europeus atuais. As populações mais variáveis foram os suínos americanos (crioulos e brasileiros); essa conclusão aparentemente surpreendente pode ser explicada pela natureza miscigenada

dessas populações, uma vez que esses animais são o resultado de cruzamentos com muitos suínos de origens diferentes (Burgos-Paz et al. 2013).

#### **4. CONCLUSÕES**

A população genômica deste estudo mostrou que os genes de receptores do gosto amargo (Tas2Rs) apresentaram uma diversidade nucleotídica mais alta que os genes não-Tas2Rs (aminoácidos e ácidos graxos), sugerindo que o gosto amargo é uma característica plástica, possivelmente associada com a habilidade dos suínos adaptarem-se a vários ecossistemas e que pode ser passível de seleção em programas de melhoramento e conservação.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- AMARAL, A. J.; FERRETTI, L. et al. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. **PLoS One**. v. 6, n. 4, p.14782, 2011.
- ARCHIBALD, A. L.; BOLUND, L. et al. Pig genome sequence analysis and publication strategy. **BMC Genomics**. n. 11, p. 438, 2010.
- AXELSSON, E. et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. **Nature**. v. 495, n.7441, p. 360-364, 2013.
- BACHMANOV, A.A.; BEAUCHAMP, G.K: **Taste receptor genes**. *Annual review of nutrition*. n.27, p.389-414, 2007.
- BOSSE, M.; MEGENS, H-J.; MADSEN, O. et al. Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. **PLoS Genet**. v. 8, n.11, p.1003100, 2012.
- BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v.324, n.5926, p.522–528, 2009.
- BURGOS-PAZ, W.; SOUZA, C. A. et al. Porcine colonization of the Americas: a 60k SNP story. **Heredity**. v. 110, n. 4, p. 321-330, 2013.
- CONTE, C.; EBELING, M.; MARCUZ, A. et al. Identification and characterization of human taste receptor genes belonging to the TAS2R family. **Cytogenet Genome Res**. v. 98, n.1, p.45-53, 2002.
- DANECEK, P.; AUTON, A. et al. The variant call format and VCFtools. **Bioinformatics**. v.27, n.15, p. 2156-2158, 2011.
- DONG, D.; JONES, G.; ZHANG, S. Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evolutionary Biology*, n 9, 12p. 2009.
- EARL, D. A.; VON HOLDT, B.M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. **Conservation Genetics Resources**, 10.1007/s12686-011-9548-7, 2011.

- ESTEVE-CODINA, A.; PAUDEL, Y. et al. Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. **BMC Genomics**. v.14, p.148, 2013.
- EVANNO, G.; REGNAUT, S. et al. Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology**. v. 14, n.8, p. 2611-2620, 2005.
- FANG, X.; MOU, Y.; HUANG, Z. et al. The sequence and analysis of a Chinese pig genome. **GigaScience**. v.1, n.1, p.16, 2012.
- FERRETTI, L.; RAMOS-ONSINS, S. E. et al. Population genomics from pool sequencing. **Mol Ecol**. n.22, p.5561-76, 2013.
- FLICEK, P. A. M.; BARRELL, D.; BEAL, K. et al. ENSEMBL 2012. **Nucleic Acids Res**. 2012 Jan;40(Database issue):D8490. doi: 10.1093/nar/gkr991. 2013.
- GAUT, B. S. Molecular clocks and nucleotide substitution rates in higher plants. in M. K. Hecht, ed. *Evolutionary biology*. Plenum Press, p.93–120, New York, 1998.
- GROENEN, M. A.; ARCHIBALD, A. L. et al. Analyses of pig genomes provide insight into porcine demography and evolution. **Nature**. v. 491, n.7424, 393-398, 2012.
- International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution International Chicken Genome Sequencing Consortium. **Nature**, n.432, 695–716, 2004.
- KINNAMON, S. C.; CUMMINGS, T. A. Chemosensory transduction mechanisms in taste. *Annu. Rev. Physiol*. v. 54, p 715–731, 1992.
- KOSIOL, C. et al. Patterns of positive selection in six Mammalian genomes. **PLoS Genet** v. 4, n.8, p.1000144, 2008.
- LI, D.; ZHANG, J. Diet Shapes the Evolution of the Vertebrate Bitter Taste Receptor Gene Repertoire. **Mol Biol Evol.**, v. 31, p.303-309, 2014.
- LI, H.; HANDSAKER, B. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**. v.25, v.16, p. 2078-2079, 2009.
- LIBRADO, P.; ROZAS, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**. v. 25, n.11, p. 1451-1452, 2009.
- LINDBLAD-TOH, K.; WINCHESTER, E. et al. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. **Nat Genet**. v. 24, n.4, 381-386, 2000.
- MAKALOWSKI, W.; BOGUSKI, M. S. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. **Proc Natl Acad Sci U S A**. v. 95, n.16, 9407-9412, 1998.
- NG, P. C.; HENIKOFF, S. SIFT: predicting amino acid changes that affect protein function. **Nucleic Acids Res** 31(13): 3812-3814, 2003.

- PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**. v. 155, p. 945–959, 2000.
- PURCELL, S.; Neale, B. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am J Hum Genet**. v. 81, n.3, p. 559-575, 2007.
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3900051070, URL <http://www.Rproject.org>.
- RAMOS-ONSINS, S. E.; FERRETTI, L.; RAINERI, E. et al. (unpublished). mstatspop: Statistical Analysis using Multiple Populations to pipeline with ms: mstatspop beta version, release 0.998980b (20131103) (unpublished).
- ROSENBERG, M. S.; Subramanian, S. et al. "Patterns of transitional mutation biases within and among mammalian genomes. **Mol. Biol. Evol.** n.20, p. 988-993, 2003.
- Roura, E: **Taste beyond taste**. In: Australian Pig Science Association Conference (*APSA 2011*): 2011; Adelaide, Australia: Australasian Pig Science Association. p. 106-117. 2011.
- RUBIN, C. J.; MEGENS, H. J. et al. Strong signatures of selection in the domestic pig genome. **Proc Natl Acad Sci U S A**. v. 109, n.48, p.19529-19536, 2012.
- SHI, P.; ZHANG, J.; YANG, H. et al. Adaptive Diversification of Bitter Taste Receptor Genes in Mammalian Evolution. **Mol. Biol. Evol.** v. 20, n.5, p.805–814, 2003.
- TORO, M. A.; J. RODRIGANÉZ, et al. Genealogical analysis of a closed herd of Black Hairless Iberian pigs. **Conservation Biology**. v.14, n.6, p. 1843-1851, 2000.
- WELLENDORPH, P.; JOHANSEN, L.D.; BRAUNER-OSBORNE, H. Chapter Five - The Emerging Role of Promiscuous 7TM Receptors as Chemosensors for Food Intake. In: *Vitamins & Hormones*. Edited by Gerald L, Academic Press, v. 84, p. 151-184, 2010.
- WONG, G.T.; GANNON, K.S.; MARGOLSKEE, R.F. Transduction of bitter and sweet taste by gustducin. *Nature*, v. 381, n 6585, p.796-800, 1996.
- YANG, S.; ZHANG, H. et al. The Local Origin of the Tibetan Pig and Additional Insights into the Origin of Asian Pigs. **PLoS One** 6(12): e28215. 2011.
- YU, G.; XIANG, H.; WANG, J.; ZHAO, X. The phylogenetic status of typical Chinese native pigs: analyzed by Asian and European pig mitochondrial genome sequences. *Journal of animal science and biotechnology*. v. 4, n.1, p.9, 2013.
- ZHANG, Z.; Li, J. et al. Computing Ka and Ks with a consideration of unequal transitional substitutions. **BMC Evol Biol**. n. 6, p.44, 2006.



## CAPÍTULO 5

### CONCLUSÕES

1. As análises filogeográficas utilizando os marcadores microssatélites revelaram um baixo grau de estruturação e divergência dos suínos monteiros distribuídos em subpopulações separadas por uma distância de até 100 km. A filogeografia comparada revelou forte relação do haplótipo mitocondrial pertencente aos suínos monteiros com várias populações suínas distribuídas pelo Brasil, principalmente outras raças suínas brasileiras que, provavelmente, devem compartilhar uma história comum sobre suas origens. Os resultados obtidos com esse estudo demonstraram que o padrão de diversidade genética e a relação genética entre os suínos monteiros e as demais raças suínas não pode ser apenas de interesse para a biologia evolutiva, mas também muito importante para a conservação dessa raça, uma vez que foi destacada a grande importância do fluxo gênico para evitar a fragmentação dessa população. Além disso, medidas de controle genético devem ser empregadas para evitar o aumento de consanguinidade que pode levar a problemas de depressão endogâmica e comprometer a sobrevivência e capacidade de adaptação dos animais em seu *habitat* comum no Brasil que é o Pantanal Mato-grossense.
2. A análise filogenética usando sequências completas e da região controle do mtDNA separou adequadamente as amostras de suínos e javalis de acordo com suas origens maternas sem contradições e foi muito informativa na compreensão da relação genética entre as raças amostradas nesse estudo. Nas análises de diversidade genética, os javalis europeus apresentaram uma diversidade nucleotídica nas sequências completas do mtDNA muito inferior ( $0,08 \pm 0,01$ ) as obtidas com as sequências da região controle ( $\% \pi = 1,33 \pm 0,18$ ). Porém, em geral, as diversidades nucleotídicas e haplotípicas calculadas, com sequências completas ou com as

sequências da região controle, foram altas nos javalis de origem asiática e europeia. Esses resultados são esperados considerando que os javalis são um ponto de origem dos suínos domésticos. A alta diversidade nucleotídica nos genes codificadores de proteína do mtDNA aumentou o poder estatístico no momento de detectar genes sujeitos a seleção positiva.

3. O estudo de genes de receptores do gosto e de nutrientes, além de contribuir com um conjunto de novos SNPs para os bancos de dados públicos que podem contribuir ainda mais para a compreensão do processo evolutivo desses genes em suínos forneceu evidências de que a alta diversidade para o *cluster* de genes de receptores para o gosto amargo pode estar relacionada à adaptação dietética, especialmente para os componentes vegetais. Curiosamente, também foi verificada uma divergência muito menos acentuada entre haplótipos asiáticos e europeus do que a observada com marcadores distribuídos por todo o genoma; que, juntamente com a elevada variabilidade, pode ser um indicativo de seleção balanceadora nestes *loci*.

## ANEXOS

## ANEXO A: CAPÍTULO 2

## TABELAS

Tabela A1 Quantidades de reagentes utilizados nas reações de PCR e de sequenciamento.

<sup>1</sup> Reagentes	Volume (uL)	<sup>2</sup> Reagentes	Volume (uL)	<sup>3</sup> Reagentes	Volume (uL)
Água ultrapura	0,45/0,4	Tampão 10x	2,0	Big Dye	0,5
<i>Primer</i> (10 uM)	0,05/0,1	MgCl <sub>2</sub> (25	1,2	Tampão 5x	1,75
Master mix	2,5	dNTP (2,5	1,6	<i>Primer</i> R/F	2,0
Q. Solution	0,5	<i>Primer</i> R	1,0	Água	4,25
DNA (3ng/uL)	1,5	<i>Primer</i> F	1,0	DNA (9ng)	1,5
-	-	Taq (1U)	0,2	-	-
-	-	Água	10	-	-
-	-	DNA	3,0	-	-
Total	5		20		10

<sup>1</sup> PCR microssatélites, <sup>2</sup> PCR *D-Loop* e <sup>3</sup> reação de sequenciamento do *D-loop*.

Tabela A2 Parâmetros de diversidade genética analisados em 19 locos de microssatélites nos suínos monteiros do Pantanal.

Locos	N	Na	Nam	cNae	Ho	cHe	F <sub>IS</sub>	EWE
S0026	180	6	4,400	3,498	0,804	0,745	-0,126	0,028*
S0155	181	5	3,000	1,597	0,358	0,393	0,043	0,000***
SW830	179	5	4,500	3,224	0,708	0,723	-0,027	0,441 <sup>ns</sup>
SW1517	180	9	6,300	4,041	0,769	0,789	-0,022	0,001**
S0002	175	5	4,000	2,756	0,628	0,669	0,014	0,036*
S0355	177	5	4,100	2,081	0,528	0,545	-0,016	0,920 <sup>ns</sup>
SW936	178	8	5,800	3,130	0,655	0,716	0,038	0,108 <sup>ns</sup>
SW2406	175	8	4,000	1,399	0,175	0,305	0,388	0,020*
OPN	180	8	5,600	3,179	0,659	0,720	0,039	0,334 <sup>ns</sup>
S0005	178	8	5,000	3,304	0,724	0,730	-0,038	0,007**
S0101	181	6	3,900	2,646	0,619	0,653	0,006	0,001***
S0068	176	9	6,600	4,733	0,855	0,824	-0,084	0,477 <sup>ns</sup>
S0090	174	6	3,600	2,142	0,519	0,561	0,027	0,014*
SW72	181	4	2,900	2,229	0,515	0,580	0,065	0,484 <sup>ns</sup>
SW455	180	11	5,400	3,557	0,740	0,753	-0,030	0,343 <sup>ns</sup>
SW911	180	7	4,700	2,713	0,603	0,664	0,045	0,261 <sup>ns</sup>
S0097	178	8	5,800	3,599	0,720	0,758	0,003	0,007**
S0228	178	5	4,500	3,696	0,666	0,768	0,087	0,464 <sup>ns</sup>
SW24	175	6	4,800	3,296	0,714	0,730	-0,025	0,925 <sup>ns</sup>
Média±SE	178,21	6,789±1,843	4,679±0,239	2,990±0,193	0,629±0,037	0,664±0,031	0,006±0,016	-
IC = 95%	-	-	4,303-5,192	2,651-3,329	0,551-0,682	0,606-0,715	-0,023-0,033	-

N = número de indivíduos tipados ; Nam = número médio de alelos; cNae = número efetivo de alelos corrigidos ; Ho = heterozigiosidade observada; cHe = heterozigiosidade esperada corrigida segundo o EHW; F<sub>IS</sub> = índice de consanguinidade intra-populacional; EWE = Equilíbrio de Hardy-Weinberg; \* = P<0,05 ; \*\* = P<0,01 ; \*\*\* = P<0,001 ; ns = não significativo ; SE = erro padrão; IC = intervalo de confiança.

Tabela A3 Frequência (Freq) dos alelos diagnósticos (AD) por população e locos.

Populações	Locos	AD	Freq.	Populações	Locos	AD	Freq.
MOB	S0155	6	0,063	MOB	S0090	1	0,143
MOB	S0155	7	0,063	MOB	SW72	7	0,063
MOB	SW1517	7	0,313	MOB	SW72	10	0,063
MOB	S0002	1	0,063	MOB	SW24	13	0,071
MOB	SW936	3	0,125	Pop3	SW1517	9	0,063
MOB	SW2406	15	0,063	Pop3	S0005	1	0,031
MOB	S0005	15	0,188	Pop8	SW936	2	0,014
MOB	S0005	20	0,063	Pop8	S0005	13	0,014
MOB	S0005	22	0,063	Pop10	SW72	5	0,083
MOB	S0068	10	0,063	Pop10	S0097	10	0,028
MOB	S0068	12	0,125	-	-	-	-

Tabela A4 Designação dos suínos monteiros à população de origem.

Populações	População de origem	Outras populações
MOB	5	3
Pop1	-	5
Pop2	3	13
Pop3	4	12
Pop4	3	15
Pop5	8	18
Pop6	5	5
Pop7	10	25
Pop8	-	4
Pop9	3	12
Pop10	10	26
Total	51	138
Percentagem	27%	73%

Tabela A5 Matriz de distância geográfica (km) entre os 10 pontos de coletas representados pelas populações (Pop) amostradas no Pantanal-MS.

	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	Pop9	Pop10
Pop1	-									
Pop2	68,021	-								
Pop3	38,456	32,669	-							
Pop4	93,513	33,363	55,057	-						
Pop5	53,209	15,037	18,027	43,372	-					
Pop6	62,497	29,241	25,465	33,721	24,775	-				
Pop7	67,232	28,577	45,144	60,149	31,115	54,817	-			
Pop8	78,627	10,742	43,319	29,203	25,778	36,253	31,205	-		
Pop9	52,542	38,951	40,791	72,306	33,785	58,358	18,218	45,265	-	
Pop10	53,343	17,281	15,863	41,041	5,779	19,157	36,807	27,699	39,219	-
Média	63,048	30,431	34,976	51,302	27,875	38,253	41,473	36,454	44,381	28,465

Tabela A6 Estimativas de  $F_{ST}$  par a par abaixo na diagonal e valores de P acima na diagonal com os 19 locos de microssatélites.

	MOB	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	Pop9	Pop10
MOB	-	0,030*	0,003*	0,000*	0,000*	0,000*	0,002*	0,000*	0,043*	0,000*	0,000*
Pop1	0,080	-	0,025*	0,001*	0,067	0,006*	0,404	0,000*	0,594	0,018*	0,009*
Pop2	0,055	0,055	-	0,640	0,256	0,128	0,001*	0,464	0,957	0,353	0,420
Pop3	0,069	0,063	0,018	-	0,118	0,020*	0,004*	0,277	0,837	0,023*	0,055
Pop4	0,067	0,045	0,019	0,019	-	0,181	0,119	0,502	0,771	0,076	0,040*
Pop5	0,066	0,048	0,016	0,019	0,014	-	0,003*	0,453	0,794	0,139	0,009*
Pop6	0,070	0,044	0,040	0,034	0,024	0,029	-	0,000*	0,512	0,001*	0,000*
Pop7	0,066	0,054	0,013	0,013	0,011	0,009	0,029	-	0,828	0,268	0,003*
Pop8	0,088	0,062	0,031	0,033	0,035	0,030	0,048	0,027	-	0,646	0,820
Pop9	0,073	0,052	0,020	0,025	0,022	0,018	0,043	0,014	0,033	-	0,003*
Pop10	0,057	0,048	0,014	0,018	0,016	0,014	0,036	0,014	0,031	0,024	-

\* $P < 0,05$ ; 10100 permutações

Tabela A7 Distâncias genéticas D de Nei (Nei, 1972) com os 19 locos de microsatélites.

	MOB	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	Pop9	Pop10
MOB	-										
Pop1	0,403	-									
Pop2	0,280	0,224	-								
Pop3	0,349	0,250	0,064	-							
Pop4	0,369	0,176	0,074	0,072	-						
Pop5	0,370	0,189	0,066	0,071	0,055	-					
Pop6	0,406	0,167	0,170	0,138	0,102	0,125	-				
Pop7	0,351	0,214	0,049	0,047	0,041	0,034	0,120	-			
Pop8	0,437	0,212	0,103	0,112	0,121	0,106	0,177	0,094	-		
Pop9	0,367	0,190	0,069	0,083	0,075	0,062	0,171	0,047	0,119	-	
Pop10	0,299	0,200	0,052	0,065	0,062	0,055	0,153	0,051	0,105	0,083	-

Tabela A8 Proporção de cada população designada à cada cluster para K2 e K4.

Pop	K=2		K=4			
	1	2	1	2	3	4
MOB	0,7588	0,2412	0,5146	0,08	0,2032	0,202
Pop1	0,1202	0,8798	0,0704	0,2558	0,2034	0,4702
Pop2	0,1898	0,8102	0,207	0,287	0,2344	0,2714
Pop3	0,0594	0,9406	0,1004	0,3674	0,2922	0,24
Pop4	0,0146	0,9854	0,0778	0,3122	0,245	0,3654
Pop5	0,017	0,983	0,1118	0,3358	0,2328	0,3194
Pop6	0,0082	0,9918	0,0536	0,2622	0,2138	0,4704
Pop7	0,0474	0,9526	0,1218	0,3644	0,2532	0,261
Pop8	0,0184	0,9816	0,1164	0,3546	0,2532	0,2762
Pop9	0,0982	0,9018	0,1786	0,355	0,2206	0,2454
Pop10	0,1486	0,8514	0,1354	0,3228	0,2714	0,27



Tabela A9 Índices de diversidade genética para as sequências da região controle do mtDNA de suínos de cada população.

Estatísticas	Monteiros_Pantanal	Monteiros_DF	Suínos_locais	Suínos_Mestiços	Suínos_comerciais	Total
Tamanho da amostra	122	6	168	6	29	331
S	10	2	19	10	17	21
$\pi$	0,00045±0,00044	0,00184±0,00119	0,01149±0,00157	0,01662±0,00358	0,01874±0,00186	0,00797±0,00093
$h$	0,016±0,016	0,333±0,215	0,693±0,693	0,600±0,129	0,884±0,030	0,483±0,041

S = sítios com substituição;  $\pi$  = diversidade nucleotídica;  $h$  =diversidade haplotípica

Tabela A10 Índices de diversidade para as sequências da região controle do mtDNA de suínos para as quatro regiões fisiográficas amostradas.

Estatísticas	Centro-oeste	Nordeste	Sudeste	Sul
Tamanho da amostra	182	71	36	42
S	6	15	14	17
$\pi$	0,02018±0,0029	0,01934±0,0018	0,01897±0,0034	0,01736±0,0041
$h$	1,000±0,063	0,964±0,051	1,000± 0,096	1,000±0,045

S = sítios com substituição;  $\pi$  = diversidade nucleotídica;  $h$  =diversidade haplotípica

Tabela A11 Estimativa de fluxo gênico baseado em  $F_{ST}$  pelo método pairwise entre as cinco populações deste estudo abaixo na diagonal. Os valores de P estão acima na diagonal e os asteriscos indicam valores significativos ( $P < 0,05$ ).

	SMT_P	SMT_DF	LOCAIS	MESTIÇOS	COMERCIAIS
SMT_P	-	0,09910	0,00000*	0,00000*	0,00000*
SMT_DF	0,16288	-	0,28829	0,20721	0,02703*
LOCAIS	0,20515	0,00984	-	0,01802	0,00000*
MESTIÇOS	0,85847	0,37714	0,19150	-	0,14414
COMERCIAIS	0,58517	0,16743	0,11429	0,10069	-

Tabela A12 Estimativa de fluxo gênico baseado em  $F_{ST}$  pelo método pairwise entre as quatro regiões fisiográficas do Brasil abaixo na diagonal. Os valores de P estão acima na diagonal e os asteriscos indicam valores significativos ( $P < 0,05$ ). Análise realizada com a região controle do mtDNA

	CO	NE	SE	S
CO	-	0,0000*	0,0000*	0,0000*
NE	0,21652	-	0,0000*	0,0090*
SE	0,40966	0,08714	-	0,0000*
S	0,41541	0,07985	0,22873	-

CO = Centro-Oeste; NE = Nordeste; SE = Sudeste; S = Sul

## FIGURAS

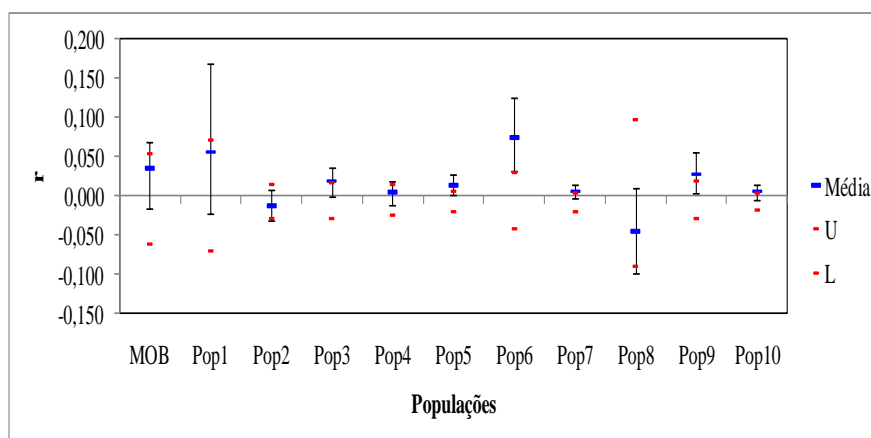


Figura A1 Teste de parentesco individual (relatedness) para as onze populações de Monteiros deste estudo baseado em dados de microssatélites.

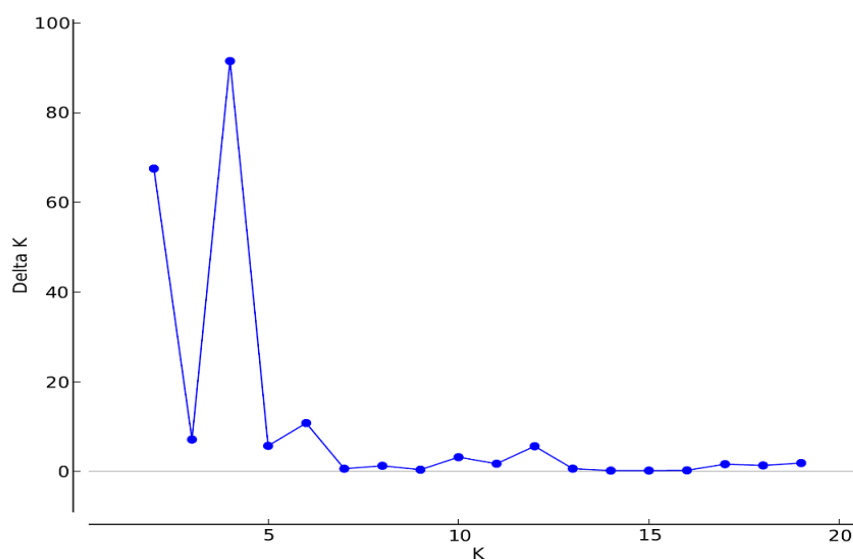


Figura A2 Estatística Delta K baseada no método de clusterização de Evanno et al. (2005).

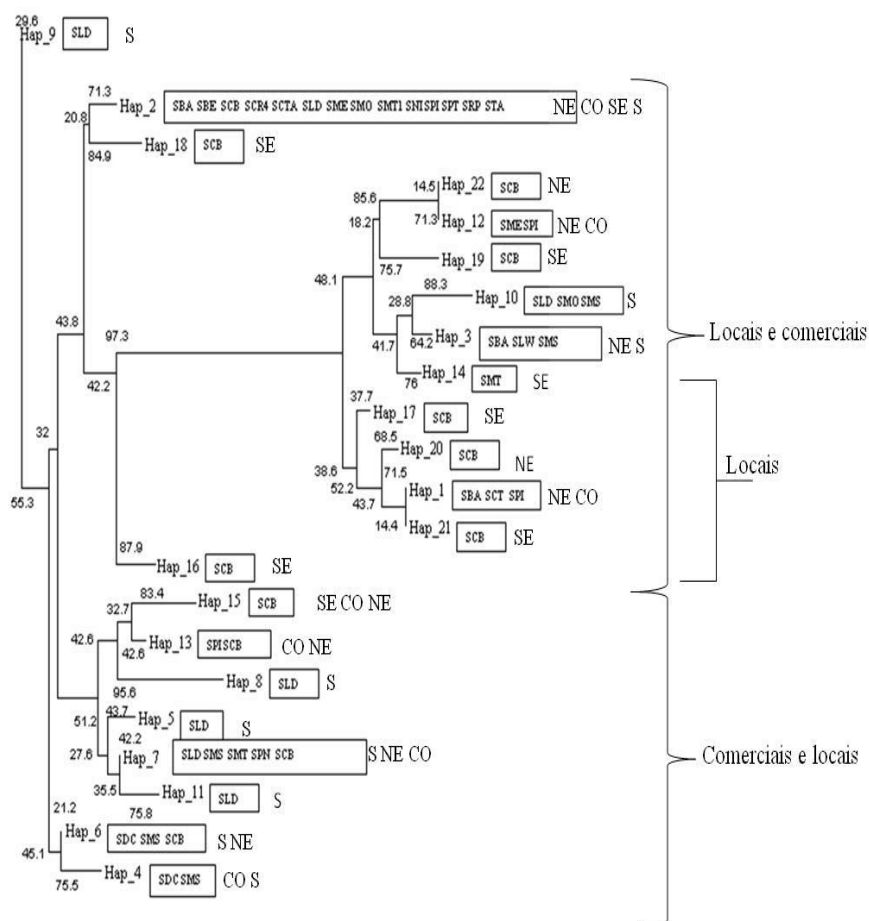


Figura A3 Relação genética e geográfica entre os 22 haplótipos retratada pela árvore construída pelo método Neighbour-Joining.

## ANEXO B: CAPÍTULO 3

## TABELAS

Tabela B13 Valores de distância genética e erro padrão ( $d \pm SE$ ) entre populações obtidos com o DNA mitocondrial completo (abaixo na diagonal) e a região controle (acima na diagonal).

	AME	EUD	EUD1	ASD	EUWB	ASWB
AME	-	0,009 $\pm$ 0,002	0,006 $\pm$ 0,002	0,015 $\pm$ 0,003	0,007 $\pm$ 0,002	0,025 $\pm$ 0,005
EUD	0,006 $\pm$ 0,000	-	-	0,015 $\pm$ 0,003	0,010 $\pm$ 0,002	0,024 $\pm$ 0,005
EUD1	0,004 $\pm$ 0,000	-	-	0,015 $\pm$ 0,004	0,007 $\pm$ 0,002	0,026 $\pm$ 0,005
ASD	0,010 $\pm$ 0,001	0,009 $\pm$ 0,001	0,012 $\pm$ 0,001	-	0,021 $\pm$ 0,004	0,016 $\pm$ 0,003
EUWB	0,003 $\pm$ 0,000	0,005 $\pm$ 0,000	0,001 $\pm$ 0,000	0,012 $\pm$ 0,001	-	0,026 $\pm$ 0,005
ASWB	0,013 $\pm$ 0,001	0,012 $\pm$ 0,001	0,015 $\pm$ 0,001	0,007 $\pm$ 0,000	0,014 $\pm$ 0,001	-

AME- Americanos; EUD=Domésticos europeus; 1Europeus sem haplótipos asiáticos; ASD=Domésticos asiáticos; EUWB=Javalis europeus e Javalis asiáticos.

Tabela B14 Posição de cada mutação não-sinônima deletéria e frequência nas populações suínas. Essa frequência não incluiu mutações polimórficas ou fixadas no outgroup.

Posição da mutação	Genes	Frequência por população					
		AME	EUD	EUD1	ASD	EUWB	ASWB
MT:4087	NADH1	0,000	0,036	0,053	0,000	0,000	0,000
MT:4112		0,000	0,036	0,053	0,000	0,000	0,000
MT:4154		0,000	0,036	0,053	0,000	0,000	0,000
MT:4180		0,000	0,000	0,000	0,000	0,214	0,182
MT:5298	NADH2	0,000	0,036	0,053	0,000	0,000	0,000
MT:5391		0,000	0,036	0,053	0,000	0,000	0,000
MT:5643		0,000	0,036	0,053	0,000	0,000	0,000
MT:5609		0,000	0,000	0,000	0,000	0,000	0,091
MT:5330		0,000	0,000	0,000	0,037	0,000	0,000
MT:5718		0,214	0,321	0,000	1,000	0,000	0,364
MT:7081	COX1	0,000	0,071	0,105	0,000	0,000	0,000
MT:7154		0,000	0,036	0,000	0,000	0,000	0,000
MT:7173		0,000	0,036	0,000	0,000	0,000	0,000
MT:7181		0,000	0,036	0,000	0,000	0,000	0,000
MT:7184		0,000	0,036	0,000	0,000	0,000	0,000
MT:7189		0,000	0,036	0,000	0,000	0,000	0,000
MT:7262		0,000	0,036	0,000	0,000	0,000	0,000
MT:7418		0,000	0,036	0,053	0,000	0,000	0,000
MT:7532		0,000	0,036	0,000	0,000	0,000	0,000
MT:8203	COX2	0,000	0,000	0,000	0,037	0,000	0,000
MT:8301		0,000	0,036	0,000	0,000	0,000	0,000
MT:8342		0,000	0,000	0,000	0,000	0,000	0,182
MT:8389		0,000	0,036	0,000	0,000	0,000	0,000
MT:8819		0,000	0,036	0,000	0,000	0,000	0,000
MT:8959	ATPase8	0,000	0,000	0,000	0,074	0,000	0,000
MT:9043		0,000	0,036	0,000	0,000	0,000	0,000
MT:9076		0,000	0,000	0,000	0,000	0,000	0,182
MT:9151		0,000	0,000	0,000	0,000	0,000	0,091
MT:9354		0,000	0,000	0,000	0,000	0,000	0,091
MT:9370	ATPase6	0,000	0,071	0,000	0,000	0,000	0,000
MT:9406		0,000	0,000	0,000	0,000	0,000	0,091
MT:9442		0,000	0,036	0,053	0,000	0,000	0,000
MT:9445		0,000	0,036	0,053	0,000	0,000	0,000
MT:9882	COX3	0,000	0,000	0,000	0,000	0,071	0,000

MT:9989		0,000	0,036	0,000	0,000	0,000	0,000
MT:9996		0,000	0,036	0,000	0,000	0,000	0,000
MT:10071		0,000	0,000	0,000	0,037	0,000	0,000
MT:10271		0,143	0,000	0,000	0,000	0,000	0,000
MT:10355		0,000	0,036	0,053	0,000	0,000	0,000
MT:10424		0,000	0,000	0,000	0,037	0,000	0,000
MT:10692	NADH3	0,000	0,036	0,000	0,000	0,000	0,000
MT:11279	NADH4L	0,000	0,000	0,000	0,000	0,000	0,182
MT:11300		0,000	0,071	0,000	0,000	0,000	0,000
MT:11368	NADH4	0,000	0,000	0,000	0,000	0,214	0,000
MT:11458		0,000	0,000	0,000	0,000	0,214	0,000
MT:11744		0,000	0,036	0,053	0,000	0,000	0,000
MT:11801		0,000	0,036	0,053	0,000	0,000	0,000
MT:12055		0,000	0,000	0,000	0,000	0,000	0,091
MT:12194		0,000	0,036	0,000	0,000	0,000	0,000
MT:12269		0,000	0,036	0,000	0,000	0,000	0,000
MT:12349		0,000	0,036	0,000	0,000	0,000	0,000
MT:12524		0,000	0,036	0,053	0,000	0,000	0,000
MT:13643	NADH5	0,000	0,036	0,053	0,000	0,000	0,000
MT:13761		0,000	0,036	0,000	0,000	0,000	0,000
MT:14070		0,000	0,036	0,053	0,000	0,000	0,000
MT:14084		0,000	0,036	0,053	0,000	0,000	0,000
MT:14180		0,000	0,036	0,053	0,000	0,000	0,000
MT:14366		0,000	0,000	0,000	0,037	0,000	0,000
MT:14694		0,000	0,036	0,000	0,000	0,000	0,000
MT:14999	NADH6	1,000	1,000	1,000	1,000	1,000	0,818
MT:15130		0,000	0,036	0,053	0,000	0,000	0,000
MT:15254		0,000	0,071	0,105	0,000	0,000	0,000
MT:15400	Cytb	0,000	0,036	0,000	0,000	0,000	0,000
MT:15474		0,000	0,000	0,000	0,000	0,000	0,273
MT:15830		0,000	0,036	0,000	0,000	0,000	0,000
MT:15933		0,000	0,036	0,053	0,000	0,000	0,000
MT:16140		0,000	0,036	0,053	0,000	0,000	0,000

---

Tabela B15 D de Tajima para mutações sinônimas (Ts) e não-sinônimas (Ta) por gene e por população. Valores em negrito ou sublinhado foram significativos a 5% e 1%, respectivamente.

Genes	Ts	Ta	Ts	Ta	Ts	Ta	Ts	Ta	Ts	Ta	Ts	Ta
	AME		EUD		EUD1		ASD		EUWB		ASWB	
NADH1	0,349	0,416	1,223	-1,239	-0,590	<b>-1,966</b>	-1,233	n,a,	n,a,	-0,565	0,009	0,223
NADH2	0,399	0,319	1,274	0,005	-1,165	-1,661	-2,146	-2,004	-0,532	-0,959	0,208	0,113
COX1	0,632	-1,155	0,557	<u>-2,283</u>	<b>-1,849</b>	-1,096	-1,815	-1,153	-1,155	n,a,	0,280	-0,127
COX2	0,578	n,a,	0,831	-1,733	n,a	n,a	n,a,	-1,153	n,a,	0,324	-0,043	-0,142
ATPase8	0,469	0,416	1,825	0,050	n,a	n,a	-1,733	-1,520	n,a,	-1,155	0,362	-0,812
ATPase6	0,314	-0,337	0,441	-0,581	-0,607	-1,715	-1,520	-1,307	0,184	1,212	-0,155	-1,053
COX3	-0,023	-0,011	0,183	-1,139	-1,050	-1,511	-1,233	-1,512	-0,013	-1,155	n,a	n,a
NADH3	0,416	0,505	-0,341	-0,150	-0,562	-1,511	-0,728	-1,153	-1,155	-1,481	n,a	n,a
NADH4L	n,a,	0,290	-1,151	1,4558	-1,165	-0,562	n,a,	-0,394	0,324	n,a,	-0,100	-0,129
NADH4	-0,244	0,327	-1,085	0,154	<b>-1,861</b>	<b>-1,861</b>	n,a,	-1,560	-1,155	-0,921	-0,142	-0,207
NADH5	0,184	-0,056	0,357	-1,297	<b>-1,841</b>	<b>-2,046</b>	-1,396	-1,718	-0,877	-1,670	-0,058	-0,687
NADH6	n,a,	0,324	-0,741	1,310	-0,562	-1,165	n,a,	-0,932	n,a	n,a	-0,778	-0,002
cytb	0,302	0,505	0,328	-0,737	-1,715	-1,762	-1,709	-1,555	-1,670	-1,155	0,095	0,423

## FIGURAS

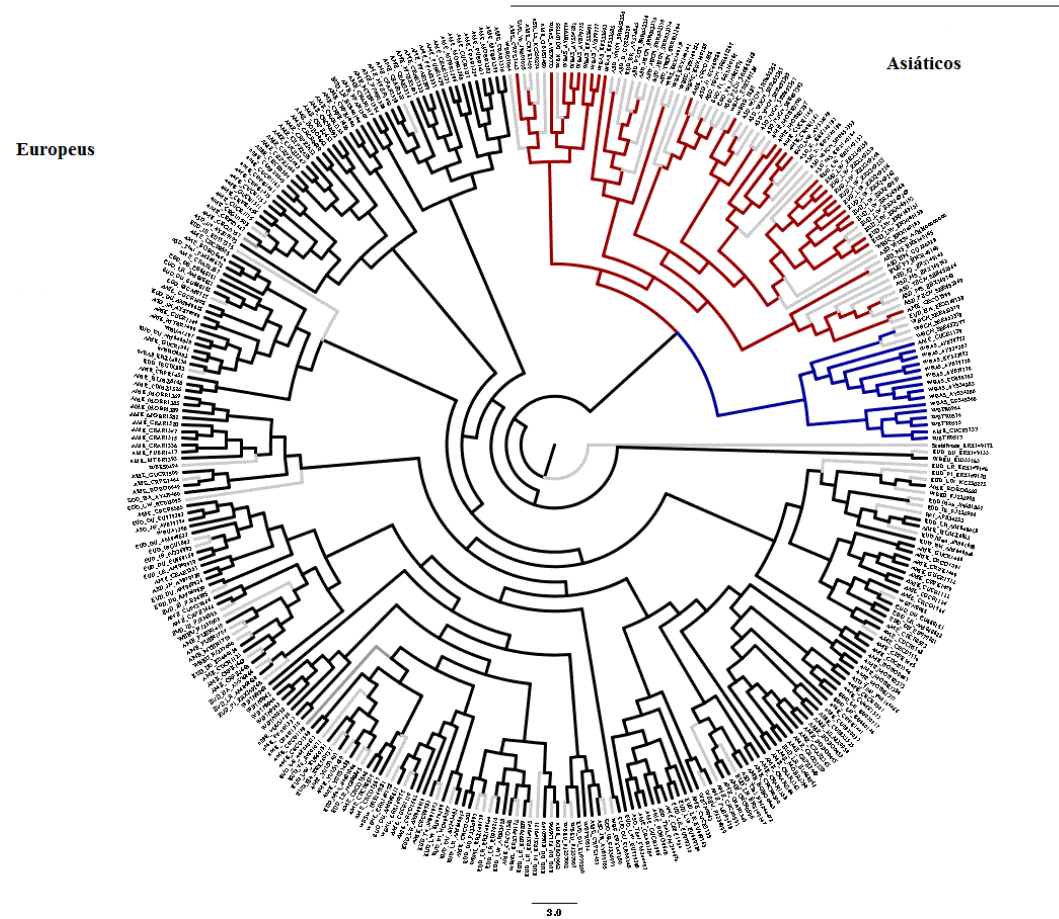


Figura B4 Árvore consenso Bayesianas de 341 seqüências da região controle do mtDNA. As cores dos ramos indicam designações diferentes dos cladros (preto são europeus e vermelho e azul são ASD e ASWB, respectivamente). Cinza claro representa as seqüências usadas para o estudo de mtDNA completo.



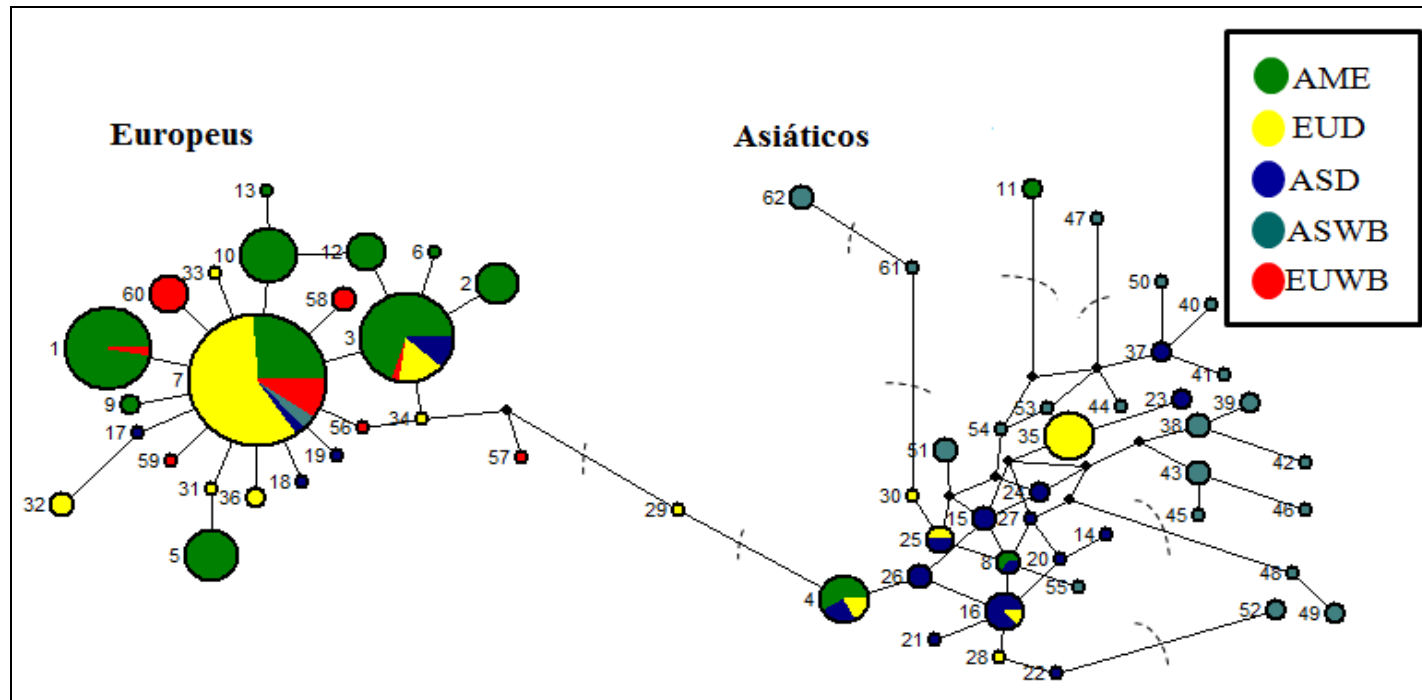


Figura B5 Relação entre 62 haplótipos pelo método de network Median-Joining usando 575 pb da região controle do mtDNA. Os diferentes círculos coloridos correspondem às populações e os tamanhos proporcionais às frequências das amostras. Pontos pretos pequenos são vetores medianos que representam haplótipos inferidos e linhas tracejadas representam o número de mutações. Um ASWB-TR0514 da Turquia sequenciado nesse estudo apresentou um haplótipo europeu, sugerindo evento de introgressão ou falta de informação genética para inferir haplótipos de um fragmento de mtDNA pequeno. Treze haplótipos foram relacionados aos suínos AME, dos quais cinco e oito foram compartilhados e exclusivos, respectivamente. Dos haplótipos AME compartilhados, o H3 foi, principalmente, compartilhado com sete EUD, um EUWB da França e com cinco ASD (cinco suínos tailandeses que têm haplótipo europeu). O H1 conteve apenas um suíno ibérico (EUD\_IB\_FJ236994), e o H7 foi compartilhado entre as cinco populações.

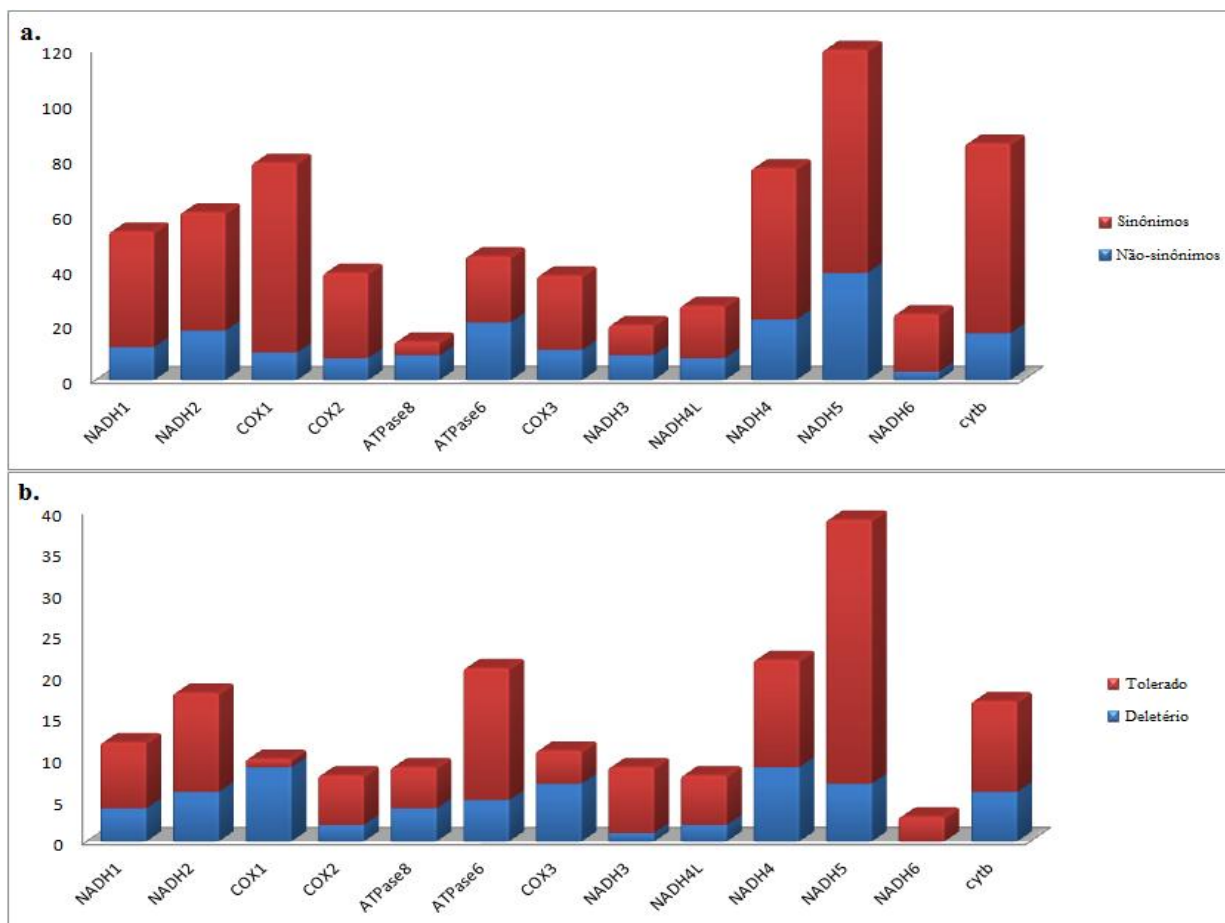


Figura B6 Número de sítios sinônimos e não-sinônimos (a) ocorrendo em cada um dos 13 genes codificadores de proteína. Distribuição das mutações não-sinônimas de cada gene em tolerado e deletério que foram determinado por escores SIFT (b).

## ANEXO C: CAPÍTULO 4

## TABELAS

Tabela C16 Números de acesso do GenBank das sequências incluídas nas análises.

Amostra	Origem	Número GenBank	Amostra	Origem	Número GenBank
LW22F01_Large White	na	ERX149148	PI21M20_Pietrain	na	ERX149170
LW22F02_Large White	na	ERX149149	PI21M21_Pietrain	na	ERX149171
LW22F03_Large White	na	ERX149150	HA20U01_Hampshire	na	ERX149137
LW22F04_Large White	na	ERX149151	HA20U02_Hampshire	na	ERX149138
LW22F06_Large White	na	ERX149152	JQ01U02_Jiangquhai	China	ERX149142
LW22F07_Large White	na	ERX149153	MS20U10_Meishan	China	ERX149162
LW22M04_Large White	na	ERX149154	MS20U11_Meishan	China	ERX149163
LW22M07_Large White	na	ERX149155	MS21M07_Meishan	China	ERX149164
LW36F01_Large White	na	ERX149156	MS21M14_Meishan	China	ERX149165
LW36F02_Large White	na	ERX149157	WB20U02_Wild Boar	Japão	ERX149175
LW36F03_Large White	na	ERX149158	WB21F05_Wild Boar	Holanda	ERX149176
LW36F04_Large White	na	ERX149159	WB21M03_Wild Boar	Holanda	ERX149177
LW36F05_Large White	na	ERX149160	WB22F01_Wild Boar	Holanda	ERX149178
LW36F06_Large White	na	ERX149161	WB22F02_Wild Boar	Holanda	ERX149179
LR21M03_Landrace	na	ERX149143	WB25U11_Wild Boar	França	ERX149180
LR24F01_Landrace	na	ERX149144	WB26M09_Wild Boar	Suécia	ERX149181
LR24F08_Landrace	na	ERX149145	WB29U04_Wild Boar	Sul da China	ERX149182
LR30F02_Landrace	na	ERX149146	WB29U12_Wild Boar	Sul da China	ERX149183
LR30F03_Landrace	na	ERX149147	WB30U01_Wild Boar	Norte da China	ERX149184
DU23M01_Duroc	na	ERX149133	WB30U08_Wild Boar	Norte da China	ERX149185
DU23M02_Duroc	na	ERX149134	WUCN1800- _Wuzhishan	China	AJKK00000000
DU23M03_Duroc	na	ERX149135	XI01U03_Xiang	China	ERX149186
DU23M04_Duroc	na	ERX149136	XI01U04_Xiang	China	ERX149187
PI21F02_Pietrain	na	ERX149167	INDO22_Wild Boar	Sumatra	ERX149139
PI21F06_Pietrain	na	ERX149168	INDO33_Wild Boar	Sumatra	ERX149141
PI21M17_Pietrain	na	ERX149169			

Na= não disponível

Tabela C17 Diversidade nucleotídica total ( $\pi_t$ ) e região do gene ( $\pi_g$ ) por população.

Grupo de genes	Genes	Internacional (n=31)		Ibéricos (n=04)		Crioulo (n=14)		Brasileiros (n=03)		Asiáticos (n=08)		ASWB (n=06)		EUWB (n=09)	
		$\pi_t$	$\pi_g$	$\pi_t$	$\pi_g$	$\pi_t$	$\pi_g$	$\pi_t$	$\pi_g$	$\pi_t$	$\pi_g$	$\pi_t$	$\pi_g$	$\pi_t$	$\pi_g$
Amargo	TAS2R20	1,6±0,6	2,6±0,2	0,3±0,0	0,3±0,0	2,7±0,3	3,8±0,5	2,4±0,5	3,5±0,6	0,3±0,0	0,6±0,0	2,5±0,7	0,7±0,1	2,7±0,4	3,9±1,6
	TAS2R9	2,4±0,5	1,1±0,3	0,7±0,0	0,5±0,1	4,4±1,4	1,1±0,0	2,6±0,6	1,7±0,2	3,7±0,6	1,3±0,5	4,4±0,6	1,5±0,4	1,6±0,3	0,6±0,1
	TAS2R10	1,9±0,3	0,5±0,0	1,5±0,5	1,7±0,5	3,3±0,9	1,1±0,2	2,3±0,3	1,1±0,1	2,4±0,2	0,1±0,1	4,6±1,8	1,5±0,4	1,1±0,1	1,1±0,1
	TAS2R42	1,9±0,3	4,4±0,6	0,3±0,0	1,0±0,2	3,1±0,6	6,3±2,4	2,5±0,6	6,4±2,9	1,8±0,3	4,4±1,9	2,3±0,3	4,2±1,2	1,9±0,3	4,9±0,7
	TAS2R16	2,1±0,3	1,5±0,3	2,2±0,3	1,5±0,4	1,6±0,2	1,3±0,1	7,8±1,8	6,0±2,1	2,2±0,2	1,7±0,5	3,2±0,5	1,0±0,1	6,2±0,7	4,4±1,4
	TAS2R38	0,8±0,1	5,9±0,7	0,8±0,1	5,6±0,9	0,9±0,1	5,2±0,6	0,9±0,1	3,7±1,0	0,9±0,0	4,6±0,6	0,7±0,0	1,5±0,3	0,8±0,0	6,3±2,6
	TAS2R39	2,8±0,5	4,7±1,8	1,7±0,2	3,1±0,8	3,4±0,5	5,4±0,8	4,1±0,3	6,9±1,8	3,9±0,9	5,5±1,9	3,5±1,1	6,5±1,5	2,0±0,2	2,2±0,4
	TAS2R41	0,5±0,1	2,4±0,5	0,6±0,1	2,9±0,5	1,0±0,1	3,2±0,8	0,5±0,1	2,0±0,0	0,8±0,0	3,5±0,5	0,8±0,1	4,3±1,6	0,8±0,0	2,6±0,6
	TAS2R60	1,0±0,2	3,6±0,3	1,4±0,3	3,3±0,3	1,5±0,5	3,7±1,8	0,9±0,2	2,3±0,3	1,0±0,0	3,0±0,9	0,6±0,1	2,6±0,6	0,9±0,1	2,1±0,3
TAS2R7C	3,2±0,3	1,7±0,5	0,4±0,0	0,0±0,0	4,9±1,3	2,4±0,5	4,0±0,5	2,2±0,7	2,8±0,6	1,4±0,3	2,8±0,8	1,5±0,3	3,2±1,4	2,7±0,7	
Amino-ácidos	GRM4	4,9±0,3	5,1±0,5	0,2±0,0	0,2±0,0	4,2±0,6	4,2±0,5	5,1±0,7	5,3±0,9	3,1±0,9	3,0±1,0	4,4±1,6	4,5±0,7	2,5±0,4	2,6±0,3
	GPRC6A	1,4±0,2	0,2±0,0	0,2±0,0	0,1±0,0	0,7±0,0	0,5±0,0	1,4±0,3	0,9±0,1	1,9±0,5	1,4±0,3	1,7±0,3	1,5±0,3	0,2±0,0	0,1±0,0
	MGLUR1	2,0±0,5	2,0±0,6	0,8±0,2	0,8±0,0	1,5±0,6	1,5±0,1	2,5±0,6	2,6±0,6	3,1±1,4	3,2±1,2	3,3±1,2	3,3±1,8	1,7±0,2	1,7±0,3
	TAS1R1	1,0±0,1	1,1±0,2	0,4±0,0	0,4±0,0	1,2±0,2	1,2±0,3	1,6±0,2	1,7±0,2	1,2±0,2	1,3±0,3	1,2±0,2	1,3±0,2	0,2±0,0	0,2±0,0
Ácidos graxos	GPR40	1,3±0,3	0,8±0,1	0,0±0,0	0,0±0,0	2,9±0,5	2,3±0,5	3,4±0,7	2,2±0,7	1,6±0,5	0,9±0,1	1,6±0,1	0,7±0,1	0,2±0,0	0,4±0,1
	GPR43	1,2±0,2	0,8±0,1	0,1±0,0	0,0±0,0	2,6±0,5	2,8±0,3	3,2±0,3	2,2±0,3	2,7±0,5	3,3±0,7	2,6±0,4	2,9±0,7	0,2±0,0	0,3±0,0
	GPR41	2,5±0,3	0,9±0,2	0,2±0,0	0,0±0,0	4,3±0,3	1,1±0,1	5,0±0,7	1,9±0,2	2,8±0,4	2,8±0,8	1,3±0,3	1,4±0,3	0,1±0,0	0,3±0,0
	GPR120	1,6±0,5	2,1±0,5	1,4±0,2	1,2±0,3	2,1±0,3	2,7±0,5	2,2±0,3	2,3±0,5	2,2±0,3	2,7±0,7	3,0±0,3	3,5±0,9	1,2±0,1	1,1±0,3
	GPR84	0,4±0,0	0,2±0,0	0,4±0,0	0,0±0,0	0,6±0,1	0,0±0,0	0,7±0,2	0,0±0,0	1,2±0,4	0,0±0,0	0,9±0,1	0,0±0,0	0,4±0,0	0,2±0,0
Mean	2,0±0,3	2,4±0,4	0,7±0,1	1,1±0,2	2,8±0,6	3,0±0,6	3,1±0,6	3,2±0,8	2,2±0,5	2,5±0,7	2,6±0,6	2,5±0,7	1,9±0,4	2,5±0,6	

Tabela C18 Distribuição de mutações não-sinônimas por gene, posição no cromossomo, mudança de alelo, identificação de SNP do dbSNPs, SIFT e frequência alélica por população. Escore SIFT  $\leq 0.05$  é considerado como deletério na função da proteína e valores  $> 0.05$  são considerados tolerados sem nenhum dano à função da proteína.

Genes	Posição em pb	Mudança de Nucleotídeo	Mudança de AA	ID SNP	SIFT	Frequência Alélica						
						Internacional	Ibericos	Crioulos	Brasileiros	Domésticos asiáticos	ASWB	EUWB
TAS2R20	63904194	G /A		rs336714661	na	0,0323	0,7500	0,1071	0,0000	0,0000	0,0000	0,1875
	63904542	G /A		rs340207639	na	0,0161	0,0000	0,0357	0,0000	0,0000	0,0000	0,0000
	63904557	G /C		rs80827179	na	0,3065	0,0000	0,2500	0,3333	0,0000	0,0000	0,0000
	63904569	G /A		rs80911144	na	0,0806	1,0000	0,2143	0,1667	0,0000	0,0000	0,6250
	63904681	T /G		rs339125673	na	0,0000	0,0000	0,0357	0,0000	0,4286	0,0833	0,0000
	63904848	C /T		rs341034956	na	0,1129	1,0000	0,2143	0,1667	0,0000	0,0000	0,6250
	63904854	A /G		rs322322534	na	0,0323	0,0000	0,0000	0,1667	0,0000	0,0000	0,0000
	63904949	G /T		rs346100475	na	0,1129	0,8750	0,2143	0,1667	0,0000	0,0000	0,6875
TAS2R9	63977077	T /G	S113R	-	0,01	0,0000	0,2500	0,0000	0,0000	0,0000	0,0000	0,0000
	63977391	C /G	S218C	rs323930780	0,25	0,0000	0,0000	0,0000	0,0000	0,2857	0,1667	0,0000
	63977261	C /A	H175N	rs81490582	0,30	0,1290	0,0000	0,0357	0,1667	0,8571	0,1667	0,0000
	63976924	G /A	M62I	rs341774888	1,00	0,1774	1,0000	0,5000	0,5000	1,0000	1,0000	0,5000
	63977076	G /A	S113N	rs81384489	1,00	0,1452	0,0000	0,3214	0,1667	1,0000	1,0000	0,1250
	63977246	A /G	M170V	rs329437592	1,00	0,0806	0,0000	0,0000	0,1667	0,1429	0,0000	0,0000
TAS2R10	63965528	G /A		-	na	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	63965543	T /C		rs328033077	na	0,0323	0,2500	0,0357	0,0000	0,0000	0,0000	0,0625

	63965603	C /T	rs338533492	na	0,0323	0,2500	0,0357	0,1667	0,0000	0,0000	0,0625
	63965779	G /T	-	na	0,0000	0,0000	0,0357	0,0000	0,0000	0,0000	0,0000
	63966178	G /A	rs336151567	na	0,0000	0,0000	0,1786	0,0000	0,0000	0,2500	0,0000
	63966296	A /G	rs320357876	na	0,0000	0,0000	0,1786	0,0000	0,0000	0,2500	0,0000
	63966333	A /T	rs331570129	na	0,0645	0,1250	0,0357	0,1667	0,0000	0,0000	0,1250
TAS2R42	63867101	T /C	rs327293870	na	0,2419	0,7500	0,4643	0,3333	0,2857	0,3333	0,5625
	63867102	G /A	rs341683063	na	0,2419	0,7500	0,4643	0,3333	0,2857	0,3333	0,5625
	63867473	A /G	rs340194692	na	0,2581	1,0000	0,4286	0,3333	0,2857	0,5000	0,6875
	63867563	T /C	rs332299843	na	0,2742	1,0000	0,3929	0,3333	0,2857	0,3333	0,5625
	63867827	T /C	rs342354538	na	0,0968	0,8750	0,1786	0,1667	0,0000	0,0000	0,6875
	63867841	T /C	rs322012964	na	0,1452	0,0000	0,2857	0,1667	0,2857	0,5000	0,0000
	63867877	A /G	-	na	0,0000	0,0000	0,0000	0,0000	0,0000	0,0833	0,0000
	63867975	A /C	rs324960785	na	0,0968	0,0000	0,2143	0,1667	0,1429	0,2500	0,0000
TAS2R16	25883755	C /T	rs341218110	na	0,5161	0,1250	0,1786	0,3333	0,5000	0,3333	0,2500
	25883684	T /A	-	na	0,0000	0,0000	0,0000	0,3333	0,0000	0,0000	0,2500
	25883851	G /C	rs323409960	na	0,0000	0,0000	0,0000	0,5000	0,0000	0,0000	0,1875
	25883916	C /G	-	na	0,0000	0,0000	0,0000	0,0000	0,0000	0,0833	0,0000
	25883983	G /T	rs323772679	na	0,0000	0,0000	0,0000	0,5000	0,0000	0,0000	0,2500
	25884094	A /C	-	na	0,0000	0,0000	0,0000	0,0000	0,0714	0,0000	0,0000
	25884101	A /T	rs328024691	na	0,0323	0,0000	0,0000	0,1667	0,0000	0,0000	0,0000
	25884264	A /G	rs333868526	na	0,0000	0,0000	0,0000	0,5000	0,0000	0,0000	0,2500
TAS2R38	8357545	G /A	-	na	0,0000	0,0000	0,0000	0,0000	0,0714	0,0000	0,0000
	8357565	C /A	rs344373699	na	0,1935	0,0000	0,3214	0,0000	0,0000	0,0000	0,1875

	8357605	A /C		rs340228133	na	0,3387	0,3750	0,8214	0,6667	0,8571	0,6667	0,5000
	8357647	G /A		-	na	0,0000	0,0000	0,0000	0,0000	0,0000	0,1667	0,0000
	8358039	A /T		rs332855976	na	0,0000	0,0000	0,0000	0,0000	0,2857	0,0000	0,0000
	8358202	G /A		rs336060799	na	0,4516	0,3750	0,7500	0,6667	0,7143	0,8333	0,5000
	8358221	G /A		-	na	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0625
	8358348	C /G		rs337735554	na	0,1452	0,3750	0,4286	0,1667	0,6429	0,0000	0,2500
	8358485	T /C		rs319573015	na	0,3710	0,6250	0,5357	0,8333	0,6429	0,5000	0,3750
	8358491	G /A		rs340948118	na	0,5323	0,7500	0,7500	0,8333	0,7143	0,8333	0,4375
TAS2R39	7358957	T /C	L114S	rs331114472	0,09	0,3387	0,6250	0,5000	0,3333	0,0000	0,0000	0,6875
	7359852	G /C	E335D	rs319025082	0,10	0,1129	0,0000	0,2500	0,1667	0,5000	0,2500	0,0000
	7359394	C /A	H183N	rs336119089	0,11	0,0161	0,0000	0,1071	0,0000	0,0000	0,0000	0,0000
	7358913	C /A	S22R	-	0,19	0,0161	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	7359424	T /C	F193L	rs328351764	0,29	0,0484	0,0000	0,0000	0,3333	0,0714	0,1667	0,0000
	7359355	G /T	G170C	rs345532192	0,38	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0625
	7359467	C /T	T207I	rs337700679	0,40	0,4032	0,2500	0,5357	0,8333	0,4286	0,3333	0,7500
	7358980	A /G	I45V	rs323606521	0,42	0,1290	0,0000	0,1429	0,0000	0,0714	0,0833	0,0000
	7359526	G /A	A227T	rs81209905	0,42	0,0161	0,0000	0,1071	0,0000	0,3571	0,0000	0,0000
	7358903	T /C	I19T	rs335568369	0,44	0,1129	0,0000	0,2143	0,5000	0,7143	0,4167	0,0625
	7359212	G /A	C112Y	-	0,44	0,0161	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	7359437	A /G	K197R	-	0,44	0,0000	0,0000	0,0000	0,0000	0,0000	0,0833	0,0000
	7359190	G /A	G115S	rs323238022	0,47	0,4677	0,6250	0,7143	0,6667	0,8571	0,5833	0,8125
	7358965	G /A	V40I	rs345006834	0,51	0,2419	0,0000	0,3571	0,5000	0,9286	0,6667	0,0625
	7359059	A /C	N71T	rs342835508	0,67	0,7097	0,7500	0,8929	1,0000	1,0000	0,8333	0,8750

	7358972	C /T	A42V	rs339810026	0,76	0,0161	0,0000	0,0714	0,0000	0,0000	0,0000	0,0000
	7359404	A /G	N186S	rs344640384	0,88	0,1129	0,0000	0,0000	0,3333	0,1429	0,2500	0,0625
	7359253	G /T	V136L	-	0,89	0,0000	0,5000	0,0000	0,0000	0,0000	0,0000	0,0000
	7359213	T /C		rs335079257	na	0,0000	0,0000	0,0000	0,1667	0,0714	0,1667	0,0625
TAS2R41	7018811	C /T	A307T	rs320709106	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,0833	0,0000
	7018883	G /A	L283F	rs342189509	0,01	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,1250
	7018818	G /C	F304L	rs342228000	0,03	0,0323	0,0000	0,0000	0,0000	0,0714	0,0000	0,0000
	7019528	A /G	Y68H	rs345262132	0,03	0,0000	0,0000	0,0000	0,1667	0,0714	0,1667	0,0000
	7019414	G /C	L106V	rs333683448	0,05	0,0161	0,0000	0,0000	0,0000	0,2143	0,0000	0,0000
	7019154	A /C	I192M	rs323873992	0,12	0,0484	0,2500	0,2143	0,0000	0,0714	0,0000	0,0625
	7018816	C /T	R305Q	rs331758880	0,27	0,0484	0,0000	0,0000	0,0000	0,0000	0,3333	0,0000
	7019271	G /C	H153Q	rs342853167	0,32	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	7019623	G /A	A107V	rs325525037	0,48	0,2742	0,3750	0,0714	0,3333	0,3571	0,2500	0,0000
	7019146	G /A	S195F	-	0,60	0,0000	0,2500	0,0000	0,0000	0,0000	0,0000	0,0000
	7018916	C /T	V272I	rs322495269	1,00	0,0968	0,2500	0,1429	0,0000	0,2143	0,0000	0,3750
	7019387	A /T	S115T	rs326069450	1,00	0,2742	0,0000	0,2500	0,1667	0,6429	0,4167	0,2500
TAS2R60	7046531	T /C	I23V	rs339482728	0,04	0,0968	0,0000	0,0714	0,0000	0,1429	0,0000	0,0000
	7045964	T /C	S212G	rs337678249	0,37	0,1290	0,0000	0,0714	0,0000	0,0714	0,0000	0,0000
	7045265	C /G	E313D	-	0,42	0,0000	0,3750	0,0000	0,0000	0,0000	0,0000	0,0000
	7046591	C /T	A3T	rs322889806	0,56	0,0484	0,0000	0,0000	0,0000	0,0000	0,1667	0,0000
TAS2R7C	63985159	G /C		rs336392315	na	0,0000	0,0000	0,0357	0,0000	0,0000	0,0000	0,1250
	63985428	T /C		rs325458119	na	0,1290	0,0000	0,0714	0,1667	0,5000	0,7500	0,0000
	63985814	G /C		rs344408296	na	0,2097	0,7500	0,4643	0,3333	0,8571	0,8333	0,5625



	63985874	A /G		rs327810383	na	0,0806	0,0000	0,2857	0,0000	0,7143	0,5833	0,0000
	63986034	A /G		rs335556860	na	0,2258	0,7500	0,4643	0,3333	0,8571	0,5000	0,5000
GRM4	34850792	C /T	A908T	rs338034384	0,69	0,1774	0,7500	0,2857	0,0000	0,0714	0,0000	0,7500
	34839257	G /C	T697S	rs330398200	0,86	0,0000	0,0000	0,0000	0,1667	0,0000	0,0000	0,0000
GPRC6A	50131432	C /T	R18C	rs324254536	0,08	0,0000	0,0000	0,0714	0,1667	0,2143	0,1667	0,0000
	50131471	A /G	S31G	rs325860095	0,18	0,0000	0,0000	0,0357	0,1667	0,1429	0,0000	0,0000
	50131438	C /T	P20S	rs334455020	0,96	0,0000	0,0000	0,0357	0,0000	0,0000	0,0833	0,0000
GRM1	21476805	C /T	M791I	-	0,00	0,0161	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
TAS1R1	62360925	A /G		rs331670639	1,00	0,0161	0,0000	0,1429	0,3333	0,6429	0,5000	0,0000
	62362410	C /G	L579V	rs330666697	0,01	0,0161	0,0000	0,0714	0,1667	0,4286	0,0000	0,0000
	62357814	C /A	S354Y	rs325274060	0,05	0,0000	0,0000	0,0000	0,0000	0,0000	0,1667	0,0000
	62357499	C /T	P249L	rs334245415	0,27	0,0484	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	62362761	C /G	L696V	rs339177207	1,00	0,0161	0,0000	0,0714	0,1667	0,3571	0,0000	0,0000
FFAR1	40339838	G /A	R211C	rs323728911	0,01	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,1875
	40340399	A /C	C25G	rs329339162	0,07	0,1129	0,0000	0,2143	0,3333	0,7143	0,3333	0,0000
	40339583	C /T	G296S	rs324621867	0,26	0,0161	0,0000	0,0357	0,1667	0,0000	0,0000	0,0000
	40339952	C /A	A173S	-	0,71	0,0161	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
FFAR2	40282557	T /G	M159L	rs320946055	0,78	0,0000	0,0000	0,1429	0,1667	0,5714	0,0000	0,0000
FFAR3	40334700	G /A	P68L	rs318787211	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,1875
	40334287	G /T	L206M	-	0,15	0,0000	0,0000	0,0000	0,1667	0,0000	0,0000	0,0000
	40334023	T /C	T294A	rs336566065	1,00	0,0323	0,0000	0,0357	0,1667	0,0000	0,0000	0,0000

na= não analisados. Análise de SIFT foi realizada com Variant Effector Predictor (VEP) que usa as informações localizadas no Ensembl, assim para genes sem informações disponíveis não foi possível conseguir os escores ou predições das mutações na função protéica.

## FIGURAS

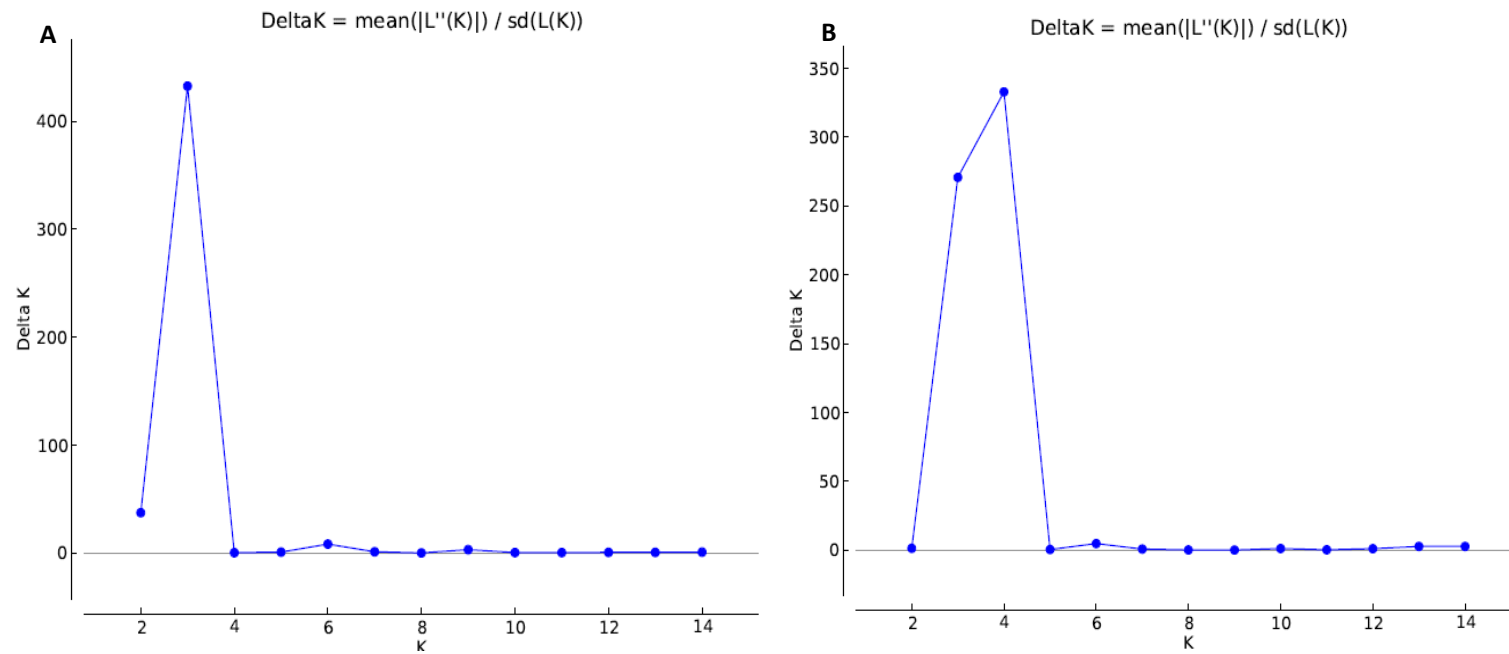


Figura C7 Resultados da análise de STRUCTURE para SNPs de genes de receptores do gosto amargo: o melhor K de 1 a 15 pela estatística Delta K para o conjunto de SNPs (A) e para SNPs não-sinônimos (B).

## ANEXO D: CAPÍTULO 4

Arquivo no formato GFF3 que foi utilizado na anotação das variantes.

1	GPRC6A	start_codon	50131381	50131383	.	+	.	;
1	GPRC6A	CDS	50131381	50131574	.	+	0	;
1	GPRC6A	exon	50131244	50131574	.	+	.	;
1	GPRC6A	CDS	50133335	50133638	.	+	1	;
1	GPRC6A	exon	50133335	50133638	.	+	.	;
1	GPRC6A	CDS	50135971	50136081	.	+	0	;
1	GPRC6A	stop_codon	50136082	50136084	.	+	.	;
1	GPRC6A	exon	50135971	50136085	.	+	.	;
1	GPRC6A	intergenic	50121244	50131243	.	.	.	;
1	GPRC6A	intergenic	50136086	50146085	.	.	.	;
1	MGLUR1	intron	50131575	50133334	.	.	.	;
1	MGLUR1	intron	50133639	50135970	.	.	.	;
1	MGLUR1	stop_codon	21476514	21476516	.	-	.	;
1	MGLUR1	CDS	21476517	21477448	.	-	2	;
1	MGLUR1	exon	21476379	21477448	.	-	.	;
1	MGLUR1	CDS	21489473	21489599	.	-	0	;
1	MGLUR1	exon	21489473	21489599	.	-	.	;
1	MGLUR1	CDS	21502830	21502998	.	-	1	;

1	MGLUR1	exon	21502830	21502998	.	-	.	;
1	MGLUR1	CDS	21507277	21507523	.	-	2	;
1	MGLUR1	exon	21507277	21507523	.	-	.	;
1	MGLUR1	CDS	21552147	21552382	.	-	1	;
1	MGLUR1	exon	21552147	21552382	.	-	.	;
1	MGLUR1	CDS	21709765	21710014	.	-	2	;
1	MGLUR1	exon	21709765	21710014	.	-	.	;
1	MGLUR1	CDS	21813937	21814636	.	-	0	;
1	MGLUR1	start_codon	21814634	21814636	.	-	.	;
1	MGLUR1	exon	21813937	21814964	.	-	.	;
1	MGLUR1	intergenic	21466379	21476378	.	-	.	;
1	MGLUR1	intergenic	21814965	21824964	.	.	-	;
1	MGLUR1	intron	21477449	21489472	.	-	.	;
1	MGLUR1	intron	21489600	21502829	.	-	.	;
1	MGLUR1	intron	21502999	21507276	.	-	.	;
1	MGLUR1	intron	21507524	21552146	.	-	.	;
1	MGLUR1	intron	21552383	21709764	.	-	.	;
1	MGLUR1	intron	21710015	21813936	.	-	.	;
1	MGLUR1	UTR	21476379	21476516	.	-	.	;
1	MGLUR1	UTR	21814636	21814964	.	-	.	;
5	GPR84	stop_codon	20410225	20410227	.	-	.	;
5	GPR84	exon	20410225	20411145	.	-	.	;
5	GPR84	CDS	20410225	20411145	.	-	.	;

5	GPR84	exon	20411148	20411195	.	-	.	;
5	GPR84	intergenic	20404644	20410224	.	-	.	;
5	GPR84	intergenic	20411149	20416116	.	-	.	;
5	GPR84	intron	20411146	20411195	.	-	.	;
5	TAS2R7C	exon	63985142	63986080	.	+	.	;
5	TAS2R7C	CDS	63985142	63986080	.	+	.	;
5	TAS2R7C	start_codon	63985142	63985144	.	+	.	;
5	TAS2R7C	stop_codon	63986078	63986080	.	+	.	;
5	TAS2R7C	intergenic	63982692	63985141	.	.	.	;
5	TAS2R7C	intergenic	63986081	63996080	.	.	.	;
5	TAS2R9	start_codon	63976739	63976741	.	+	.	;
5	TAS2R9	exon	63976739	63977674	.	+	.	;
5	TAS2R9	CDS	63976739	63977674	.	+	.	;
5	TAS2R9	stop_codon	63977672	63977674	.	+	.	;
5	TAS2R9	intergenic	63971739	63976738	.	.	.	;
5	TAS2R9	intergenic	63977675	63982674	.	.	.	;
5	TAS2R10	start_codon	63965446	63965448	.	+	.	;
5	TAS2R10	exon	63965446	63966375	.	+	.	;
5	TAS2R10	CDS	63965446	63966375	.	+	.	;
5	TAS2R10	stop_codon	63966373	63966375	.	+	.	;
5	TAS2R10	intergenic	63958146	63965445	.	.	.	;
5	TAS2R10	intergenic	63966376	63971725	.	.	.	;
5	TAS2R20	start_codon	63904140	63904142	.	+	.	;

5	TAS2R20	exon	63904140	63905054	.	+	.	;
5	TAS2R20	CDS	63904140	63905054	.	+	.	;
5	TAS2R20	stop_codon	63905052	63905054	.	+	.	;
5	TAS2R20	intergenic	63894140	63904139	.	.	.	;
5	TAS2R20	intergenic	63905055	63915054	.	.	.	;
5	TAS2R42	exon	63867091	63868041	.	+	.	;
5	TAS2R42	CDS	63867091	63868041	.	+	.	;
5	TAS2R42	start_codon	63867091	63867093	.	+	.	;
5	TAS2R42	stop_codon	63868039	63868041	.	+	.	;
5	TAS2R42	intergenic	63857091	63867090	.	.	.	;
5	TAS2R42	intergenic	63868042	63878041	.	.	.	;
6	GPR40	stop_codon	40339566	40339568	.	-	.	;
6	GPR40	CDS	40339569	40340468	.	-	.	;
6	GPR40	start_codon	40340466	40340468	.	-	.	;
6	GPR40	exon	40339566	40340468	.	-	.	;
6	GPR40	intergenic	40335081	40339565	.	-	.	;
6	GPR40	intergenic	40340469	40343065	.	-	.	;
6	GPR41	start_codon	40283026	40283028	.	-	.	;
6	GPR41	CDS	40282039	40283028	.	-	.	;
6	GPR41	stop_codon	40282039	40282041	.	-	.	;
6	GPR41	exon	40282039	40283031	.	-	.	;
6	GPR41	intergenic	40272039	40282038	.	-	.	;
6	GPR41	intergenic	40283032	40293031	.	-	.	;

6	GPR42	stop_codon	40333922	40333924	.	-	.	;
6	GPR42	CDS	40333925	40334902	.	-	.	;
6	GPR42	start_codon	40334900	40334902	.	-	.	;
6	GPR42	exon	40333922	40334902	.	-	.	;
6	GPR42	intergenic	40323922	40333921	.	.	-	.
6	GPR42	intergenic	40334903	40335070	.	.	-	.
6	TAS1R1	start_codon	62350603	62350605	.	+	.	;
6	TAS1R1	CDS	62350603	62350793	.	+	0	;
6	TAS1R1	exon	62350603	62350793	.	+	.	;
6	TAS1R1	CDS	62354649	62354955	.	+	1	;
6	TAS1R1	exon	62354649	62354955	.	+	.	;
6	TAS1R1	CDS	62357248	62357343	.	+	0	;
6	TAS1R1	exon	62357248	62357343	.	+	.	;
6	TAS1R1	CDS	62357348	62358007	.	+	0	;
6	TAS1R1	exon	62357348	62358007	.	+	.	;
6	TAS1R1	CDS	62360362	62360583	.	+	0	;
6	TAS1R1	exon	62360362	62360583	.	+	.	;
6	TAS1R1	CDS	62360879	62360999	.	+	0	;
6	TAS1R1	exon	62360879	62360999	.	+	.	;
6	TAS1R1	CDS	62362273	62363201	.	+	2	;
6	TAS1R1	stop_codon	62363202	62363204	.	+	.	;
6	TAS1R1	exon	62362273	62363204	.	+	.	;
6	TAS1R1	intergenic	62349877	62350602	.	.	.	;

6	TAS1R1	intergenic	62363205	62363714	.	.	.	;
6	TAS1R1	intron	62350794	62354648	.	.	.	;
6	TAS1R1	intron	62354956	62357247	.	.	.	;
6	TAS1R1	intron	62357344	62357347	.	.	.	;
6	TAS1R1	intron	62358008	62360361	.	.	.	;
6	TAS1R1	intron	62360584	62360878	.	.	.	;
6	TAS1R1	intron	62361000	62362272	.	.	.	;
6	TAS1R3	start_codon	58111944	58111946	.	+	.	;
6	TAS1R3	CDS	58111944	58112128	.	+	0	;
6	TAS1R3	exon	58111612	58112128	.	+	.	;
6	TAS1R3	CDS	58112201	58112501	.	+	1	;
6	TAS1R3	exon	58112201	58112501	.	+	.	;
6	TAS1R3	CDS	58112585	58113367	.	+	0	;
6	TAS1R3	exon	58112585	58113367	.	+	.	;
6	TAS1R3	CDS	58113482	58113685	.	+	0	;
6	TAS1R3	exon	58113482	58113685	.	+	.	;
6	TAS1R3	CDS	58113818	58113938	.	+	0	;
6	TAS1R3	exon	58113818	58113938	.	+	.	;
6	TAS1R3	CDS	58114082	58115052	.	+	2	;
6	TAS1R3	stop_codon	58115053	58115055	.	+	.	;
6	TAS1R3	exon	58114082	58115907	.	+	.	;
6	TAS1R3	intergenic	58109541	58111611	.	.	.	;
6	TAS1R3	intergenic	58115908	58116535	.	.	.	;



6	TAS1R3	intron	58112129	58112200	.	.	.	;
6	TAS1R3	intron	58112502	58112584	.	.	.	;
6	TAS1R3	intron	58113368	58113481	.	.	.	;
6	TAS1R3	intron	58113686	58113817	.	.	.	;
6	TAS1R3	intron	58113939	58114081	.	.	.	;
6	TAS1R3	UTR	58111612	58111944	.	+	.	;
6	TAS1R3	UTR	58115055	58115907	.	+	.	;
7	GRM4	stop_codon	34839241	34839243	.	-	.	;
7	GRM4	exon	34839241	34839290	.	-	.	;
7	GRM4	CDS	34839241	34839290	.	-	.	;
7	GRM4	exon	34844404	34844650	.	-	.	;
7	GRM4	CDS	34844404	34844650	.	-	.	;
7	GRM4	exon	34850439	34851374	.	-	.	;
7	GRM4	CDS	34850439	34851374	.	-	.	;
7	GRM4	exon	34854475	34854611	.	-	.	;
7	GRM4	CDS	34854475	34854611	.	-	.	;
7	GRM4	exon	34854855	34855055	.	-	.	;
7	GRM4	CDS	34854855	34855055	.	-	.	;
7	GRM4	exon	34865610	34865750	.	-	.	;
7	GRM4	CDS	34865610	34865750	.	-	.	;
7	GRM4	exon	34867710	34867864	.	-	.	;
7	GRM4	CDS	34867710	34867864	.	-	.	;
7	GRM4	exon	34870300	34870435	.	-	.	;

7	GRM4	CDS	34870300	34870435	.	-	.	;
7	GRM4	exon	34893117	34893333	.	-	.	;
7	GRM4	CDS	34893117	34893333	.	-	.	;
7	GRM4	CDS	34927186	34927704	.	-	.	;
7	GRM4	start_codon	34927702	34927704	.	-	.	;
7	GRM4	exon	34927186	34928071	.	-	.	;
7	GRM4	intergenic	34829241	34839240	.	-	.	;
7	GRM4	intergenic	34928072	34938071	.	-	.	;
7	GRM4	intron	34839291	34844403	.	-	.	;
7	GRM4	intron	34844651	34850438	.	-	.	;
7	GRM4	intron	34851375	34854474	.	-	.	;
7	GRM4	intron	34854612	34854854	.	-	.	;
7	GRM4	intron	34855056	34865609	.	-	.	;
7	GRM4	intron	34865751	34867709	.	-	.	;
7	GRM4	intron	34867865	34870299	.	-	.	;
7	GRM4	intron	34870436	34893116	.	-	.	;
7	GRM4	intron	34893334	34927185	.	-	.	;
7	GRM4	3_UTR	34839243	34839290	.	-	.	;
7	GRM4	5_UTR	34927704	34928071	.	-	.	;
13	CASR	stop_codon	147907932	147907934	.	-	.	;
13	CASR	CDS	147907935	147909439	.	-	2	;
13	CASR	exon	147907932	147909439	.	-	.	;
13	CASR	CDS	147910886	147911009	.	-	0	;

13	CASR	exon	147910886	147911009	.	-	.	;
13	CASR	CDS	147916007	147916237	.	-	0	;
13	CASR	exon	147916007	147916237	.	-	.	;
13	CASR	CDS	147927407	147928291	.	-	0	;
13	CASR	exon	147927407	147928291	.	-	.	;
13	CASR	CDS	147932029	147932335	.	-	1	;
13	CASR	exon	147932029	147932335	.	-	.	;
13	CASR	CDS	147934886	147935070	.	-	0	;
13	CASR	start_codon	147935068	147935070	.	-	.	;
13	CASR	exon	147934886	147935070	.	-	.	;
13	CASR	intergenic	147897932	147907931	.	.	.	;
13	CASR	intergenic	147935071	147940283	.	.	.	;
13	CASR	intron	147909440	147910885	.	.	.	;
13	CASR	intron	147911010	147916006	.	.	.	;
13	CASR	intron	147916238	147927406	.	.	.	;
13	CASR	intron	147928292	147932028	.	.	.	;
13	CASR	intron	147932336	147934885	.	.	.	;
14	GPR120	exon	114743575	114744148	.	+	.	;
14	GPR120	CDS	114743582	114744148	.	+	.	;
14	GPR120	start_codon	114743582	114743584	.	+	.	;
14	GPR120	exon	114752983	114753111	.	+	.	;
14	GPR120	CDS	114752983	114753111	.	+	.	;
14	GPR120	exon	114764725	114765158	.	+	.	;

14	GPR120	CDS	114764725	114765158	.	+	.	;
14	GPR120	stop_codon	114765112	114765114	.	+	.	;
14	GPR120	intergenic	114733575	114743574	.	+	.	;
14	GPR120	intergenic	114765159	114769008	.	+	.	;
14	GPR120	intron	114744149	114752982	.	+	.	;
14	GPR120	intron	114753112	114764724	.	+	.	;
18	TAS2R16	start_codon	25883452	25883454	.	+	.	;
18	TAS2R16	exon	25883452	25884354	.	+	.	;
18	TAS2R16	CDS	25883452	25884354	.	+	.	;
18	TAS2R16	stop_codon	25884352	25884354	.	+	.	;
18	TAS2R16	intergenic	25873452	25883451	.	.	.	;
18	TAS2R16	intergenic	25884355	25894354	.	.	.	;
18	TAS2R38	exon	8357518	8358525	.	+	.	;
18	TAS2R38	CDS	8357518	8358525	.	+	.	;
18	TAS2R38	start_codon	8357518	8357521	.	+	.	;
18	TAS2R38	stop_codon	8358523	8358525	.	+	.	;
18	TAS2R38	intergenic	8347518	8357517	.	.	.	;
18	TAS2R38	intergenic	8358526	8368525	.	.	.	;
18	TAS2R39	start_codon	7358848	7358851	.	+	.	;
18	TAS2R39	CDS	7358848	7359852	.	+	0	;
18	TAS2R39	stop_codon	7359853	7359855	.	+	.	;
18	TAS2R39	exon	7358848	7359855	.	+	.	;
18	TAS2R39	intergenic	7348848	7358847	.	.	.	;

18	TAS2R39	intergenic	7359856	7369855	.	.	.	;
18	TAS2R41	stop_codon	7018806	7018808	.	-	.	;
18	TAS2R41	CDS	7018809	7019729	.	-	0	;
18	TAS2R41	start_codon	7019727	7019729	.	-	.	;
18	TAS2R41	exon	7018806	7019729	.	-	.	;
18	TAS2R41	intergenic	7008806	7018804	.	-	.	;
18	TAS2R41	intergenic	7019730	7029729	.	-	.	;
18	TAS2R60	stop_codon	7045247	7045249	.	-	.	;
18	TAS2R60	CDS	7045250	7045282	.	-	0	;
18	TAS2R60	exon	7045247	7045282	.	-	.	;
18	TAS2R60	CDS	7045677	7046597	.	-	0	;
18	TAS2R60	start_codon	7046595	7046597	.	-	.	;
18	TAS2R60	exon	7045677	7046597	.	-	.	;
18	TAS2R60	intergenic	7046598	7056597	.	-	.	;
18	TAS2R60	intergenic	7035247	7045246	.	-	.	;
18	TAS2R60	intron	7045283	7045676	.	-	.	;