

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

**VARREDURAS GENÔMICAS PARA A DETECÇÃO DE
VARIANTES GENÉTICAS ASSOCIADAS À REPRODUÇÃO DE
CÃES**

Rafaela Beatriz Pintor Torrecilha

Médica Veterinária

2018

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

**VARREDURAS GENÔMICAS PARA A DETECÇÃO DE
VARIANTES GENÉTICAS ASSOCIADAS À REPRODUÇÃO DE
CÃES**

Rafaela Beatriz Pintor Torrecilha

Orientador: José Fernando Garcia

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para obtenção do título de Doutor em Medicina Veterinária (Reprodução Animal).

2018

T689v Torrecilha, Rafaela Beatriz Pintor
 Varreduras Genômicas para a Detecção de Variantes
 Genéticas Associadas à Reprodução de Cães / Rafaela
 Beatriz Pintor Torrecilha. -- Jaboticabal, 2018
 67 p. : il., tabs.

 Tese (doutorado) - Universidade Estadual Paulista
 (Unesp), Faculdade de Ciências Agrárias e Veterinárias,
 Jaboticabal
 Orientador: José Fernando Garcia

 1. Dog. 2. GDF9. 3. Haplotype. 4. Deleterious. 5.
 Reproduction. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos
pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: VARREDURAS GENÔMICAS PARA A DETECÇÃO DE VARIANTES GENÉTICAS ASSOCIADAS À REPRODUÇÃO DE CÃES

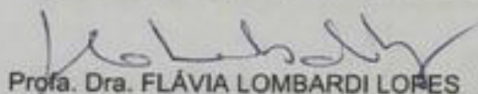
AUTORA: RAFAELA BEATRIZ PINTOR TORRECILHA

ORIENTADOR: JOSÉ FERNANDO GARCIA

Aprovada como parte das exigências para obtenção do Título de Doutora em MEDICINA VETERINÁRIA, área: Reprodução Animal pela Comissão Examinadora:



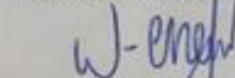
Prof. Dr. JOSÉ FERNANDO GARCIA
Departamento de Apoio, Produção e Saúde Animal / FMVA-Araçatuba/SP



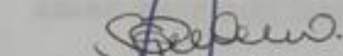
Profa. Dra. FLÁVIA LOMBARDI LOPES
Departamento de Apoio, Produção e Saúde Animal / FMVA/UNESP - Araçatuba



Profa. Dra. MARICY APARECIDO FERREIRA
Depto de Medicina Veterinária Preventiva e Reprodução Animal / FCAV / UNESP - Jaboticabal



Prof. Dr. WAGNER LUIS FERREIRA
Departamento de Clínica, Cirurgia e Reprodução Animal / Faculdade de Medicina Veterinária - Câmpus de Araçatuba/Unesp



Dra. SILVANA DE CÁSSIA PAULAN
Bióloga Autônoma / Santa Fé do Sul/SP

Jaboticabal, 25 de outubro de 2018

DADOS CURRICULARES DA AUTORA

RAFAELA BEATRIZ PINTOR TORRECILHA- filha de Sidney Torrecilha Basso e Maria Aparecida Pintor Dias Torrecilha, nascida em Araçatuba-SP, no dia 18 de setembro de 1989. Graduada em Medicina Veterinária pela Universidade de Marília (2012) e portadora do título de Mestre (2015) em Ciência Animal pela Faculdade de Medicina Veterinária de Araçatuba (FMVA), Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP). Durante o mestrado obteve bolsa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), sob orientação do Prof. Dr. Paulo César Ciarlini. No ano de 2015, ingressou no curso de Doutorado do Programa de Medicina Veterinária da Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal (FCAV- UNESP) na área de Reprodução Animal. Durante o doutorado obteve bolsa da CAPES e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo nº 2017/08373-1, sob orientação do Prof. Dr. José Fernando Garcia.

AGRADECIMENTOS

Gostaria de agradecer a Deus por tudo de maravilhoso que tem feito em minha vida, sempre me abençoando com o prazer de ter pessoas boas ao meu redor e por permitir o contato diário com animais, que são os seres mais puros e inocentes que podemos conhecer.

Aos meus pais, por me amarem, educarem e aconselharem de uma maneira tão generosa que só tenho palavras boas para descrever minha vida com vocês. Tudo que sou e realizo eu agradeço a vocês. Obrigada por serem os melhores pais que eu poderia ter. Vocês sempre me ensinaram a amar e a respeitar o próximo, o que eu sempre levei muito a sério. Por isso inclusive, que já passaram por nossa casa vários gatos, cães, pássaros, porquinho da índia, gambás, etc. Mãe, sei que isso gerava algumas broncas, mas cada animal resgatado que você me ajudava, havia um novo aprendizado e um destino mudado. Mãe, obrigada por me ajudar nessa jornada com os animais, se eu me tornar 1% da mulher que você é eu poderia sair desta vida muito feliz e realizada.

Aos meus irmãos, Maria Helena e Francis, vocês são os irmãos mais velhos mais amados desse mundo.

Yuri, meu amor, creio que essa tese não existiria sem você. Nem tenho palavras para descrever meu agradecimento, não só pela orientação no trabalho, mas por ser quem você é no nosso dia a dia. Nossa rotina não é fácil, trabalhar e morar juntos, saber separar trabalho de nossa vida pessoal, só é possível porque é com você. Da mesma forma que minha mãe é meu exemplo de mulher, você é meu exemplo de vida profissional, não conheço alguém mais dedicado e apaixonado pelo que faz do que você. Obrigada por tudo, meu eterno companheiro! Te amo!

A todos meus amigos e familiares, que diariamente me fazem sorrir e me ajudam a relaxar nos momentos mais estressantes, não citarei nomes para não ser injusta de esquecer alguém.

Aos pesquisadores e pós-graduandos do Laboratório de Bioquímica e Biologia Molecular de Araçatuba (LBBMA) por toda ajuda no decorrer do doutorado, principalmente ao Marco Milanesi, que auxiliou muito no desenvolvimento desta tese. Aos Professores José Fernando Garcia e Cárís Maroni Nunes, por permitirem minha estadia no laboratório. Prof. José Fernando Garcia, agradeço por me orientar, apoiar minha decisão em continuar trabalhando com cães, mesmo não sendo sua área. Obrigada por todo auxílio e por permitir que eu me apaixonasse por essa linda área de genética de cães.

À Universidade de BOKU em Viena (Áustria), em especial ao Prof. Johann Sölkner e seus alunos, que me receberam e foram de grande importância no decorrer deste trabalho.

À Universidade de Berna na Suíça, aos pesquisadores envolvidos na coleta e processamento de dados genômicos de cães, e em especial aos Professores Tosso Leeb e Claude Schelling, por toda colaboração e paciência, que foram essenciais para a realização desta tese.

À Prof^a Maricy e suas pós-graduandas Mari e Carol, que me receberam tão carinhosamente e auxiliaram no decorrer do trabalho. À Prof^a Silvia Crusco, por toda colaboração com experiências reprodutivas de cães.

À seção de pós-graduação da Universidade Estadual Paulista (UNESP) da Faculdade de Ciências Agrárias e Veterinárias (FCAV).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Agradeço também à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo nº 2017/08373-1, pelo suporte financeiro através de bolsa de estudo.

À todas pessoas que participaram e apoiaram direta e indiretamente para a realização deste trabalho.

Meus sinceros agradecimentos a todos!

“Leve na sua memória para o resto de sua vida, as coisas boas que surgiram no meio das dificuldades. Elas serão uma prova de sua capacidade em vencer as provas e lhe darão confiança na presença divina, que nos auxilia em qualquer situação, em qualquer tempo, diante de qualquer obstáculo.”

Chico Xavier

SUMÁRIO

| | |
|---|----|
| RESUMO | iv |
| ABSTRACT | v |
| LISTA DE ABREVIACÕES | vi |
| LISTA DE TABELAS | ix |
| LISTA DE FIGURAS | x |
| CAPÍTULO 1: Considerações Gerais | 1 |
| 1. A seleção artificial promoveu o aumento da frequência de variantes de alto impacto funcional no genoma canino..... | 1 |
| 2. Ferramentas genômicas podem auxiliar na caracterização de variantes funcionais associadas à reprodução em cães..... | 3 |
| 3. Variantes funcionais associadas à reprodução através de deficiência de homozigotos..... | 5 |
| 4. Variantes com alto impacto funcional por meio de análise de associação genômica ampla em cães..... | 7 |
| 5. Objetivos do presente trabalho..... | 7 |
| 6. Referências | 8 |
| CAPÍTULO 2: Association of a missense variant in <i>GDF9</i> with decreased average litter size in Entlebucher Mountain dogs | 12 |
| Abstract..... | 12 |
| 1. Introduction..... | 13 |
| 2. Material and Methods..... | 14 |
| 2.1. Summary of the data..... | 14 |
| 2.2. Genotype filtering..... | 14 |
| 2.3. Regression model..... | 15 |
| 2.4. Genome-wide association analysis..... | 16 |
| 2.5. Analysis of whole genome sequence data | 17 |
| 2.6. Orthology analyses | 17 |
| 3. Results..... | 18 |

| | |
|--|-----------|
| 3.1. Single-step GWAS maps litter size associations to a chromosomal region sheltering <i>GDF9</i> | 18 |
| 3.2. Whole genome sequence data reveal a novel missense variant in <i>GDF9</i> | 19 |
| 4. Discussion..... | 22 |
| 5. Conclusions..... | 24 |
| 6. Funding..... | 24 |
| 7. Acknowledgments | 24 |
| 8. Author contributions | 25 |
| 9. References..... | 25 |
| CAPÍTULO 3: High frequency canine haplotypes with exceptional deficiency of homozygosity..... | 28 |
| Abstract..... | 28 |
| 1. Introduction..... | 29 |
| 2. Material and Methods..... | 30 |
| 2.1. Genotypes..... | 30 |
| 2.2. Inbreeding analysis..... | 31 |
| 2.3. Choice of haplotype size..... | 32 |
| 2.4. Detection of haplotype with deficit of homozygotes..... | 32 |
| 2.5. Power analysis | 33 |
| 2.6. Analysis of carrier sequence data..... | 33 |
| 2.7. Gene annotation..... | 34 |
| 2.8. Association between haplotypes and average litter size..... | 34 |
| 3. Results..... | 35 |
| 3.1. Detection power at sample size below 1,000 dogs is constrained to HDH with frequency > 10%..... | 35 |
| 3.2. Low effective population size and high inbreeding supports high frequency HDH in dogs..... | 35 |
| 3.3. Genome-wide screening identifies sixteen HDHR across four dog breeds | 36 |
| 3.4. HDHR9 validates the <i>FOXI3</i> homozygous-lethal frameshift variant in | |

| | |
|--|----|
| Chinese Crested dogs..... | 42 |
| 3.5. HDHR8 indicates selection against the homozygous wild type <i>CBD103</i> genotype in Golden Retriever dogs..... | 43 |
| 3.6. Black-and-tan coat color as the putative unfavorable recessive phenotype underlying HDHR10 in Labrador and Golden Retriever dogs..... | 45 |
| 3.7. HDHR16 points to <i>COL25A1</i> as another candidate gene harboring unfavorable alleles in retrievers dogs..... | 46 |
| 3.8. HDHR3 contains the largest CNV in the dog genome spanning <i>AMY2B</i> | 47 |
| 3.9. HDHR1, HDHR13 and HDHR15 contain candidate genes implicated in retinal disease..... | 48 |
| 3.10. HDHR2 points to an obesity-related gene with high differentiation between dogs and wild canids..... | 48 |
| 3.11. The common HDHR4 locus maps to a complex copy number variant encompassing <i>ARHGAP27</i> | 49 |
| 3.12. HDHR5 found in Entlebucher Mountain dogs encompasses a gene implicated in reduced fertility..... | 50 |
| 3.13. HDHR6, HDHR7, HDHR11, HDHR12 and HDHR14 may be linked with homozygous-lethal mutations..... | 50 |
| 4. Discussion..... | 51 |
| 5. Conclusion..... | 52 |
| 6. Funding..... | 52 |
| 7. Acknowledgments | 53 |
| 8. References | 53 |
| APÊNDICES | 59 |
| APÊNDICE A - Capítulo 2 Supplementary Figures..... | 60 |
| APÊNDICE B - Capítulo 3 Supplementary Figures..... | 63 |

VARREDURAS GENÔMICAS PARA DETECÇÃO DE VARIANTES GENÉTICAS ASSOCIADAS À REPRODUÇÃO DE CÃES

RESUMO - Variante funcional (*functional variant* - FV) é um polimorfismo da sequência de DNA que possui efeitos sobre a função ou mesmo sobre o nível de expressão de um ou mais genes. Por apresentarem alta probabilidade de interferência em fenótipos celulares, teciduais ou mesmo sistêmicos, as FVs são alvos frequentes de pressão de seleção negativa ou positiva em uma população. Algumas FVs podem, inclusive, causar perda parcial ou total da função gênica, caso no qual estas recebem a denominação de *loss-of-function variant* (LoFV). Os cães domésticos (*Canis lupus familiaris*) possuem baixa variabilidade genética decorrente da intensa seleção artificial feita pelo homem para fenótipos extremos de características morfológicas e comportamentais. Os altos níveis de endogamia nas populações caninas favorecem a intensificação da ação da deriva genética e, por consequência, o aumento da prevalência de FVs e LoFVs que seriam mantidas em baixa frequência se estas populações apresentassem maior variabilidade genética. Assim, o cão doméstico é um modelo valioso para estudos de processos fisiológicos e de doenças genéticas humanas. O recente surgimento de ferramentas para análise genômica têm auxiliado na identificação de FVs e LoFVs de interesse zootécnico e biomédico em cães. Estas variantes podem ser mapeadas através de duas estratégias distintas, porém complementares, de varredura genômica. A primeira, denominada de estudo de associação genômica ampla (*Genome-Wide Association Study* - GWAS), consiste no teste de associações entre genótipos e fenótipos. A segunda consiste na busca por combinações alélicas, chamadas de haplótipos, que ocorram em alta frequência, porém que apresentem deficiência de homozigose. No presente trabalho, utilizamos ambas as estratégias no mapeamento de FVs e LoFVs de alto impacto funcional associadas à reprodução de cães. No primeiro estudo, um GWAS para tamanho de ninhada foi conduzido em Cães Boiadeiros de Entlebuch, o qual revelou uma FV no gene do fator de crescimento e diferenciação 9 (growth/differentiation factor 9 – *GDF9*). No segundo estudo, reportamos haplótipos de alta frequência com deficiência de homozigotos nas raças Cão da Crista Chinesa, Cão Boiadeiro de Entlebuch, Labrador Retriever e Golden Retriever. Ambas as abordagens permitiram a identificação de variantes que podem estar associadas a prejuízos reprodutivos.

Palavras chave: Polimorfismo de sítio único (SNP), haplótipo, variante funcional, reprodução, *Canis lupus familiaris*

GENOME-WIDE SCANS FOR THE DETECTION OF GENETIC VARIANTS ASSOCIATED WITH REPRODUCTIVE PERFORMANCE IN DOGS

ABSTRACT- Functional variant (FV) is a DNA sequence polymorphism that has effects on the function or even the level of expression of one or more genes. FVs are frequent targets of negative or positive selection in a population due to its high probability of interfering in cell, tissue or system phenotypes. Some FVs may even cause partial or total loss of gene function, in which case they are called loss-of-function variants (LoFVs). Domestic dogs (*Canis lupus familiaris*) have low genetic variability resulting from intense human-driven artificial selection for extreme phenotypes of morphological and behavioral traits. High levels of inbreeding in canine populations favor the intensification of genetic drift and, consequently, an increase in the prevalence of FVs and LoFVs that would be maintained at low frequencies if these populations presented higher genetic variability. Thus, the domestic dog is a valuable model for studying physiological processes and human genetic diseases. The recent emergence of tools for genomic analysis has assisted in the identification of FVs and LoFVs of zotechnical and biomedical interest in dogs. These variants can be mapped through two distinct but complementary genomic scanning strategies. The first, known as Genome-Wide Association Study (GWAS), involves testing of phenotype-genotype associations. The second consists in the search for allelic combinations, called haplotypes, that occur in high frequency, but that present deficiency of homozygosity. In the present work, we used both strategies in the mapping of FVs and LoFVs of high functional impact associated with reproduction in dogs. In the first study, a GWAS for average litter size was conducted in Entlebucher Mountain dogs, which revealed a FV in the growth differentiation factor 9 gene (*GDF9*). In the second study, we reported high frequency haplotypes with deficiency of homozygosity in the Chinese Crested, Entlebucher Mountain, Labrador Retriever and Golden Retriever dog breeds. Both approaches allowed the identification of variants that may be associated with reproductive performance.

Keywords: Single nucleotides variants (SNV), haplotype, functional variant, reproduction, *Canis lupus familiaris*

LISTA DE ABREVIÇÕES

| | |
|-----------------|--|
| AI-REML | Average Information REstricted Maximum Likelihood |
| ALS | Average litter size |
| <i>AMY2B</i> | Pancreatic alpha-amylase 2b gene |
| <i>ARHGAP27</i> | Rho GTPase activating protein 27 gene |
| <i>BMP15</i> | Bone morphogenetic protein 15 gene |
| bp | Base pairs |
| BWA | Burrows-Wheeler Alignment |
| CanFam3.1 | Genoma de referência <i>Canis familiaris</i> 3.1 |
| <i>CBD103</i> | Canine β -defensin gene |
| CC | Chinese Crested |
| CFA | <i>Canis familiaris</i> autosome |
| <i>CNOT3</i> | CCR4-NOT transcription complex subunit 3 gene |
| CNV | Copy number variation |
| <i>COL11A1</i> | Collagen type XI alpha 1 chain gene |
| <i>COL25A1</i> | Collagen type XXV alpha 1 chain gene |
| <i>CRTC3</i> | CREB regulated transcription coactivator 3 gene |
| DNA | Ácido Desoxirribonucleico |
| EM | Entlebucher Mountain |
| EM-REML | Expectation Maximization REstricted Maximum Likelihood |
| F _L | Locus autozygosity |
| <i>FOXI3</i> | Forkhead box I3 gene |
| FV | Functional Variant |
| GBLUP | Genomic Best Linear Unbiased Prediction |
| GC | Genomic Control |
| <i>GDF9</i> | growth differentiation factor 9 gene |
| GR | Golden Retriever |
| GWAS | Genome-Wide Association Study |
| HDHR | Homozygous-deficient regions |

| | |
|---------------|--|
| HMM | Hidden Markov Model |
| HURP | Highly unfavorable recessive phenotype |
| HWE | Hardy-Weinberg Equilibrium |
| IBD | Identical-by-descent |
| IBS | Identical-by-state |
| INDELS | Insertions-deletions |
| LD | Linkage disequilibrium |
| lincRNA | Long intergenic non-coding RNA |
| LoFV | Loss of function variant |
| LR | Labrador Retriever |
| Mc1r | Melanocortin receptor 1 |
| <i>MC1R</i> | Melanocortin receptor 1 gene |
| MGI | Mouse Genome Informatics |
| <i>NCAPH2</i> | Non-SMC condensin II complex subunit H2 gene |
| Ne | Tamanho efetivo de população |
| OMIA | Online Mendelian Inheritance in Animals |
| OMIM | Online Mendelian Inheritance in Men |
| p.Cys90Tyr | Substituição de um aminoácido Cisteína por Tirosina no 90º resíduo da proteína |
| p.Pro77Ser | Substituição de um aminoácido Prolina por Serina no 77º resíduo da proteína |
| PCR | Polymerase Chain Reaction |
| <i>PDE6D</i> | Phosphodiesterase 6D gene |
| PRA | Progressive retinal atrophy |
| <i>PRPF31</i> | Pre-mRNA processing factor 31 homolog gene |
| QTL | Loco de característica quantitativa |
| REML | REstricted Maximum Likelihood |
| <i>RNPC3</i> | RNA binding region containing 3 gene |
| ROH | Corrida de homozigose |
| SAG | S-antigen visual arrestin gene |
| SHAPEIT2 | Segmented HAPlotype Estimation & Imputation Tool |

| | |
|---------------|--|
| SINE | Short interspersed nuclear elements |
| SMAD | Sma mothers against decapentaplegic |
| <i>SMC3</i> | Structural maintenance of chromosomes 3 gene |
| SNV | Variante de sítio único |
| <i>SYCE3</i> | Synaptonemal complex central elemento protein 3 gene |
| <i>TBX15</i> | T-box 15 gene |
| TGF- β | Transforming growth factor-beta |
| WGS | Whole-genome sequences |
| α -MSH | Alpha melanocyte stimulating hormone |
| Δ G23 | Deletion of codon 23, encoding glycine |

LISTA DE TABELAS

| | |
|---|----|
| CAPÍTULO 3: High frequency canine haplotypes with exceptional deficiency of homozygosity | 28 |
| Table 1. Sources of Illumina® CanineHD genotypes used in this study..... | 31 |
| Table 2. Observed-to-expected ratios and <i>p</i> -values (inside brackets) for homozygous-deficient haplotype region (HDHR)..... | 38 |

LISTA DE FIGURAS

| | |
|---|----|
| CAPÍTULO 2: Association of a missense variant in <i>GDF9</i> with decreased average litter size in Entlebucher Mountain dogs | 12 |
| Figure 1. Genome-wide association scan for major loci affecting litter size in Entlebucher Mountain dogs..... | 19 |
| Figure 2. Effect and conservation of the detected missense variant (g.11:21147009G>A, p.Pro77Ser) in the growth differentiation factor 9 gene (<i>GDF9</i>) | 21 |
| CAPÍTULO 3: High frequency canine haplotypes with exceptional deficiency of homozygosity | 28 |
| Figure 1. Sample size required to detect homozygous-deficient haplotypes (HDH) as a function of their frequency..... | 36 |
| Figure 2. Manhattan plots of genome-wide screenings for homozygous-deficient haplotypes of length 100 kbp..... | 37 |
| Figure 3. Detection of a homozygous-deficient haplotype (HDH) tagging the homozygous-lethal mutation of <i>FOXI3</i> in Chinese Crested (CC) dogs..... | 43 |
| Figure 4. Proposed model for epistatic interactions between <i>MC1R</i> and <i>CBD103</i> mutations in dogs..... | 45 |
| APÊNDICES | 59 |
| Figure 1A. Histograms and summary statistics of average litter size and number of observed litters for 892 females of the Entlebucher Mountain dog breed..... | 60 |
| Figure 2A. Convergence plot of standard errors averaged across 87855 SNV markers | 61 |
| Figure 3A. Genotype status at deletions near <i>SMAD2</i> in eight Entlebucher Mountain dogs..... | 62 |
| Figure 1B. Genomic autozygosity estimated with runs of homozygosity larger than 1 Mbp..... | 63 |
| Figure 2B. Genomic autozygosity estimated with runs of homozygosity larger than 2 Mbp..... | 63 |
| Figure 3B. Genomic autozygosity estimated with runs of homozygosity larger than | |

| | |
|--|----|
| 4 Mbp..... | 64 |
| Figure 4B. Genomic autozygosity estimated with runs of homozygosity larger than 8 Mbp..... | 64 |
| Figure 5B. Genomic autozygosity estimated with runs of homozygosity larger than 16 Mbp..... | 65 |
| Figure 6B. Manhattan plot of genome-wide screenings for homozygous-deficient haplotypes with window sizes of 500 kbp | 65 |
| Figure 7B. Manhattan plot of genome-wide screenings for homozygous-deficient haplotypes with window sizes of 1 Mbp..... | 66 |
| Figure 8B. Manhattan plot of genome-wide screenings for homozygous-deficient haplotypes with window sizes of 2 Mbp..... | 66 |
| Figure 9B. Manhattan plot of genome-wide screenings for homozygous-deficient haplotypes with window sizes of 5 Mbp..... | 67 |

CAPÍTULO 1 - Considerações gerais

1. A seleção artificial promoveu o aumento da frequência de variantes de alto impacto funcional no genoma canino

O cão doméstico (*Canis lupus familiaris*) foi a primeira espécie animal a ser domesticada pelo homem, entre 20 e 40 mil anos atrás (Pendleton et al., 2018). Nos últimos três séculos, a intensa seleção artificial para fenótipos extremos de características morfológicas e comportamentais deu origem às mais de 400 raças catalogadas atualmente (Ostrander et al., 2017). Esta intensa seleção favoreceu a redução da diversidade genética e consequente intensificação da ação da deriva genética, favorecendo o aumento da frequência de variantes de alto impacto funcional no genoma canino (Xue et al., 2012; Marsden et al., 2016).

Deriva genética (*genetic drift*) é um mecanismo de evolução neutra que consiste na flutuação aleatória das frequências alélicas. Um alelo pode tornar-se prevalente ou extinguir-se da população por meio da deriva genética por consequência de fenômenos como gargalos populacionais, redução da variabilidade genética ou endogamia. O arrasto genético (*genetic draft*) é caracterizado por aumento da frequência de alelos neutros que pegam “carona” (*hitchhiking*) com alelos vizinhos que estão sob pressão de seleção. Alguns alelos podem ainda ter efeitos pleiotrópicos, que indicam que podem afetar mais de uma característica, podendo ser uma delas vantajosa e a outra desvantajosa.

Variante funcional (*functional variants- FV*) é definida como um polimorfismo da sequência de DNA capaz de afetar a função ou o nível de expressão de um ou mais genes. Algumas FVs podem, inclusive, causar perda parcial ou total da função gênica, caso no qual estas recebem a denominação específica de *loss-of-function variant* (LoFV) (Pagel et al., 2017). Por apresentarem alta probabilidade de interferência em fenótipos celulares, teciduais ou mesmo sistêmicos, tanto as FVs como as LoFVs são alvos frequentes de pressão de seleção negativa em uma população (Boyle et al., 2017; Zeng et al., 2018). A análise de genomas completos de milhares de seres humanos revelou que, em média, cada pessoa hígida é portadora de aproximadamente 100 LoFVs, cada

uma destas apresentando baixa frequência populacional e, em muitos casos, sendo privativas (Macarthur et al., 2012). Muitas LoFVs não induzem diferenças fenotípicas significativas quando em heterozigose, porém, a herança de duas cópias de uma LoFV pode gerar fenótipos extremos ou até ser incompatíveis com a vida. Assim, a diminuição da variabilidade genética de uma população, seja por meio de endogamia, deriva genética ou gargalo populacional, pode favorecer o aumento da frequência e a eventual homozigose destas variantes.

O genoma canino (*Canis lupus familiaris*) apresenta alta ortologia com o genoma humano, fazendo do cão doméstico um excelente modelo para o estudo de processos fisiológicos e patológicos (Karlsson e Lindblad-toh, 2008). Apesar da grande diversidade morfológica observada nos cães (Akey et al., 2010), o processo de formação de raças foi caracterizado pelo uso de pequeno número de animais fundadores (Parker et al., 2004; 2017) e drásticos gargalos populacionais proporcionados pela seleção humana (Lindblad-toh et al., 2005, Ostrander et al., 2017). Conseqüentemente, as diversas raças caninas apresentam reduzida variabilidade genética em comparação com outras espécies (Vila et al., 1999), corroborando para o aumento indireto da frequência de FVs e LoFVs e da ocorrência de doenças genéticas nessa espécie (Cruz et al., 2008; Ostrander, 2012; Parker et al., 2010).

A endogamia em cães, além de ser responsável pelo aumento da frequência de doenças genéticas e perdas embrionárias, também afeta drasticamente o desempenho reprodutivo da espécie. Alterações como criptorquidismo (Smit et al., 2018), doenças uterinas (Voorwald, 2015), tamanho de ninhada reduzidos e problemas de acasalamentos (Schrack et al., 2017) são constantemente encontrados dentro de canis. Cães da raça Boiadeiro de Entlebuch, por exemplo, possuem alta consanguinidade devido ao seu baixo número de fundadores e gargalo populacional decorrente da Segunda Guerra Mundial. Por consequência, nos últimos anos houve uma queda de eficiência de acasalamentos e uma redução do número médio de filhotes por ninhada na raça (Schrack et al., 2017).

2. Ferramentas genômicas podem auxiliar na caracterização de variantes funcionais associadas à reprodução em cães

Na última década, a finalização do genoma canino de referência, juntamente à descoberta de polimorfismos de sítio único (*single nucleotide variants* – SNVs ou *single nucleotide polymorphisms*- SNPs) comuns entre as raças caninas e ao surgimento de microarranjos de DNA com paralelismo massivo (SNV chips), viabilizou o processo de mapeamento de regiões cromossômicas envolvidas em diferenças fenotípicas em cães (Lindblad-toh et al., 2005; Karlsson et al., 2007). Estes avanços na área de genômica impulsionaram inúmeros estudos populacionais de associação entre variantes genômicas e características de interesse em cães (Pfhaler e Distl, 2012; Philipp et al., 2012; Seppälä et al., 2012, Hayward et al., 2016).

Marcadores SNV podem auxiliar no mapeamento de FVs associados a características reprodutivas em cães por meio de duas estratégias distintas porém complementares: i) através da busca de combinações alélicas (i.e., haplótipos) com moderada ou alta frequência que raramente ocorrem em homozigose (Vanraden et al., 2011); ou ii) por meio da busca de associações entre fenótipo e genótipo (Bush e Moore, 2012). Em ambos os casos, a correlação não aleatória entre variantes próximas em um segmento cromossômico (i.e., desequilíbrio de ligação) permite a captura indireta da informação de FVs não observadas da sequência de DNA por meio de alta densidade de marcadores SNV. Por sua vez, as coordenadas genômicas dos SNVs viabilizam o sequenciamento alvo da região afetada para posterior caracterização da FV. Portanto, a aplicação destas abordagens em cães podem auxiliar i) no entendimento de enfermidades e processos fisiológicos nos animais de companhia; ii) na geração de informação para o desenvolvimento de ferramentas diagnósticas; iii) no desenho e potencialmente personalização de condutas terapêuticas; iv) no aconselhamento genético para controle da incidência de casos; e, por fim, v) na elucidação de processos similares em humanos.

De fato, atualmente existem 719 enfermidades genéticas descritas em cães, acometendo praticamente todos os sistemas orgânicos, sendo que apenas 243 possuem

causa conhecida e 420 como modelo para doenças humanas (*Online Mendelian Inheritance in Animals*- OMIA, disponível em: <http://omia.angis.org.au/home/>). Dentre essas anotações catalogadas, algumas são associadas a perdas embrionárias e/ou a redução do tamanho de ninhada. Um exemplo, seria o Cão da Crista Chinesa, raça com marcante fenótipo de alopecia corporal (i.e. promovida por displasia ectodérmica). Este fenótipo está associado a uma LoFV descrita no gene *Forkhead Box I3 (FOXI3)* (Drögemüller et al., 2014). Ao mesmo tempo, esta LoFV em estado homozigoto leva à morte embrionária, sendo sempre recomendada a reprodução entre animais “powderpuff” (peludos) e animais heterozigotos para a LoFV do *FOXI3* (alopécicos). Por simples herança Mendeliana, o acasalamento entre cães alopécicos (heterozigotos) gera morte de 25% dos embriões produzidos. A LoFV do *FOXI3* foi encontrada através de estudos de associação genômica ampla (*Genome-wide association study*- GWAS) por Drögemüller e colaboradores (2014), mas também poderia ser identificada através de busca por haplótipos de alta frequência com ausência de homozigotos, já que estes morrem em estado embrionário.

Apesar de todas as raças de cães possuírem altos índices de consanguinidade (endogamia), como já descrito anteriormente, algumas raças de cães são mais endogâmicas que outras (Boyko et al., 2010). Esta observação pode ser revelada por análises de tamanho efetivo de população (*Effective Population Size* - N_e) e corridas de homozigose (*Runs of Homozygosity* - ROH). O tamanho efetivo de população (N_e) é um parâmetro importante na genética populacional, sendo utilizado como ferramenta na conservação e quantificação de variabilidade genética (Barbato et al., 2015; Husemann et al., 2016). Já a análise de ROH permite avaliar segmentos genômicos homozigotos ao longo dos cromossomos (Mcquillan et al., 2008; Curik et al., 2014). Blocos homozigotos de grande extensão estão geralmente associados com endogamia recente, enquanto que blocos menores são relacionados à endogamia ou fixação de haplótipos herdados de ancestrais remotos (Mcquillan et al., 2008; Boyko et al., 2010).

3. Variantes funcionais associadas à reprodução através de deficiência de homozigotos

Como discutido anteriormente, a deriva genética favorece o aumento da frequência de LoFVs em populações com baixa variabilidade genética, tornando-as suscetíveis à incidência de variantes deletérias em estado homozigoto. No entanto, ela não é o único fenômeno capaz de manter LoFVs em alta frequência em uma população. Algumas variantes deletérias podem apresentar pleiotropia, sendo mantidas em alta frequência por simultaneamente causarem um fenótipo positivamente selecionado. Um exemplo típico, é o já citado Cão da Crista Chinesa, no qual a LoFV do *FOXI3* é deletéria em homozigose, causando morte embrionária, e ao mesmo tempo responsável pelo fenótipo de alopecia corporal em heterozigose, o qual é positivamente selecionado (Drögemüller et al., 2014). Outro fenômeno é o arrasto genético, no qual a LoFV pode estar em desequilíbrio de ligação com uma outra FV positivamente selecionada. Independente do fenômeno de manutenção de LoFVs, espera-se que haplótipos carregadores de LoFVs desfavoráveis em homozigose apresentem ausência ou baixa frequência de homozigotos na genotipagem de número suficientemente grande de indivíduos (VanRaden et al., 2011). A metodologia requer apenas genótipos de indivíduos não afetados, não sendo necessários dados fenotípicos de perdas embrionárias.

A detecção indireta de variantes com impacto funcional que causam fenótipos recessivos tem se mostrado viável em outras espécies através da detecção de haplótipos com deficiência de homozigotos (VanRaden et al., 2011; Sahana et al., 2013; Hoff et al., 2017; Häggman e Uimari, 2017). O método inicia-se com a determinação de haplótipos em janelas cromossômicas de tamanhos pré-definidos. Um teste de hipótese é aplicado para cada haplótipo para comparar o número observado de homozigotos com o que seria esperado em Equilíbrio de Hardy-Weinberg (*Hardy-Weinberg Equilibrium*- HWE). Haplótipos com número de homozigotos observados muito abaixo do esperado são então considerados candidatos a albergar variantes que causam fenótipos recessivos desfavoráveis. A técnica tem sido particularmente útil na predição de variantes que provocam morte embrionária, uma vez que esta é de difícil mensuração e raramente

observada. Considerando a existência de ferramentas genômicas de alta densidade em cães, como o SNV chip, esta estratégia pode ser aplicada ao genoma canino de modo a detectar variantes de alta frequência e de alto impacto funcional nesta espécie.

O mapeamento de haplótipos com ausência de homozigotos foi aplicado com sucesso na identificação de variantes letais em bovinos (VanRaden et al., 2011; Fritz et al., 2013; Kadri et al., 2014). Nesta espécie, a semelhança do que ocorreu em cães, a forte pressão de seleção para aumento de produção contribuiu para o acúmulo de LoFVs, as quais são responsáveis por expressiva parcela de falhas reprodutivas e, conseqüentemente, perdas econômicas na pecuária (VanRaden et al., 2011; Kadri et al., 2014).

Por exemplo, VanRaden e colaboradores (2011) buscaram haplótipos com ausência de homozigotos em 65.732 bovinos de três raças diferentes genotipados para cerca de 50.000 SNVs. No estudo, onze haplótipos candidatos foram identificados, entre os quais esperavam-se sob equilíbrio de Hardy-Weinberg de 7 a 90 indivíduos homozigotos, sendo que nenhum foi observado. Estes achados culminaram na posterior caracterização da variante causal de um destes haplótipos, por meio de dados de sequenciamento (Sonstegard et al., 2013), e o desenvolvimento de ferramentas diagnósticas para direcionar acasalamentos nestes rebanhos (http://aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html).

Apesar do evidente acúmulo de LoFVs no genoma canino (Marsden et al., 2016), estudos semelhantes em cães ainda não foram conduzidos. Comparado a outras espécies, o cão doméstico sofreu gargalos populacionais mais intensos, requerendo assim tamanhos amostrais de magnitude menores em estudos genômicos (Karlsson et al., 2007).

4. Variantes com alto impacto funcional por meio de análise de associação genômica ampla em cães

Estudos de GWAS surgiram como uma ferramenta eficaz para identificar regiões cromossômicas associadas a variações fenotípicas (Chung e Chanock, 2011). Loci de características quantitativas (*Quantitative trait loci*- QTLs) já foram detectados para diversas características herdáveis em cães, como por exemplo condrodisplasia (Parker et al., 2009), variações de caudas e orelhas (Vaysse et al., 2011), braquicefalia (Bannasch et al., 2010), diferenças de pelagens (Cadieu et al., 2009), características reprodutivas (Inanç et al., 2018), entre diversos outros (Hayward et al., 2016; Machiela e Chanock, 2014).

Da mesma forma que a abordagem de haplótipos com deficiência de homozigotos, este método com a utilização de fenótipos visa a identificação de regiões e possíveis variantes genéticas candidatas associadas ao fenótipo fornecido. A utilização destas ferramentas e confirmação de variantes com impacto funcional viabilizam o desenvolvimento de um painel de diagnóstico molecular para a identificação de animais carreadores dentro de canis. Este tipo de ferramenta pode auxiliar no direcionamento de acasalamentos com maior eficiência e produtividade.

5. Objetivos do presente trabalho

Com base nas informações descritas, o segundo capítulo desta tese objetivou a utilização da análise de GWAS para a identificação de variantes genéticas relacionadas ao tamanho de ninhada em Cães Boiadeiros de Entlebuch, tendo em vista que esta raça apresentou recentes problemas reprodutivos, com perdas de filhotes, acasalamentos malsucedidos e redução do tamanho médio de ninhadas. O terceiro capítulo, por sua vez, apresenta a abordagem de busca por variantes com impacto funcional por intermédio de ausência ou deficiência de homozigotos em cães das raças Cão da Crista Chinesa, Boiadeiro de Entlebuch, Golden Retriever e Labrador Retriever.

6. Referências

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW (2010) Tracking footprints of artificial selection in the dog genome. **PNAS** 107:1160-1165.

Bannasch D, Young A, et al (2010) Localization of canine brachycephaly using an across breed mapping approach. **PLoS ONE** 5:e9632.

Barbato M, Orozco-terwengel P, Tapio M, Bruford W (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. **Front Genet** 6:109.

Boyle EA, Li YI, Pritchard JK (2017) An expanded View of Complex Traits: From Polygenic to Omnigenic. **Cell** 169:1177-1186.

Boyko AR, Quignon P, et al (2010) A Simple Genetic Architecture Underlies Morphological Variation in Dogs. **PLoS Biol** 8:e1000451.

Bush WS, Moore JH (2012) Chapter 11: Genome-Wide association studies. **PLoS Comput Biol** 8:e1002822.

Cadiou E, Neff MW, et al (2009) Coat variation in the domestic dog is governed by variants in three genes. **Science** 326:150-153.

Chung CC, Chanock SJ (2011) Current status of genome-wide association studies in cancer. **Hum Genet** 130:59-78.

Cruz F, Vilà C, Webster MT (2008) The legacy of domestication: accumulation of deleterious mutations in the dog genome. **Mol. Biol. Evol** 25:2331-2336.

Curik I, Ferenčaković M, Sölkner J (2014) Inbreeding and runs of homozygosity: A possible solution to an old problem. **Livestock Science** 166:26-34.

Drögemüller C, Karlsson EK, Hytönen MK, Perloski M, Dolf G, Sainio K, Lohi H, Lindblad-toh K, Leeb T (2008) A mutation in hairless dogs implicates *FOXI3* in ectodermal development. **Science** 321:1462.

Fels L, Distl O (2014) Identification and validation of Quantitative Trait Loci (QTL) for Canine Hip Dysplasia (CHD) in German Shepherd Dogs. **PLoS ONE** 9:e99618.

Fritz S, Capitan A, et al (2013) Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutation in *HART*, *SHBG* and *SLC37A2*. **PLoS ONE** 8:e65550.

Hayward JJ, Castelhana MG, et al (2016) Complex disease and phenotype mapping in the domestic dog. **Nat. Commun** 7:10460.

Häggman J, Uimari P (2017) Novel harmful recessive haplotypes for reproductive traits in pigs. **J Anim Breed Genet** 132:129-135.

Hoff JL, Decker JE, Schnabel RD, Taylor JF (2017) Candidate lethal haplotypes and causal mutations in Angus cattle. **BMC Genomics** 18:799.

Husemann M, Zachos FE, Paxton RJ, Habel JC (2016) Effective population size in ecology and evolution. **Heredity** 177:191-192.

Inanç ME, Tekin K, et al (2018) Genome-wide association of male reproductive traits in Aksaray Malakli dogs. **Reprod Dom Anim** 53:1555-1562.

Kadri NK, Sahana G, et al (2014) A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: Additional evidence for the common occurrence of balancing selection in livestock. **PLoS ONE** 10:e1004049.

Karlsson EK, Baranowska I, et al (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. **Nat Genet** 39:1321- 1328.

Karlsson EK, Lindblad-toh K (2008) Leader of the pack: gene mapping in dogs and other model organisms. **Nature Genetics** 9:713- 725.

Lindblad-toh K, Wade CM, et al (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. **Nature** 438:804-819.

Macarthur DG, Balasubramanian S, et al (2012) A systematic survey of loss-of-function variants in human protein-coding genes. **Science** 335:823-828.

Machiela MJ, Chanock SJ (2014) GWAS is going to the dogs. **Genome Biol** 15:105.

Marsden C, Johnson GS, Wayne R (2015) Population bottlenecks led to the accumulation of deleterious variants in domestic and wild canids. In: PLANT AND ANIMAL GENOME XXIII CONFERENCE (PAG), 2015. **Anais eletrônicos**. Disponível em: <<https://pag.confex.com/pag/xxiii/webprogram/Paper16185.html>>. Acesso em 21 de ago. 2018.

Marsden CD, Ortega-del Vecchyo D, O'brien DP, Taylor JF, Ramirez O, Vilà C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE (2016) Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. **PNAS** 113:152-157.

Mcquillan R, Leutenegger AL, et al (2008) Runs of homozygosity in European populations. **Am J Hum Genet** 83:359-372.

Ostrander EA, Wayne RK, Freedman AH, Davis BW (2017) Demographic history, selection and functional diversity of the canine genome. **Nat Rev Genet** 18:705-720.

Ostrander EA (2012) Both ends of the leash- The human links to good dogs with bad genes. **N. Engl. J. Med** 367:636-346.

Pagel KA, Pejaver V, Lin GN, Nam H, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD, Radivojac P (2017) When loss-of-function in loss of function: assessing mutational signatures and impact of loss-of-function genetic variants. **Bioinformatics** 33:i389-i398.

Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, Ostrander EA (2017) Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. **Cell Reports** 19:697-708.

Parker HG, Vonholdt BM, et al (2009) An expressed *FGF4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. **Science** 325:995-998.

Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, Defrance HB, Ostrander EA, Kruglyak L (2004) Genetic structure of the purebred domestic dog. **Science** 304:1160-1664.

Parker HG, Shearin AL, Ostrander EA (2010) Man's best friend becomes biology's best in show: Genome analysis in the domestic dog. **Annu. Rev. Genet** 44:309-336.

Pendleton AL, Shen F, Taravella AM, Emery S, Veeramah KR, Boyko AR, Kidd JM (2018) Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. **BMC Biology** 16:64.

Pfahler S, Distl O (2012) Identification of Quantitative Trait Loci (QTL) for canine hip dysplasia and canine elbow dysplasia in Bernese Mountain Dogs. **PLoS ONE** 7:e49782.

Philipp U, Vollmar A, Hänggström J, Thomas A, Distl O (2012) Multiple loci are associated with dilated cardiomyopathy in Irish Wolfhounds. **PLoS ONE** 7:e36691.

Sahana G, Nielsen US, Aamand GP, Lund MS, Guldbrandtsen B (2013) Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. **PLoS ONE** 8: e82909.

Seppälä EH, Koskinen LL, et al (2012) H. Identification of novel idiopathic epilepsy locus in Belgian Shepherd dogs. **PLoS ONE** 7:e33549.

Sonstegard TS, Cole JB, VanRaden PM, Curtis P, Van Tassell CP, Null DJ, Schroeder SG, Bickhart D, McClure MC (2013) Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. **PLoS ONE** 8:e54872.

Schrack J, Dolf G, Reichler LM, Schelling C (2017) Factors influencing litter size and puppy losses in the Entlebucher Mountain dog. **Theriogenology** 95:163-170.

Smit MM, Ekenstedt KJ, Minor KM, Lim CK, Leegwater PAJ, Furrow E (2018) Prevalence of the *AMHR2* mutation in Miniature Schnauzers and genetic investigation of a Belgian Malinois with persistent Müllerian duct syndrome. **Reprod Dom Anim** 53:371-376.

VanRaden PM, Olson KM, Null DJ, Hutchison JL (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. **J. Dairy Sci** 94:6153-6161.

Vaysse A, Ratnakumar A, et al (2011) Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. **PLoS Genet** 7:e1002316.

Vilà C, Maldonado JE, Wayne RK (1999) Phylogenetic relationships, evolution and genetic diversity of the domestic dog. **J. Hered** 90:71-77.

Voorwald FA, Marchi FA, Villacis RA, Alves CE, Toniollo GH, Amorim RL, Drigo SA, Rogatto SR (2015) Molecular expression profile reveals potential biomarkers and therapeutic targets in canine endometrial lesions. **PLoS ONE** 10:e0133894.

Xue Y, Chen Y, et al (2012) Deleterious- and disease- allele prevalence in healthy individuals: Insights from current predictions, mutation database, and populations- scale resequencing. **Am J Hum Genet** 91:1022-1032.

Zeng J, Vlaming R, et al (2018). Signatures of negative selection in the genetic architecture of human complex traits. **Nature Genetics** 50:746-753.

CAPÍTULO 2- Association of a missense variant in *GDF9* with decreased average litter size in Entlebucher Mountain dogs

Rafaela Beatriz Pintor Torrecilha; Marco Milanese; Milena Gallana; Sarah Sugiato; Ann-Kritin Besold; Iris Reichler; Vidhya Jagannathan; Gaudenz Dolf; Tosso Leeb; Johann Sölkner; José Fernando Garcia; Claude Schelling; Yuri Tani Utsunomiya

ABSTRACT- The Entlebucher Mountain (EM) dog is a Swiss canine breed that experienced a rapid decrease in average litter size (ALS) in the last decade, suggesting genetic drift of variants with major effects on the trait. Here, we conducted a genome-wide association study (GWAS) for ALS using the single-step methodology to take advantage of 1632 pedigree records, 892 phenotypes and 372 genotypes (173662 markers) for which only a small percentage of the dogs (~12%) had both phenotypes and genotypes available. Our analysis revealed genome-wide significant ($p < 5.69 \times 10^{-7}$) associations mapping to the vicinity of the growth differentiation factor 9 gene (*GDF9*), which encodes for a ligand of the transforming growth factor-beta (TGF- β) superfamily that regulates the expansion and differentiation of cumulus cells and promotes ovarian follicle growth. The trait heritability was estimated at 43.1%, much higher than in other dog breeds, from which ~15% was accountable by the *GDF9* locus alone. Therefore, markers spanning the *GDF9* region explained around 6.5% of the variance in ALS in EM dogs, opening a tangible opportunity to improve fertility in the breed. Analysis of whole genome sequences of 8 EM dogs revealed a novel *GDF9* missense substitution (g.11:21147009G>A) affecting a highly conserved nucleotide in vertebrates. The derived allele A was associated with decreased ALS, and was predicted to cause a p.Pro77Ser amino acid change to the *GDF9* protein. Intriguingly, residue 77 was immediately followed by a 6-residues deletion that is fixed in the canine species but absent in non-canids. Moreover, given that canids uniquely ovulate oocytes at the Metaphase II stage of the first meiotic division, requiring maturation in the oviduct, we conjecture that the p.Pro77Ser substitution and the 6-residues deletion of *GDF9* may serve as a model for future insights into the dynamics of ovarian follicle growth and oocyte maturation in canids.

Keywords: *Canis lupus familiaris*, number of offspring, multiple ovulation, single nucleotide polymorphism, single-step GBLUP

1. Introduction

The Entlebucher Mountain (EM) dog is the smallest and youngest of the four Alpine dog breeds (NEMDA, 2018). Inbreeding depression is a major concern in the breed, since it suffered two important historical population bottlenecks: (i) one in its formation about 100 years ago, which relied on a small number of founders; and (ii) another during World War I and II, when its development was put on hold. Inbreeding in the EM population kept increasing slowly and constantly since, which contributed to a recent decline in the breed's average litter size (Schrack et al., 2017). These observations tease the hypothesis of recent genetic drift of variants with large effects on the trait, which could be prospected through a genome-wide association study (GWAS).

As the EM population size is small, conducting a standard GWAS analysis for average litter size (ALS) may be difficult since the ideal sample would comprise phenotypes and genotypes from hundreds of females, each with at least one litter observed. Currently, these data exist only in a fragmented way. From the studbooks of the Swiss national kennel club, genealogy and ALS data are recoverable for several hundreds of registered females. However, sources of DNA may be only available for a small percentage of these dogs. Access to DNA samples from living EM dogs is fairly easy, but the number of living females with recorded litters is limited. In order to counteract the discontinuity between DNA and phenotype availability, pedigree relationships could be used to link genotypes from either males or females to phenotypes from non-genotyped females. This type of strategy has been developed for the analysis of complex traits in livestock, where the combination of large sets of phenotypes with smaller sets of genotypes is a frequent scenario (Misztal et al., 2009; Aguilar et al., 2010). The methodology used to that end is known as single-step GBLUP, which is capable of fitting a linear mixed model to a heterogeneous data set including pedigree, genotypes and phenotypes. The statistical model estimates the heritability of the trait and the aggregate of additive genetic effects of each animal, which can then be converted into marker effects through a procedure called single-step GWAS (Wang et al., 2012; 2014; Zhang et al., 2016).

Here, we conducted a single-step GWAS analysis of ALS using 1632 pedigree records, 892 phenotypes and 372 Illumina® CanineHD genotypes (173662 markers). Combined with whole-genome sequences (WGS) of 8 dogs, these data revealed a missense variant in the growth differentiation factor 9 gene (*GDF9*) associated with decreased ALS in EM dogs.

2. Material and Methods

2.1 Summary of the data

Data from 1632 EM dogs were available for analysis, from which 111 had both phenotypes and genotypes, 261 had only genotypes, 781 had only phenotypes and 479 did not have genotypes or phenotypes but were revealed by screening of 5556 pedigree records as relatives of animals with phenotypes. These data were obtained from a previous report (Schrack et al., 2017), which were complemented with WGS from 8 EM dogs. The distributions of ALS and number of recorded litters per female for the 892 phenotyped dogs are presented in **Apêndices- Figure 1A**.

2.2 Genotype filtering

Genotypes for 177 males and 195 females were derived from the Illumina® CanineHD BeadChip assay, which included 173662 single nucleotide variant (SNV) markers. These genotypes were previously generated and used for a standard GWAS for ectopic ureters in the EM breed (Gallana et al., 2018). The PLINK v1.9b4.6 software (Purcell et al., 2007; Chang et al., 2015) was used to filter markers exhibiting call rate greater than 95% and minor allele frequency of at least 2%.

2.3 Regression model

The blupf90 suite of programs (Misztal et al., 2002) was used to fit the following linear mixed model to the data:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where \mathbf{y} is the 892 x 1 vector of ALS, $\mathbf{1}$ is a 892 x 1 vector of ones, σ is a scalar representing the overall mean, $\mathbf{u} = [\mathbf{u}_n \ \mathbf{u}_g]^T$ is the 1632 x 1 vector of unobserved random effects for non-genotyped (\mathbf{u}_n) and genotyped (\mathbf{u}_g) animals, \mathbf{Z} is the 892 x 1632 incidence matrix relating animals to phenotypes, and \mathbf{e} is the 892 x 1 vector of random residual effects. Following the single-step GBLUP approach (Misztal et al., 2009; Aguilar et al., 2010), random effects were assumed independent and distributed as:

$$\mathbf{u} \sim \text{MVN}(0, \mathbf{H}\sigma_u^2)$$

$$\mathbf{e} \sim \text{MVN}(0, \mathbf{W}\sigma_e^2)$$

where \mathbf{W} is a diagonal matrix of number of recorded litters and \mathbf{H} is a composite matrix combining pedigree (\mathbf{A}) and genomic (\mathbf{G}) relationships among animals. All elements from \mathbf{H} were identical to those in \mathbf{A} , except that the subset of \mathbf{A} containing pedigree relationships among genotyped animals (\mathbf{A}_{gg}) was replaced by $\omega\mathbf{G} + (1 - \omega)\mathbf{A}_{gg}$, where $\omega = 0.95$. The genomic relationship matrix was defined as $\mathbf{G} = \mathbf{MDM}^Tq$, where \mathbf{M} is the centered genotype matrix with animals in rows and markers in columns, \mathbf{D} is a diagonal matrix of marker weights, and $q = 1/[2\sum p_i(1 - p_i)]$, where p_i is the reference allele frequency at marker i (VanRaden, 2008). Variance components σ_u^2 and σ_e^2 were estimated via REstricted Maximum Likelihood (REML) using 10 iterations of the Expectation-Maximization (EM-REML) algorithm followed by maximization until convergence by the Average-Information (AI-REML) algorithm (Patterson e Thompson, 1971; Gilmour et al., 1995; Yang et al., 2011). The heritability of the trait was obtained from REML estimates of variance components as $h^2 = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$. Solutions for μ and \mathbf{u} were obtained by solving the mixed model equations.

2.4 Genome-wide association analysis

The genome-wide scan for ALS associations was based on the single-step GWAS approach (Wang et al., 2012; 2014; Zhang et al., 2016) and comprised converting animal effects into marker effects by iterating the steps below:

1. Initialize \mathbf{D} as an identity matrix
2. Fit the mixed model
3. Update allele substitution effects: $\hat{\mathbf{a}} = \mathbf{qDM}'\mathbf{G}^{-1}\hat{\mathbf{u}}_g$
4. Update marker-specific variances: $\sigma_i^2 = 2\rho_i(1 - \rho_i)\hat{a}_i^2$
5. Update the genome-wide variance: $\sigma_a^2 = \sum\sigma_i$
6. Update the contribution of each marker to the genome-wide variance: $\sigma_i = \sigma_i^2/\sigma_a^2$
7. Update marker weights: $\mathbf{D}_{ii} = \pi_i N$
8. Exit if the maximum number of iterations has been reached, otherwise loop to 2

The analysis was carried out using two iterations in order to achieve a balanced relationship between increased power to detect large effects and shrinkage of small effects. Significance of SNV markers was assessed using a subsampling procedure, as suggested by Wang et al. (2014). Briefly, the entire single-step analysis was re-run on multiple subsamples of the original data, and marker effects obtained in these samples were used to derive standard errors as follows:

$$SE(\hat{a}_i) = \sqrt{\frac{m}{n(b-1)} \sum_{k=1}^b (\hat{a}_{ik} - \bar{a}_i)^2}$$

where, relative to marker i , b is the number of subsamples, m is the subsample size, n is the original sample size, \hat{a}_{ik} is the estimated effect at the k^{th} subsample and \bar{a}_i is the marker effect averaged across all subsamples. Here we assumed $b = 150$, $n = 896$ and $m = 625$ (i.e., ~70% of the original phenotypic data), and adequacy of the choices of b and m were verified by evaluating the convergence of standard error estimates (**Appendices-Figure 2A**). Standard errors were then used to compute the conventional chi-squared

statistic $t_i^2 = [\hat{\alpha}_i / SE(\hat{\alpha}_i)]^2$. The genomic control method (Devlin e Roeder, 1999) was used to correct the test statistics for systematic inflation, which was measured as the slope of a simple linear regression of observed onto theoretical chi-squared quantiles. A Bonferroni-corrected significance level of $0.05/N$ was adopted, where N is the number of markers. The R v3.4.4 software (R CORE TEAM, 2018) and the ggplot2 package (Wickham, 2016) were used for graphical visualization of the results.

2.5 Analysis of whole genome sequence data

A total of 8 dogs were whole genome sequenced at an average coverage of 13x. Paired-end reads of 2 x 101 bp were generated with a Illumina[®] HiSeq 2000 instrument, following the manufacturer's protocol. Read alignments to the CanFam3.1 assembly were performed with the Burrows-Wheeler Alignment (BWA) algorithm v0.7.10-r789 (Li e Durbin, 2009). Optical and PCR duplicates were marked with PicardTools v1.119 (available at: <http://broadinstitute.github.io/picard/>). Sequence variants were extracted from aligned reads using the mpileup algorithm from SAMtools v1.3.1 and BCFtools v1.3.1 (Li et al., 2009). Variant effects were predicted and annotated with Ensembl Variant Effect Predictor (Mclaren et al., 2016). The Integrative Genomics Viewer v2.3 software (Thorvaldsdóttir et al., 2013) was used to visually inspect sequence alignments to search for additional variants that were not detected by the mpileup algorithm, such as large structural variants.

2.6 Orthology analyses

The UCSC (Casper et al., 2018) and Ensembl (Zerbino et al., 2018) databases were used to retrieve inter-species alignments of protein and genomic sequences of candidate genes, as well as PhyloP scores of nucleotide conservation in 100 vertebrates.

3. Results

3.1 Single-step GWAS maps litter size associations to a chromosomal region sheltering *GDF9*

We combined pedigree data from 1632 dogs with 892 phenotypes and 372 genotypes using the single-step GWAS method to map genetic loci associated with ALS in the EM breed. A total of 87855 SNV markers were screened for association assuming a significance level of 5.69×10^{-7} , and three markers on canine autosome 11 (CFA11) were declared significant (**Figure 1 - A**). Relative to the CanFam3.1 reference genome assembly, the most significant marker was g.11:20806345C>T (rs22084797, $p = 4.21 \times 10^{-7}$), which presented a reference allele frequency of 33.3%. Associations on CFA11 mapped to a chromosomal region proximal to a ~570 kb gap in marker coverage (**Figure 1 - B**). Test statistics used to compute p -values were only slightly inflated and were therefore corrected using genomic control (**Figure 1 - C**). The sheep and goat sequences orthologous to the gap segment spanning CFA11:21076652-21646082 were previously shown to affect litter size, with the growth differentiation factor 9 gene (*GDF9*) being pointed as the causal gene (Silva et al., 2011; Feng et al., 2011; Våge et al., 2013; An et al., 2013). Notably, the topology of $-\log_{10}(p)$ values on CFA11 suggested peaking associations towards the gap region in EM dogs, where the canine *GDF9* is located (CFA11:21143982-21147237). Additive genetic and residual variances of ALS were estimated at 1.143 ± 0.252 and 1.506 ± 0.138 , respectively, resulting in a heritability of 43.1% in EM dogs. Associations on CFA11 accounted for ~15% of this heritability, translating into ~6.5% of the phenotypic variance.

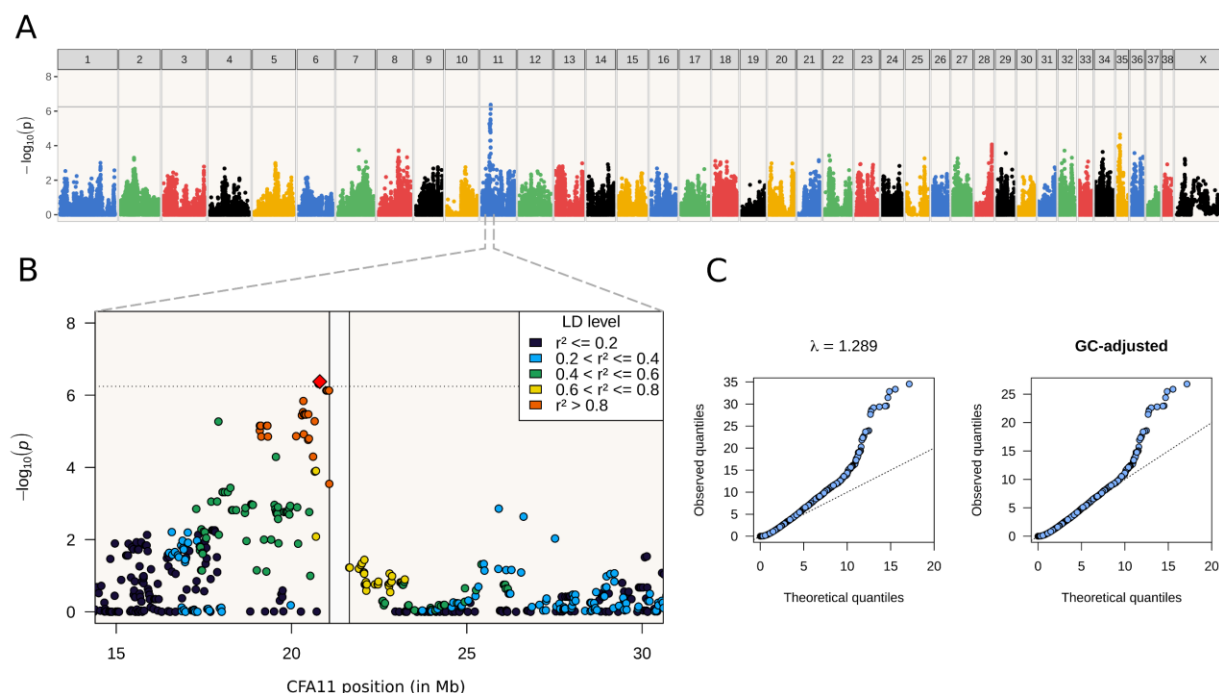


Figure 1. Genome-wide association scan for major loci affecting litter size in Entlebucher Mountain dogs. **(A)** Manhattan plot showing significant associations ($p < 5.69 \times 10^{-7}$) mapping to canine autosome 11 (CFA11). **(B)** Regional association plot of CFA11 demonstrating linkage disequilibrium to the top scoring marker g.11:20806345C>T (rs22084797, red diamond) and a gap in marker coverage between 21.08 and 21.65 Mb (white rectangle). **(C)** Quantile-quantile plots of association test statistics before (left panel) and after (right panel) genomic control (GC) adjustment for the inflation factor (λ).

3.2 Whole genome sequence data reveal a novel missense variant in *GDF9*

We analyzed WGS data of 8 EM dogs to prospect candidate causal variants underlying the CFA11 association signal. We detected two missense variants in *GDF9* that were in perfect linkage disequilibrium with the most significant marker g.11:20806345C>T, namely g.11:21146969C>T (rs850899575) and g.11:21147009G>A. The Ensembl track for variant g.11:21146969C>T indicated that the alternative allele T underlying the tolerated p.Cys90Tyr amino acid substitution was fairly common across dog breeds, as it had a frequency of 19% in 218 dog genomes. However, the g.11:21147009G>A variant associated with the amino acid change p.Pro77Ser was not cataloged and is therefore novel, implying it is not as commonly found across dog breeds

as g.11:21146969C>T. From the 8 sequenced EM dogs, 5 were homozygous AA and 3 were heterozygous GA at the novel variant, indicating that the alternative allele A is the major one in the breed. Based on the perfect linkage disequilibrium with rs22084797, we predicted that allele A at the novel variant was associated with decreased litter size. We further observed that the residue 77 was positioned immediately proximal to a 6-residues deletion in the canine *GDF9* protein sequence that is not found in non-canid species (**Figure 2 - A**). Additionally, residues 77 (Pro) and 78 (Ala) flanking the canine deletion were found to be conserved across mammals, indicating that further changes to this part of the protein sequence could alter its function and trigger natural negative selection. Examination of the codon sequences revealed that the first two bases of codon 77 were highly conserved (**Figure 2 - B**), whereas the third base was not, which was expected given that CCN encodes for the conserved proline residue. In particular, the g.11:21147009G>A variant in EM dogs alters the most conserved base of that codon (**Figure 2 - C**).

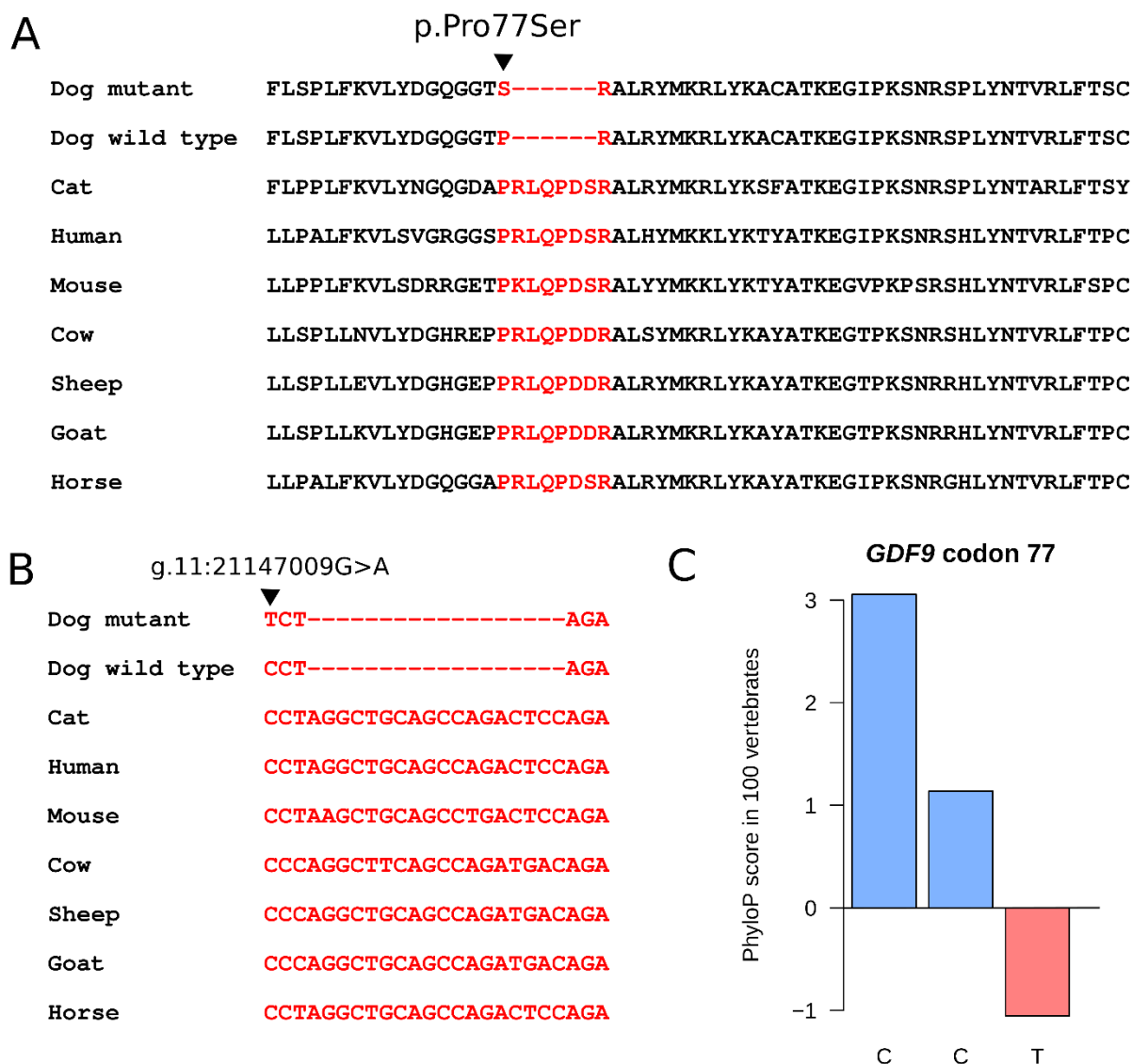


Figure 2. Effect and conservation of the detected missense variant (g.11:21147009G>A, p.Pro77Ser) in the growth differentiation factor 9 gene (*GDF9*). **(A)** The proline to serine substitution at residue 77 occurs immediately proximal to a 6-residues deletion that is not observed in non-canid species. **(B)** Orthologous genomic sequences (negative strand) of the *GDF9* segment containing the seven codons that were deleted in dogs. **(C)** Conservation of the triplet encoding residue 77. The low conservation of thymine (T) and high conservation of the two cytosines (C) are consistent with conservation of proline, since this amino acid is encoded by CCN.

4. Discussion

Average litter size (ALS) is an indicator of fertility and prenatal survival and is defined as the mean number of offspring per litter. Leroy and collaborators (2015) estimated the heritability (h^2) of ALS in seven dog breeds and found values ranging from 6% in Basset hounds to 11% in Bernese Mountain dogs, indicating that the trait is most likely complex (i.e., affected by a large number of genes) and lowly heritable within dog breeds. However, between-breeds differences in h^2 estimates suggest that some dog breeds might have a stronger genetic contribution to ALS variation than others, implying breed-specific genetic architectures for ALS. In the present study, we found that h^2 for ALS in the Entlebucher Mountain (EM) dog, an Alpine breed closely related to Bernese Mountain dogs, could be as high as 43%. Furthermore, in spite of the assumed trait complexity, markers flanking the growth differentiation factor 9 gene (*GDF9*) could explain as much as 6% of the variance in ALS in EM dogs.

The upper limit of offspring per litter is physiologically constrained by the ovulation rate of the dam, which is partially controlled by two genes of the transforming growth factor-beta (TGF- β) superfamily: *BMP15* (bone morphogenetic protein 15) and *GDF9* (Moore et al. 2003; Shimasaki et al. 2004; Reader et al., 2011; 2016). Animal models for genetic variants in *GDF9* and *BMP15* with large effects on ALS are well exemplified in sheep (Silva et al., 2011; Våge et al., 2013; Demars et al., 2013) and goats (Feng et al., 2011; An et al., 2013). The proteins encoded by *GDF9* and *BMP15* are mainly expressed by the oocyte, and act in synergism to regulate ovarian follicle growth and to fine tune steroidogenesis, differentiation, division and expansion of cumulus cells (Reader et al., 2016). In particular, expansion of cumulus cells is a morphological marker of oocyte quality and viability, and is crucial for oocyte maturation (Lee et al., 2017). Both *GDF9* and *BMP15* signal the *Sma* mothers against decapentaplegic (SMAD) pathway, but while *GDF9* activates SMAD2/3, *BMP15* activates SMAD1/5/8 (Reader et al., 2016). The canine *BMP15* is located on chromosome X positions 43764777-43771303, and although we can not rule out the contribution of variants in *BMP15* to ALS variation in EM dogs, our study did not detect significant associations nearby that gene. Likewise, two deletions downstream of the SMAD2 gene (*SMAD2*) were previously shown to affect body size in

several dog breeds (Rimbault et al., 2013), but our analysis did not identify associations for ALS towards *SMAD2* (CFA7:43720229-43769889) in the EM breed. The absence of associations for *BMP15* and *SMAD2* could be related to a true lack of effects of variants of these genes on ALS variation in EM dogs, or to a limitation of statistical power in our study. In the latter case, either the effect size or frequency of the variants could be too low to be detected at our current sample size. In fact, the deletions were observed only in one out of the eight whole genome sequenced EM dogs (**Apêndices- Figure 3A**). Therefore, the results of the present study were restricted to significant genetic effects mapping to *GDF9*.

The most plausible candidate causal variant pinpointed by our whole genome sequence analysis was the novel g.11:21147009G>A missense substitution in *GDF9*, predicted to lead to a proline to serine change at codon 77 of the *GDF9* protein. Allele G at the variant site was found to be highly conserved in 100 species of vertebrates, implying negative selection towards substitutions of that nucleotide across species. Moreover, the derived allele A was associated with negative effects on ALS, in line with the observed decline in the mean offspring per litter in EM dogs in the last decade (Schrack et al., 2017). Therefore, the candidate variant is consistent with our initial hypothesis that inbreeding in EM dogs might have facilitated drift of variants with large negative effects on ALS. Additionally, the p.Pro77Ser residue substitution was immediately followed by a 6-residues deletion that is fixed in the domestic dog but absent in non-canid species. Canids uniquely ovulate immature oocytes at the Metaphase II stage of the first meiotic division, and oocyte maturation is only completed in the oviduct (Chastant-Maillard et al., 2011; Reynaud et al., 2012). Of note, Lee et al. (2017) demonstrated that the canine oviduct highly expresses oocyte maturation-related genes, including *BMP15* and *SMAD2*. Therefore, the investigation of differences in protein sequences and patterns of gene expression between canids and non-canids within the *GDF9/BMP15/SMAD2* pathway seems a promising source of insights into the particularities of reproductive biology in the Canidae family.

5. Conclusions

The canine *GDF9* chromosomal region shelters genetic variants explaining ~6.5% of the variation in average litter size in Entlebucher Mountain dogs. Among these variants, g.11:21147009G>A is the most plausible functional candidate since it leads to an amino acid substitution preceding a canid-specific deletion of 6 residues in the protein encoded by *GDF9*. Although a more detailed characterization of the variants in *GDF9* is necessary before marker-assisted selection for litter size is conducted in the Entlebucher Mountain dogs, our results open a new window for improvement of fertility traits in the breed. Finally, given the importance of *GDF9* to oocyte maturation, future research prioritizing differences in protein sequence and expression patterns of this gene between canids and non-canids may contribute to improve our knowledge about this yet cryptic process in canids.

6. Funding

Author R.B.P.T. was supported by scholarships from the Coordination for the Improvement of Higher Education Personnel (CAPES) and the Sao Paulo Research Foundation (FAPESP, process 2017/08373-1). This work was also funded by the Albert Heim Stiftung, the Stiftung für das Wohl des Hundes, as well as the Swiss Federal Food Safety and Veterinary Office.

7. Acknowledgments

We thank the breeders association and especially Margret Epple, Herma Cornelese and Max and Gertrud Heller for their kind support.

8. Author contributions

YTU, RBPT and CS conceived the study. JFG, JS and TL contributed to the study design. MG, SS, AB, IR and VJ sampled DNA for genotyping and sequencing. TL coordinated genotyping and sequencing. CS and GD curated the phenotypic and pedigree data. RBPT, MM and YTU performed data analyses. RBPT and YTU wrote the manuscript. All authors revised and agreed with the contents of the manuscript.

9. References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science** 93:743-752.

An XP, Hou, JX, Zhao HB, Li G, Bai L, Peng JYM, Yan Q, Song YX, Wang JG, Cao BY (2013) Polymorphism identification in goat GNRH1 and GDF9 genes and their association analysis with litter size. **Animal Genetics** 44:234-238.

Casper J, Zweig AS, et al (2018) The UCSC Genome Browser database: 2018 update. **Nucleic Acids Res** 46:D762-D769.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. **Gigascience** 4:7.

Chastant-maillard S, Viaris de Lesegno C, Chebrou M, Thoumire S, Meylheuc T, Fontbonne A, Chodkiewicz M, Saint-dizier M, Reynaud K (2011) The canine oocyte: uncommon features of in vivo and in vitro maturation. **Reproduction, Fertility and Development** 23:391-402.

Demars J, Fabre S, et al (2013) Genome-wide association studies identify two novel BMP15 mutations responsible for an atypical hyperprolificacy phenotype in sheep. **PLoS Genetics** 9:e1003482.

Devlin B, Roeder K (1999) Genomic control for association studies. **Biometrics** 55:997-1004.

Feng T, Geng CX, et al (2011) Polymorphisms of caprine GDF9 gene and their association with litter size in Jining Grey goats. **Molecular Biology Reports** 38:5189-5197.

Gallana M, Utsunomiya YT, Dolf G, Torrecilha RBP, Falbo AK, Jagannathan V, Leeb T, Reichler I, Sölkner J, Schelling C (2018) Genome-wide association study and heritability estimate for ectopic ureters in Entlebucher mountain dogs. **Animal Genetics** 49:645-650.

Lee SH, Oh HJ, Kim MJ, Kim GA, Choi YB, Jo YK, Setyawan EMN, Lee BC (2017) Oocyte maturation-related gene expression in the canine oviduct, cumulus cells, and oocytes and effect of co-culture with oviduct cells on in vitro maturation of oocytes. **Journal of Assisted Reproduction and Genetics** 34:929-938.

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: An efficient algorithm for variance parameters estimation in linear mixed models. **Biometrics** 51:1440-1450.

Leroy G, Phocas F, Hedan B, Verrier E, Rognon X (2015) Inbreeding impact on litter size and survival in selected canine breeds. **Vet J** 203:74-78.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics** 25:1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) 1000 GENOME PROJECT DATA PROCESSING SUBGROUP. The Sequence Alignment/Map format and SAMtools. **Bioinformatics** 25:2078-2079.

Mclaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F (2016) The Ensembl Variant Effect Predictor. **Genome Biol** 1:122.

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science** 92:4648-4655.

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH (2002) BLUPF90 and related programs (BGF90). Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. **Anais** 28:21-22.

Moore RK, Otsuka F, Shimasaki S (2003) Molecular basis of bone morphogenetic protein-15 signalling in granulosa cells. **The Journal of Biological Chemistry** 278:304-310.

NEMDA (National Entlebucher Mountain Dog Association) (2018) **Breed history**. Disponível em: <<https://nemda.org/breed-history/>> Acesso em: 30 ago. 2018.

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. **Biometrika** 58:545-554.

Purcell S, Neale B, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am J Hum Genet** 81:559-575.

R Core Team (2018) **R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing**, Vienna, Austria. Disponível em: <<https://www.r-project.org>>. Acesso em: 30 ago 2018.

Reader KL, Heath DA, Lun S, Mcintosh CJ, Western AH, Littlejohn RP, McNatty KP, Juengel JL (2011) Signalling pathways involved in the cooperative effects of ovine and murine GDF9+BMP15-stimulated thymidine uptake by rat granulosa cells. **Reproduction** 142:123-131.

Reader KL, Mottershead DG, Martin GA, Gilchrist RB, Heath DA, McNatty KP, Juengel J, L (2016) Signalling pathways involved in the synergistic effects of human growth differentiation factor 9 and bone morphogenetic protein 15. **Reproduction, Fertility and Development** 28:491-498.

Reynaud K, Fontbonne A, Saint-Dizier M, Thoumire S, Marnier C, Tahir MZ, Meylheuc T, Chastant-Maillard S (2012) Folliculogenesis, ovulation and endocrine control of oocytes and embryos in the dog. **Reproduction in Domestic Animals** 47:66-69.

Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, Kilroy-Glynn P, Wayne RK, Sutter NB, Ostrander EA (2013) Derived variants at six genes explain nearly half of size reduction in dog breeds. **Genome Res** 23:1985-1995.

Schrack J, Dolf G, Reichler IM, Schelling C (2017) Factors influencing litter size and puppy losses in the Entlebucher Mountain dog. **Theriogenology** 95:163-170.

Shimasaki S, Moore RK, Otsuka F, Erickson GF (2004) The bone morphogenetic protein system in mammalian reproduction. **Endocr. Rev** 25:72-101.

Våge DI, Husdal M, Kent MP, Klemetsdal G, Boman IA (2013) A missense mutation in growth differentiation factor 9 (*GDF9*) is strongly associated with litter size in sheep. **BMC Genetics** 14:1.

VanRaden PM (2008) Efficient methods to compute genomic predictions. **Journal of Dairy Science** 91:4414-4423.

Silva BD, Castr EA, et al (2011) A new polymorphism in the Growth and Differentiation Factor 9 (*GDF9*) gene is associated with increased ovulation rate and prolificacy in homozygous sheep. **Animal Genetics** 42:89-92.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. **Brief Bioinform** 14:178-192.

Wang H, Misztal I, Aguilar I, Legarra A, Fernando RL, Vitezica Z, Okimoto R, Wing T, Hawken R, Muir WM (2014) Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics** 5:134.

Wang H, Misztal I, Aguilar I, Legarra A, Muir WM (2012) Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research** 94:73-83.

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. **Springer-Verlag**: New York, 260.

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. **American Journal of Human Genetics** 88:76-82.

Zerbino DR, Achuthan P, et al (2018) Ensembl. **Nucleic Acids Res** 46:D754-D761.

Zhang X, Lourenco D, Aguilar I, Legarra A, Misztal I (2016) Weighting Strategies for Single-Step Genomic BLUP: An Iterative Approach for Accurate Calculation of GEBV and GWAS. **Frontiers in Genetics** 7:151.

CAPÍTULO 3- High frequency canine haplotypes with exceptional deficiency of homozygosity

Rafaela Beatriz Pintor Torrecilha; Marco Milanesi; Milena Gallana; Sarah Sugiarto; Ann-Kristin Besold; Iris Reichler; Vidhya Jagannathan; Claude Schelling; Gaudenz Dolf; Cord Drögemüller; Márcia Dalastra Laurenti; Luís Fábio da Silva Batista; Gábor Mészáros; Johann Sölkner; Tosso Leeb; José Fernando Garcia; Yuri Tani Utsunomiya

ABSTRACT- A recessive phenotype is produced when its causal genetic variant is at the homozygous state. A highly unfavorable recessive phenotype (HURP) is defined as a recessive phenotype that is under strong negative (purifying) selection. While most HURP-causing genetic variants are typically purged or kept at very low frequency in a population, some of these may become highly prevalent as a consequence of genetic drift, genetic draft or pleiotropy (i.e., the HURP-causing variant may concurrently determine an advantageous dominant phenotype that is positively selected). In particular, genetic drift can be potentiated by population bottlenecks and inbreeding, phenomena that occurred during the formation of modern breeds of domestic dogs. Because it is negatively selected, a HURP is usually rare in a population and thus difficult to observe or measure. However, if the HURP-causing variant is prevalent, it leaves a distinctive genomic signature characterized by chromosomal regions sheltering high frequency haplotypes with exceptional deficiency of homozygosity (HDHR). The purpose of this study was to evaluate the use of genome-wide screenings for HDHR as a mapping strategy for HURP-causing variants in dogs. In order to validate the method, we first applied the screening technique to Chinese Crested (CC, n = 137) dogs, a breed that segregates a well known pleiotropic homozygous-lethal variant. We then applied the genome-wide analysis in Golden Retriever (GR, n = 871), Labrador Retriever (LR, n = 561) and Entlebucher Mountain (EM, n = 373) dogs. By using a combination of high density single nucleotide variant (SNV) microarray and whole genome sequence (WGS) data, we were able to detect sixteen HDHR across all four canine breeds. Five HDHRs contained genes linked to either hair coat phenotypes (*FOXI3*, *CBD103* and *TBX15*) or retinal disease (*PRPF31*, *CNOT3*, *SAG*, *SMC3* and *PDE6D*). Two HDHRs overlapped with chromosomal regions that are highly differentiated between domestic and wild canids (*AMY2B* and *CRTC3* chromosomal domains). Six HDHRs had near zero homozygotes and therefore were likely tagging lethal alleles. One of these was further verified to be significantly associated with average litter size in EM dogs. This study provides a proof of concept that genetic variants underlying HURPs can be indirectly mapped using genome-wide scans for HDHRs in dog populations.

Keywords: positive selection; purifying selection; deleterious mutations; *Canis lupus familiaris*; autosomal recessive inheritance; population genetics

1. Introduction

Domestication of dogs (*Canis lupus familiaris*) may have occurred over 15 thousand years ago, but the selective breeding that gave rise to the nearly 400 dog breeds cataloged to date greatly intensified in the past few centuries (Ostrander et al., 2017). The process of formation of dog breeds involved strong population bottlenecks and intense artificial selection for behavioral and morphological traits, resulting in rapid reductions of within-breed genetic diversity and accumulation of deleterious mutations (Marsden et al., 2016). Consequently, the domestic dog became an invaluable model for studying patterns of genetic variation produced by inbreeding and drift in finite populations. Moreover, the characterization of loci harboring deleterious mutations in dogs has the potential to enlighten the function of orthologous genes in humans and uncover their relevance in disease and biomedical applications.

The indirect detection of genetic variants underlying highly unfavorable recessive phenotypes (HURPs) has successfully been demonstrated in other animal species through the search for high frequency haplotypes with exceptional deficiency of homozygosity (VanRaden et al. 2011; Sahana et al., 2013; Hoff et al., 2017; Häggman e Uimari, 2016). The method begins by determining haplotypes in chromosomal windows of a predefined length. Then, a hypothesis test (typically based on a null binomial distribution) is applied to each haplotype to compare the observed number of homozygotes with what would be expected under Hardy-Weinberg Equilibrium (HWE). Haplotypes with much lower than expected homozygosity are considered tags for putative variants causing HURPs. The technique has been found particularly useful in the prediction of early embryonic, recessive lethal phenotypes, since these are difficult to measure and barely observable.

Apart from deleterious alleles that could be under the influence of natural negative (purifying) selection, the technique may also be useful to reveal mutations that are artificially penalized at the homozygous state by human selection. In this context, “unfavorable” may be interpreted beyond the scope of the classical definition of “fitness” and embrace both natural and artificial selection. Therefore, exceptional deficiency of haplotype homozygosity may be explained by genetic variants implicated in selectively disadvantageous recessive phenotypes, which comprises a broader class of variants than simply deleterious or homozygous-lethal mutations. Considering the current availability of high density single nucleotide variant (SNV) genotype data for a number of dog breeds, this strategy could be applied to the canine genome

in order to uncover high frequency haplotypes that segregate with HURP-causing genetic variants in the species. Here, we report the first study of this kind in dogs, which relied on data from 373 Entlebucher Mountain (EM), 137 Chinese Crested (CC), 871 Golden Retriever (GR) and 561 Labrador Retriever (LR) dogs genotyped for over 148,000 SNVs. In order to propose candidate mutations, we also investigated whole genome sequence data from eight EM, four GR and three LR dogs, as well as sequence variants previously reported by other re-sequencing projects.

2. Materials and Methods

2.1 Genotypes

Illumina® CanineHD BeadChip assay genotypes from the EM, CC, GR and LR breeds were gathered from six different sources, which are listed in **Table 1**. General genotype data manipulation was performed with PLINK v1.9b4.6 (Purcell et al., 2007; Chang et al., 2015). Replicates between genotyping initiatives were avoided by excluding samples with 90% or more of their genotypes identical to another sample. Prior to analysis, all 173,662 SNV probe sequences were realigned to the CanFam3.1 reference assembly using custom scripts to retain only autosomal markers with unambiguous genomic coordinates. Next, SNVs failing a Fisher's exact test for HWE ($p < 10^{-5}$) in all four breeds were excluded. Finally, only markers exhibiting call rate greater than 95% and GenTrain Score greater than 70% were kept for downstream analyses. After filtering, genotypes were subjected to phasing with the Segmented HAPlotype Estimation & Imputation Tool v2.r837 (SHAPEIT2) (O'Connell et al., 2014) using windows of 500 kbp, 200 states and 10 burn in, 10 prune and 50 main iterations. The effective population size parameter required by SHAPEIT2 was estimated for each breed separately from linkage disequilibrium data (N_{eLD}) with SNeP v1.1 (Barbato et al., 2015).

Table 1. Sources of Illumina® CanineHD genotypes used in this study

| Breed | Abbreviation | Sample size | Samples from this study | Samples from other studies | References |
|----------------------|--------------|-------------|-------------------------|----------------------------|---|
| Chinese Crested | CC | 137 | 137 | 0 | - |
| Entlebucher Mountain | EM | 373 | 373 | 0 | - |
| Golden Retriever | GR | 871 | 72 | 799 | Arendt et al. (2015); Olsson et al. (2015); Tonomura et al (2015); Molin et al. (2015) |
| Labrador Retriever | LR | 561 | 132 | 429 | Forsberg et al. (2015); Olsson et al. (2015); Molin et al. (2015) |
| Total | - | 1948 | 720 | 1228 | - |

2.2 Inbreeding analysis

Since genetic drift is potentiated by population bottlenecks and inbreeding, we estimated genomic autozygosity at the individual level based on the percentage of the genome that was covered by Runs of Homozygosity (ROH), as proposed by McQuillan et al. (2008). Stretches of consecutive homozygous genotypes were identified for each dog using the Hidden Markov Model (HMM) implemented in BCFTools v1.3.1 (Narasimhan et al., 2016). Allele frequencies required by the model were estimated from breed-level genotype data. Uncertainty about genotype emissions was taken into account by setting the phred-scaled genotype likelihood to the default value of 30. Then, genomic autozygosity (F_{ROH}) was computed as:

$$F_{ROH} = 100 \frac{\sum_{i=1}^n R_i}{G}$$

where R_i is the length of ROH i , and G is the total size of the genome covered by SNV markers. The value of G was obtained by summing the distances between consecutive markers according to their positions in the CanFam3.1 assembly. Genomic autozygosity was calculated

based on ROH of different minimum lengths: 1, 2, 4, 8 or 16 Mbp, representing inbreeding that occurred approximately 50, 25, 13, 6 and 3 generations in the past, respectively (Howrigan et al., 2011; Ferenčaković et al., 2013). Considering a generation interval of up to 4 years (Lewis et al., 2015), segments over 1Mbp in length were considered proxies for inbreeding that occurred up to 200 years ago, which covers most breed history. Locus autozygosity (F_L) was also calculated for each SNV as the percentage of dogs presenting a ROH encompassing the SNV (Kim et al., 2013).

2.3 Choice of haplotype size

Let k be either the size of the sliding window in base pairs (bp) or the number of markers included in the sliding window. As argued by Hoff et al. (2017), the larger is the value of k , the higher is the likelihood that identical-by-state (IBS) haplotypes are also identical-by-descent (IBD). Consequently, IBS haplotypes at large k will be more likely to be identical at the level of sequence, except for a few young or *de novo* mutations. However, the likelihood of IBD is expected to decline as a function of number of generations since the common ancestor, such that choosing an excessively large value of k may give rise to many private haplotypes. At the opposite direction, a too small value of k may yield IBS haplotypes that are not IBD. The value of k may be further affected by the demographic history of the population under investigation, being specially sensitive to the extent of linkage disequilibrium (LD) in the genome. Therefore, the choice of k is not a trivial problem and thus lacks a straight-forward solution. In order to retain as much power as possible in the present study, we interrogated haplotypes of lengths 100 kbp, 500 kbp, 1 Mbp, 2 Mbp and 5 Mbp. At each step, windows moved forward one quarter of their lengths. In order to avoid large gaps in marker coverage, we discarded haplotypes with densities below 50 kbp/SNV. All haplotypes were determined from phased data using the GHap R package v1.22 (Utsunomiya et al., 2016; R CORE TEAM, 2018).

2.4 Detection of haplotypes with deficit of homozygotes

Relative to haplotype i , let p_i and h_i represent the haplotype frequency and the observed number of homozygotes, respectively. The probability of observing h_i or less homozygotes under Hardy-Weinberg Equilibrium was obtained as:

$$Pr(X \leq h_i) = \sum_{j \leq h_i} \binom{n}{j} p_i^{2j} (1 - p_i^2)^{n-j}$$

where n is the total number of dogs and X is a random draw from the binomial distribution. Haplotypes presenting p smaller than a significance level α were considered homozygous-deficient (HDH) and prioritized for further investigation. The value of α was selected based on a permutation test, as specified in the next section. Since multiple HDH may overlap as consequence of our sliding window approach, consensus segments of overlapping HDH were built, which will be hereafter referred as HDH regions (HDHR).

2.5 Power analysis

As a consequence from the assumption of a binomial distribution under the null hypothesis, the probability formulated above is essentially dependent on sample size and haplotype frequency. Therefore, one can easily compute, for a given haplotype frequency, what would be the minimum sample size required to detect deficiency of homozygosity at the chosen significance level. As our primary goal was to identify high frequency HDH, we limited our search space to haplotype frequencies between 5% and 30%. We built a curve of sample size as a function of haplotype frequency and then projected the sample sizes of CC, EM, GR and LR against that curve in order to determine the lower boundary of haplotype frequency that was detectable in each breed in the present study.

2.6 Analysis of carrier sequence data

Whole genome sequences of 8 EM, 4 GR and 3 LR were available for analysis. Paired-end libraries of 2 x 100 bp were generated with a Illumina® HiSeq 2000 instrument, following the manufacturer's protocol. Reads were aligned against the CanFam3.1 assembly using the Burrows-Wheeler Alignment (BWA) algorithm v0.7.10-r789 (Li e Durbin, 2009). After alignment, optical and PCR duplicates were marked with PicardTools v1.119 (available at: <http://broadinstitute.github.io/picard/>). Sequenced animals were classified as carriers or non-carriers for each one of the candidate HDHR, and had their variants extracted from aligned

reads using the mpileup algorithm from SAMtools v1.3.1 and BCFtools v1.3.1 (Li et al., 2009). Variant effects were predicted and annotated with Ensembl Variant Effect Predictor (McLaren et al., 2016). Variants with genotypes matching haplotype predictions were considered neighboring tags for causal variants underlying HDHR. The Integrative Genomics Viewer v2.3 software (Thorvaldsdóttir et al., 2013) was used to visually inspect sequence alignments near tag variants to confirm variant positions and search for additional variants that were not detected by the mpileup algorithm (e.g. large structural variants). Sequence variants previously reported by other re-sequencing projects that overlapped with HDHR were also considered.

2.7 Gene annotation

We used the *intersect* utility from BedTools v2.25.0 (Quinlan e Hall, 2010) to map Ensembl gene annotations (Kinsella et al., 2011) to all HDHR. We also used the UCSC genome browser for complementary annotation (Casper et al., 2018). Orthologous gene entries in the Online Mendelian Inheritance in Men (OMIM, 2018), Online Mendelian Inheritance in Animals (OMIA, 2018) and the Mouse Genome Informatics (MGI, 2018) databases were also inspected in order to identify genes for which disruptive mutations have been previously found to cause any measurable phenotype, particularly embryonic lethality, early death or anatomophysiological impairment. Finally, we mined the literature in the search for other traits that could be linked to the candidate regions, such as recessive hair coat and disease phenotypes.

2.8 Association between haplotypes and average litter size

Data on average litter size estimated over three or more litters were available for 64 EM females. Phenotypes were regressed on candidate haplotypes using a multiple linear regression analysis. Haplotype effects were evaluated using the t-test for significance of regression coefficients and a significance level of 5%.

3. Results

3.1 Detection power at sample sizes below 1,000 dogs is constrained to HDH with frequency >10%

As shown in **Figure 1**, the sample sizes available for CC, EM, LR and GR limited our search to candidate haplotypes with frequencies equal to or greater than 28%, 17%, 14% and 11% in these breeds, respectively. Therefore, we hypothesize that genetic variants underlying candidate HDH in this study were most likely: (i) contributed by highly successful sires or dams via inbreeding and genetic drift; (ii) pleiotropic and simultaneously responsible for an unfavorable recessive phenotype and a selected dominant phenotype; or (iii) in LD with a selected variant and thus under genetic draft. In cases (ii) and (iii), positive selection on the favored phenotype must have been strong enough to counteract the action of purifying selection on the unfavorable phenotype. Case (i) suggests that inbreeding to influential families in the formation of the breed may override the selective force against the unfavorable variant and keep a large number of heterozygous carriers in the population.

3.2 Low effective population size and high inbreeding supports high frequency HDH in dogs

After filtering, a total of 149383, 148965, 149357 and 149670 SNVs were analyzed in the CC, EM, GR and LR data sets, respectively. Estimates of N_{eLD} were LR = 172, CC = 154, GR = 122 and EM = 20, indicating that all four breeds are partially inbred and that EM is far more inbred than the other three breeds. The N_{eLD} estimates agreed with the ROH analysis, where EM presented consistently larger average values of F_{ROH} (**Apêndices- Figures 1B-5B**). For instance, considering only recent inbreeding (ROH > 4 Mbp), average autozygosity was estimated at $17.22 \pm 6.45\%$, $16.60 \pm 6.35\%$, $20.91 \pm 7.62\%$ and $35.07 \pm 3.84\%$ for LR, CC, GR and EM, respectively. Therefore, the existence of high frequency HDH in these breeds is plausible given their current levels of inbreeding.

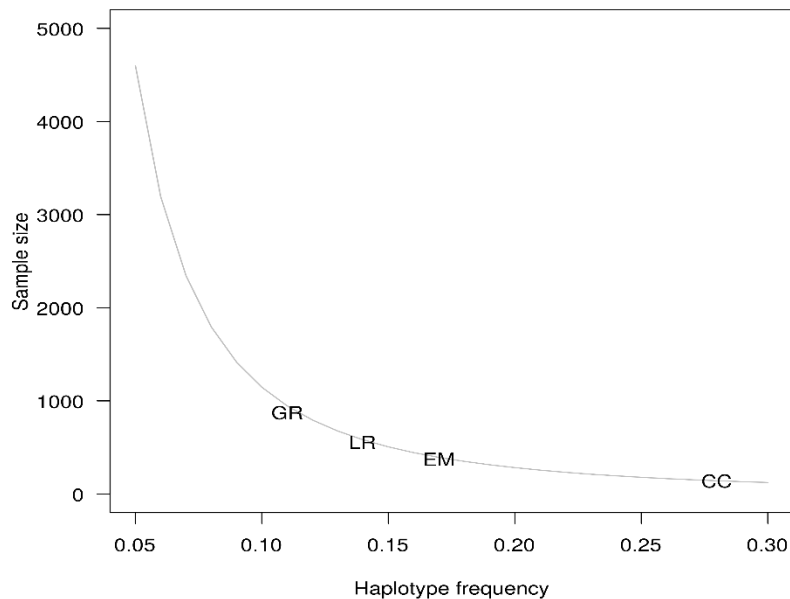


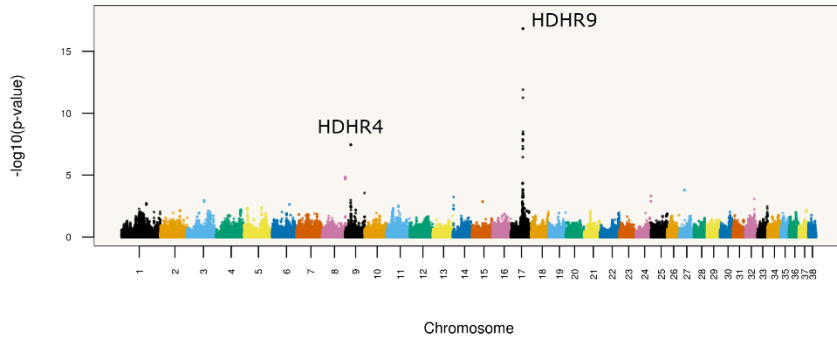
Figure 1. Sample size required to detect homozygous-deficient haplotypes (HDH) as a function of their frequency. The curve was built considering a significance level of $\alpha = 10^{-5}$. Abbreviations on the curve mark the sample sizes available for analyses in the four investigated breeds, revealing their corresponding lower bound of detectable frequency. GR = Golden Retriever, LR = Labrador Retriever, EM = Entlebucher Mountain, CC = Chinese Crested.

3.3 Genome-wide screening identifies sixteen HDHR across four dog breeds

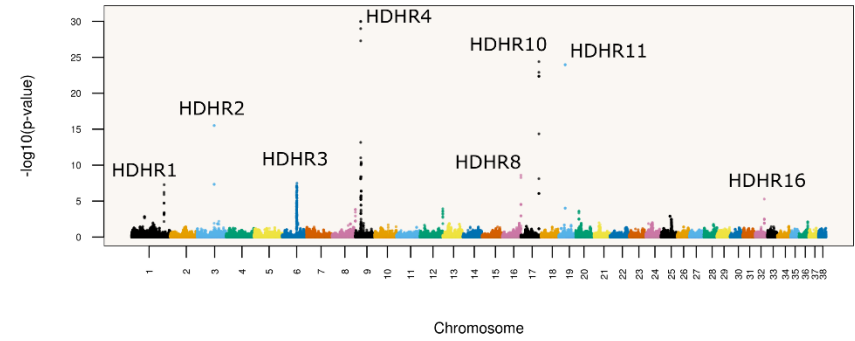
We performed genome-wide screenings for HDHRs using sliding windows of 100 kbp, 500 kbp, 1 Mbp, 2 Mbp and 5 Mbp. All analyses produced similar genome-wide significance topology, but the run using windows of 100 kbp retained more statistical power and thus presented better signal resolution (**Appendices- Figures 6B-9B**). Therefore, we focused on the results obtained with 100 kbp-long haplotypes. We found a total of sixteen HDHR across all four breeds studied here (**Figure 2**), which were numbered from HDHR1 to HDHR16 (**Table 2**). The numbers of HDHR that were specific to CC, GR, LR and EM were 1, 3, 2 and 4, respectively. We detected 4 HDHR that were shared between GR and LR, and a single HDHR that was shared between LR and EM. Remarkably, HDHR4 was shared among all four breeds.

1

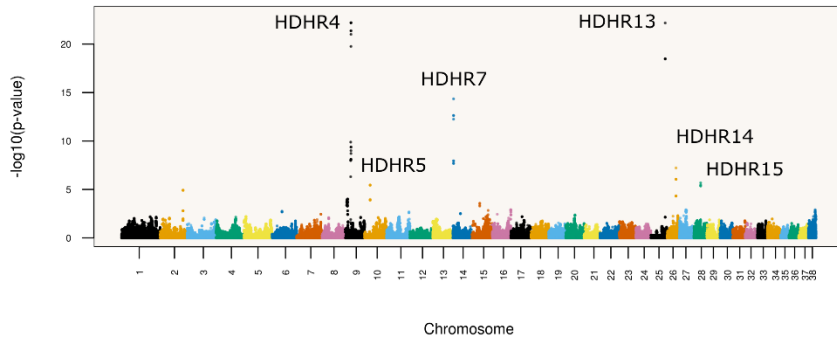
A Chinese Crested Dogs



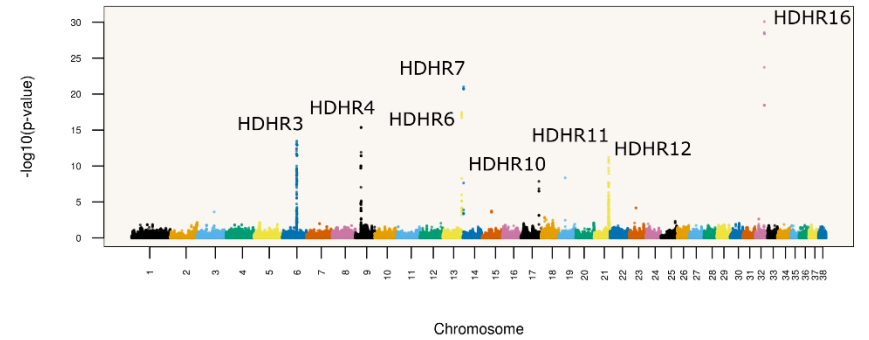
C Golden Retriever



B Entlebucher Mountain Dogs



D Labrador Retriever



2

3 **Figure 2.** Manhattan plots of genome-wide screenings for homozygous-deficient haplotypes of length 100 kbp.

4

5

6 **Table 2.** Observed-to-expected ratios and *p*-values (inside brackets) for homozygous-deficient haplotype regions (HDHR)

| Region | Chr | Start | End | Chinese Crested | Golden Retriever | Labrador Retriever | Entlebucher Mountain | Number of genes in the region | Candidate genes |
|------------|-----|---------------|---------------|-------------------------------------|---------------------------------------|---------------------------------------|--------------------------------------|-------------------------------|--|
| HDHR1 | 1 | 10262500 0 | 10322500 0 | - | 44:88 (5.30 x 10 ⁻⁸) | - | - | 29 | <i>CNOT3</i> <i>PRPF31</i> |
| HDHR2 | 3 | 52750000 | 54000000 | - | 12:64 (3.09 x 10 ⁻¹⁶) | - | - | 37 | <i>CRTC3</i> |
| HDHR3 | 6 | 44250000 | 49000000 | - | 7:33 (3.35 x 10 ⁻⁸) | 21:73 (3.37 x 10 ⁻¹⁴) | - | 22 | <i>AMY2B</i> <i>RNPC3</i> <i>COL11A1</i> |
| HDHR4 | 9 | 15000000 | 18500000 | 0:16 (3.55 x 10 ⁻⁸) | 20:116 (1.03 x 10 ⁻³⁰) | 4:46 (4.38 x 10 ⁻¹⁶) | 4:58 (2.83 x 10 ⁻²²) | 47 | <i>ARHGAP27</i> |
| HDHR5 | 10 | 15500000 | 17500000 | - | - | - | 67:108 (5.67 x 10 ⁻⁷) | 59 | <i>SYCE3</i> <i>NCAPH2</i> |
| HDHR6 | 13 | 58000000 | 60500000 | - | - | 0:39 (3.52 x 10 ⁻¹⁸) | - | 54 | - |
| HDHR7 | 14 | 1 | 6250000 | - | - | 1:50 (9.96 x 10 ⁻²²) | 0:31 (1.23 x 10 ⁻¹⁴) | 80 | - |
| HDHR8 | 16 | 58875000 | 59000000 | - | 27:69 (2.53 x 10 ⁻⁹) | - | - | 4 | <i>CBD103</i> |
| HDHR9 | 17 | 36750000 | 39000000 | 0:34 (1.49 x 10 ⁻¹⁷) | - | - | - | 61 | <i>FOXI3</i> |
| HDHR10 | 17 | 55750000 | 58500000 | - | 54:158 (4.06 x 10 ⁻²⁵) | 0:21 (4.21 x 10 ⁻¹⁰) | - | 42 | <i>TBX15</i> |
| HDHR11 | 19 | 20000000 | 20500000 | - | 1:57 (1.07 x 10 ⁻²⁴) | 0:19 (4.42 x 10 ⁻⁹) | - | 9 | - |
| HDHR12 | 21 | 45825000 | 46525000 | - | - | 0:25 (6.07 x 10 ⁻¹²) | - | 9 | - |
| HDHR13 | 25 | 43500000 | 46000000 | - | - | - | 1:51 (8.59 x 10 ⁻²³) | 55 | <i>PDE6D</i> <i>SAG</i> |
| HDHR14 | 26 | 26000000 | 28250000 | - | - | - | 0:16 (7.10 x 10 ⁻⁸) | 51 | - |
| HDHR15 | 28 | 20875000 | 21375000 | - | - | - | 49:85 (1.35 x 10 ⁻⁶) | 4 | <i>SMC3</i> |
| HDHR1 6 | 32 | 2650000 0 | 3000000 0 | - | 98:145 (5.29 x 10 ⁻⁶) | 26:123 (8.32 x 10 ⁻³¹) | - | 37 | <i>COL25A1</i> <i>SEC24B</i> |

3.4 HDHR9 validates the *FOXI3* homozygous-lethal frameshift variant in Chinese Crested dogs

A single CC-specific region was identified, namely HDHR9 (**Figure 3 - a**). This region spanned positions 36.75-39.00 Mbp on canine autosome (CFA) 17, with the peaking haplotype mapping to 37.93-38.08 Mbp. This haplotype had a sample frequency of 49.64%, and although 34 homozygotes were expected, none was observed ($p = 1.49 \times 10^{-17}$). The HDHR contained the forkhead box I3 gene (*FOXI3*, CFA17:38,032,447-38,038,363), known to harbor a 7 bp duplication in exon 1 that explains the hairless ectodermal dysplasia phenotype in CC (Drögemüller et al., 2008). This finding validates our approach of detecting high frequency HDH linked to HURPs, since the *FOXI3* mutation was originally reported by Drögemüller and collaborators (2008) using a different method that required phenotypes (i.e., genome-wide association analysis). We also found that the *FOXI3* region contained a ROH desert, consistent with strong positive selection for a homozygous-lethal mutation (**Figure 3 - b**). This *FOXI3* mutation is therefore an example of a variant causing both a recessive disadvantageous phenotype (death) and a selected dominant phenotype (hairless ectodermal dysplasia). Subsequent to this successful proof of principle, we proceeded with the exploration of the candidate regions in the remaining breeds.

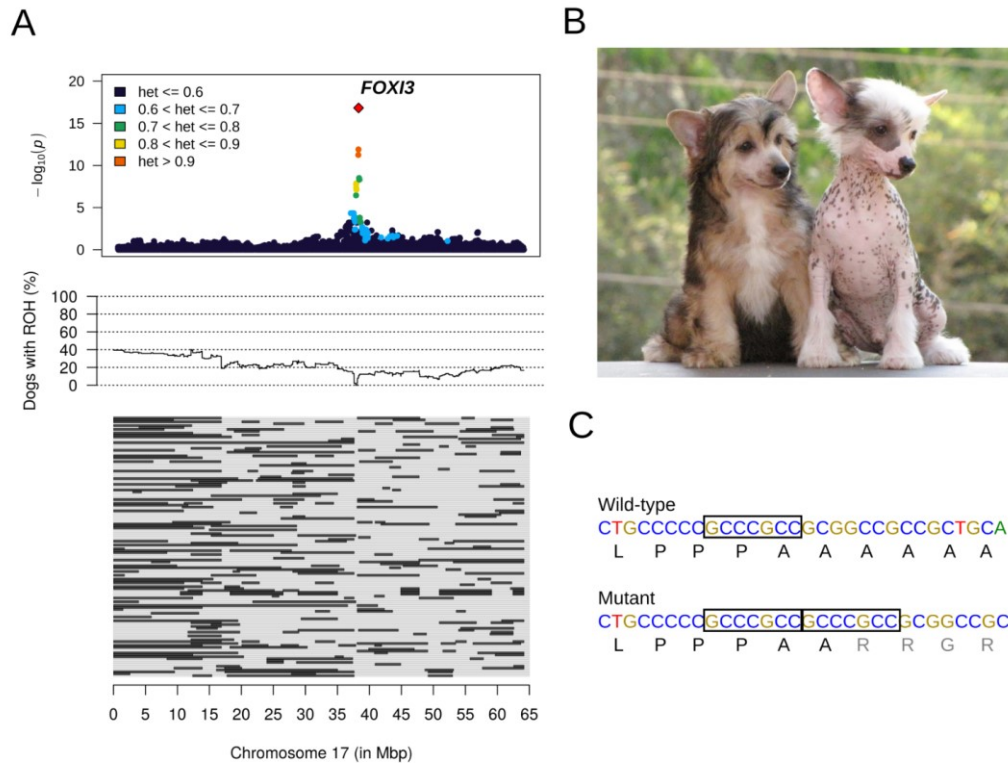


Figure 3. Detection of a homozygous-deficient haplotype (HDH) tagging the homozygous-lethal mutation of *FOXI3* in Chinese Crested (CC) dogs. **(A)** The top panel displays the significance of haplotypes (points) along coordinates of CFA17, with the highest point mapping to *FOXI3*. Colors portray the level of heterozygosity of each haplotype. The middle panel shows locus-specific inbreeding with lack of autozygosity around the position of the *FOXI3* variant. The lower panel exhibits every CC individual as a horizontal gray bar, with black areas representing runs of homozygosity (ROH). **(B)** Aspect of the dominant ectodermal dysplasia phenotype (hairless dog on the right) as compared to the wild type phenotype (powderpuff dog on the left), characterized by absent or sparse hair coat. **(C)** The deleterious variant tagged by the HDH is a 7 bp duplication in exon 1 causing a shift in the reading frame of *FOXI3* (Drögemüller et al., 2008).

3.5 HDHR8 indicates selection against the homozygous wild type *CBD103* genotype in Golden Retriever dogs

The HDHR8 segment (CFA16:58.88-59.00 Mbp) found in GR dogs contained the canine β -defensin gene (*CBD103*). The Δ G23 variant (rs851502010) responsible for the dominant black coat color phenotype (Candille et al., 2007) was found to negatively

correlate with haplotype predictions in the four sequenced GR dogs. These data indicated that the wild type allele was the unfavorable variant. The $\Delta G23$ variant is a 3-bp inframe deletion that causes the removal of a glycine residue at the N-terminus of the resulting protein, which leads to a higher binding affinity between β -defensin and the melanocortin 1 receptor (Mc1r) of melanocytes. Consequently, the $\Delta G23$ allele promotes a higher production of eumelanin, resulting in a dominant black coat color. Leonard et al. (2012) analyzed a sample of 13 GR dogs and found that 46% were homozygous for the $\Delta G23$ mutation whereas the remaining 54% were heterozygous, translating into frequencies of 73% and 27% for the $\Delta G23$ and the wild type alleles, respectively. At a first glance, these data seem inconsistent since GR dogs have yellow/brown instead of black coat color. However, GR are highly selected for a premature stop codon (p.Arg206STOP or R306ter) in the Mc1r gene (*MC1R*), which switches eumelanin production to pheomelanin. Although it has been proposed that the loss-of-function mutation at *MC1R* leads to a yellow/golden phenotype regardless of *CBD103* or Agouti (*ASIP*) genotypes (CANDILLE et al., 2007), this model does not explain why Leonard et al. (2012) could not find at least one GR dog homozygous for the *CBD103* wild type allele (as expected under HWE) or why HDHR8 presents deficiency of homozygosity for the haplotype tagging the wild type variant in the present study. In fact, it is still not yet clear whether the R306ter variant causes complete or partial loss of function of Mc1r. Indeed, the R306ter variant is predicted to cause the loss of 12 amino-acids at the C-terminal tail of Mc1r, which are required for receptor localization and trafficking but not for functioning (Newton et al., 2000). Thus, it could be hypothesized that R306ter causes a decreased localization of Mc1r to the melanocyte membrane, but not its complete absence. If this hypothesis holds, epistasis between *MC1R* and *CBD103* genotypes would be expected and the $\Delta G23$ mutation would contribute to darker hues of yellow/golden in GR dogs due to partial production of eumelanin. Since GR dogs exhibit color lightness variation, homozygous wild type *CBD103* genotypes could be associated with lighter yellow coat, which might be considered undesirable by some breeders. This hypothesis is supported by the observation of R306ter homozygous and wild type *CBD103* homozygous Akita-inu dogs being white (Oguro-Okano et al., 2011), and by the fact that the white/cream coat color in GR is most likely explained by the autosomal recessive inheritance of an unknown variant

(Schmutz e Berryere, 2007). Additionally, introgression of R306ter from domestic dogs to Newfoundland coyotes (*Canis latrans*), supposedly contributed by hybridization with GR dogs, resulted in white coyotes (Brockerville et al., 2013). **Figure 4** summarizes the proposed epistasis mechanism between *MC1R* and *CBD103* genotypes in GR dogs.

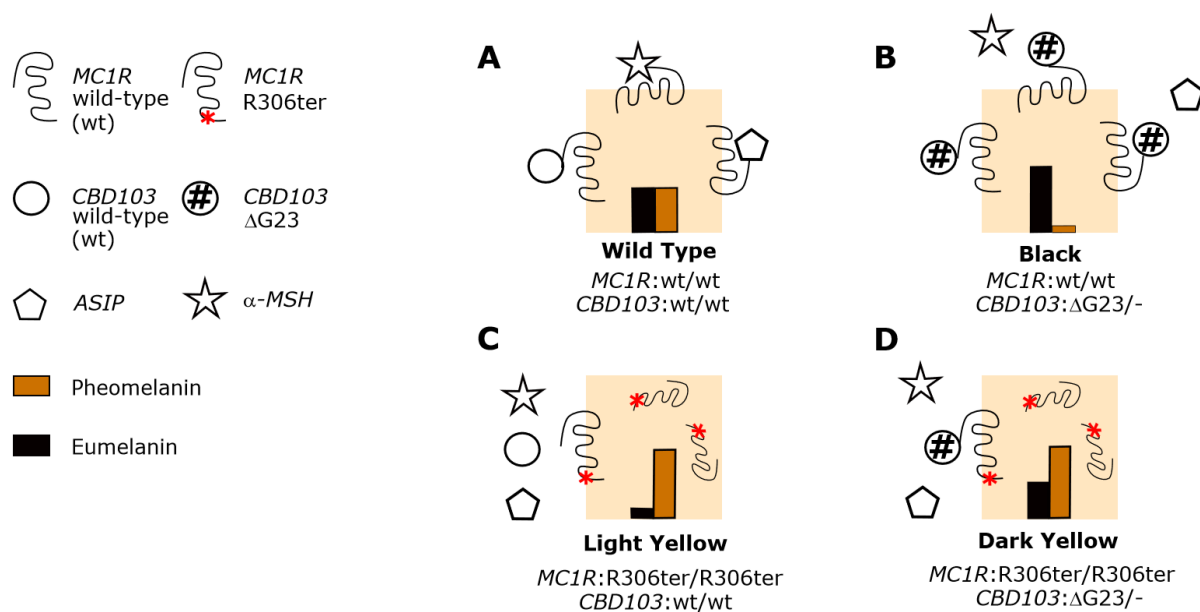


Figure 4. Proposed model for epistatic interactions between *MC1R* and *CBD103* mutations in dogs. **(A)** Competitive binding of *CBD103*, *ASIP* and α -*MSH* gene products to the melanocortin receptor produces an Agouti-like wild type coat color, characterized by mixtures of eumelanin and pheomelanin. **(B)** The Δ G23 mutation enhances binding affinity of *CBD103* and *MC1R* gene products, resulting in the predominance of eumelanin production and black coat color. **(C)** The R306ter mutation decreases trafficking and localization of *MC1R* gene products to the melanocyte membrane, yielding predominance of pheomelanin production and light yellow, white or cream coat colors. **(D)** Presence of both Δ G23 and R306ter mutations results in few melanocortin receptors in the melanocyte membrane to bind with *CBD103* ligands with higher likelihood as compared with *ASIP* or α -*MSH* gene products, producing darker hues of yellow/golden coat color.

3.6 Black-and-tan coat color patterning as the putative unfavorable recessive phenotype underlying HDHR10 in Labrador and Golden Retriever dogs

Homozygotes for the significant haplotype at HDHR10 were rare in GR dogs (three times more homozygotes were expected) and completely absent in LR dogs. Haplotype predictions were consistent with genotypes at two insertions-deletions (INDELs) in introns

1 and 2 of the T-box 15 gene (*TBX15*). These INDELs were predicted to affect two canine-specific short interspersed nuclear elements (SINE_Cf). The *TBX15* murine orthologue has been associated with black-and-tan coat color (Candille et al. 2004). Indeed, the black-and-tan coat color was prevalent in the ancestral retriever population and deemed to follow an autosomal recessive inheritance. A SINE insertion in *ASIP*, which segregates in both GR and LR, was previously incriminated as the cause of the black-and-tan phenotype in several dog breeds (Dreger e Schmutz, 2011). Although epistasis among *ASIP*, *MC1R* and *CBD103* alleles are deemed to prevent the black-and-tan phenotype in retriever dogs, black-and-tan progeny are known to occur at very low frequency in both breeds. Therefore, it is an opened question whether yet another mutation plays a role in determining black-and-tan coat color in retriever dogs. Moreover, the effects of *TBX15* on color patterning are most likely independent from *ASIP* mutations (Mills e Patterson, 2009). We thus hypothesize that a novel *TBX15* mutation, for which the two SINE_Cf identified here are candidates, also contributes epistatically to the appearance of the recessive black-and-tan phenotype in GR and LR. In this scenario, selection against animals that were homozygous for the unknown *TBX15* variant would then explain the HDHR10 signal.

3.7 HDHR16 points to *COL25A1* as another candidate gene harboring unfavorable alleles in retriever dogs

The HDHR16 signal was consistent with two closely located INDELs (4 bp deletion at CFA32:28690728-28690731 and 6 bp insertion at CFA32:29321784) within intron 1 of the collagen type XXV alpha 1 chain gene (*COL25A1*). Mutations in *COL25A1* have been linked with hereditary cranial dysinnervation and ocular muscle disorders in humans (Shinwari et al., 2015), as well as Alzheimer's disease-like pathology in mice (Tong et al., 2010). Although we could not point to a specific mutation or phenotype, putative traits involved with HDHR16 may include those related with ocular, neuronal or muscular activity.

3.8 HDHR3 contains the largest CNV in the dog genome spanning *AMY2B*

The HDHR3 locus on CFA6:44250000-49000000 was shared between GR and LR, and sheltered a ~1.5 Mbp copy number variant (CNV) spanning the pancreatic alpha-amylase 2b gene (*AMY2B*), RNA binding region containing 3 gene (*RNPC3*) and collagen type XI alpha 1 chain (*COL11A1*). This structural variant has been reported by previous re-sequencing projects (Nicholas et al., 2009; Berglund et al., 2012) and is the largest CNV in the dog genome described to date. The amylase gene shelters strong selection signatures and copy number gains in dogs, suggesting a post-domestication selection for starch-rich diets (Axelsson et al. 2013). Arendt and colleagues (2014; 2016) estimated that GR and LR carry on average ~13 and ~12 copies of *AMY2B*, respectively. Although we could not confirm carrier status in the present study, this complex CNV region seems the most plausible site sheltering the mutation causing deficiency of haplotype homozygosity at HDHR3. Given the complexity of this CNV region, we hypothesize that the intense selective pressure for ability to digest starch, which required evolutionary innovation and gene amplification in the chromosomal domain encompassing *AMY2B*, may have caused deleterious mutations to increase in frequency through genetic draft. In this case, positive or even balancing selection on the amplified copies of *AMY2B* must have been strong enough to counteract the action of purifying selection on hitchhiking disadvantageous mutations. Interestingly, most re-sequenced dogs presented a deletion of the neighboring segment flanking *RNPC3* and *COL11A1* in previous studies, suggesting the deletion to be the ancestral allele (at least in the domestic lineage), whereas samples from LR and Boxer have been observed to carry complex duplications (Nicholas et al., 2009; Berglund et al., 2012), all assumed to be derived alleles. These duplications might represent candidate deleterious mutations since *COL11A1* is involved with osteochondrodysplasia and connective tissue syndromes in humans (Kohmoto et al., 2016).

3.9 HDHR1, HDHR13 and HDHR15 contain candidate genes implicated in retinal disease

Two EM-specific regions (HDHR13 and HDHR15) and one GR-specific region (HDHR1) contained genes that could be linked with progressive retinal atrophy (PRA). This retinal disease is homologous to the human retinitis pigmentosa and is common in several dog breeds, including GR (Downs et al., 2011) and EM (Heitmann et al., 2005). The HDHR1 region spanned the pre-mRNA processing factor 31 homolog (*PRPF31*) and the CCR4-NOT transcription complex subunit 3 (*CNOT3*) genes. Epistasis between these two genes causes incomplete penetrance of retinitis pigmentosa in humans (Venturini et al., 2012; Rose et al., 2014). Retinitis only occurs when an individual carries at least one copy of the causal mutation at *PRPF31* (autosomal dominant) and two copies of a high-expressing allele of *CNOT3* (autosomal recessive). Also, murine embryos carrying null *CNOT3* or *PRPF31* alleles die during early development (Dickinson et al., 2016). Given the large number of genes and variants implicated in retinitis pigmentosa, the complexity of the mode of inheritance, and the prevalence of PRA in the GR breed (Downs et al., 2014), an involvement of *CNOT3* and *PRPF31* with HDHR1 could be hypothesized. Moreover, the HDHR13 and HDHR15 regions in EM also contained genes incriminated in retinitis pigmentosa in humans, namely S-antigen visual arrestin (*SAG*), phosphodiesterase 6D (*PDE6D*) and structural maintenance of chromosomes 3 (*SMC3*) (Sullivan et al., 2017; He et al., 2008). Since PRA can be triggered by mutations in several different genes, and is generally selected against by dog breeders, the observation that 3 out of the 16 candidate regions detected in the present study harbored genes implicated in retinal disease may be relevant.

3.10 HDHR2 points to an obesity-related gene with high differentiation between dogs and wild canids

The HDHR2 region was also exclusive to the GR breed and mapped to CFA3:52.75-54.00 Mbp. Five times more dogs were expected to be homozygous for the significant haplotype than what was observed. Out of the four sequenced GR dogs, only

one was a carrier for the significant haplotype. A total of 467 sequence variants were filtered based on genotypes of that single dog, and from which only 4 were missense variants. One of these variants (rs851527235) was located on the CREB regulated transcription coactivator 3 gene (*CRTC3*), which has been found to shelter higher structural diversity in domestic dogs as compared to wild canids (Ramirez et al., 2014). This gene is also involved with energy balance, insulin-sensitivity and adipose tissue mass, traits that are linked to obesity (Song et al., 2010). Heritable obesity is common in dogs (Switonski et al., 2013), especially in the GR breed. Although we could not confirm causality of rs851527235 or the involvement of *CRTC3* in the HDHR2 signal, associations between HDHR2 and heritable obesity should be investigated in depth in future studies.

3.11 The common HDHR4 locus maps to a complex copy number variant encompassing *ARHGAP27*

We took advantage of whole genome sequence data of 8 EM, 5 GR and 3 LR dogs to explore the HDHR4 locus (CFA9:15.00-18.50 Mbp), which appears to shelter haplotypes with deficiency of homozygosity in all four breeds studied here. In total, 7 EM, 1 GR and 0 LR were predicted to be heterozygous for the unobserved mutation, and a single EM dog was predicted to be a rare homozygous. A total of 138 variants located between CFA9:16917716-18205230 were in perfect correlation with haplotype predictions and were further investigated. Inspection of read alignments supporting these variants revealed a complex CNV spanning positions 18.19 – 18.24 Mbp on CFA9. This CNV contained the gene encoding the Rho GTPase activating protein 27 (*ARHGAP27*) and a long intergenic non-coding RNA (lincRNA). This complex CNV has been previously reported in several dog breeds (Berglund et al., 2012), which supports the appearance of the HDHR4 signal in all four breeds in this study. However, the HDHR4 region spanned over 3 Mbp and we could not narrow it down to individual genes or variants.

3.12 HDHR5 found in Entlebucher Mountain dogs encompasses a gene implicated in reduced fertility

Using EM sequence data, we found a 4 bp insertion (CFA10:16759326) within the HDHR5 segment mapping to the intronic region of two genes that overlap at the opposing DNA strands, namely the synaptonemal complex central element protein 3 gene (*SYCE3*) and the non-SMC condensin II complex subunit H2 gene (*NCAPH2*). In particular, *SYCE3* is required for the progression of homologous recombination during meiosis (LU et al., 2014), and deletion of *SYCE3* in mice causes both male and female infertility due to meiotic arrest (Schramm et al., 2011). Given that mean litter size and rate of mating success decreased in the EM breed in the past years, and that inbreeding alone has been found to contribute with only small negative effects to fertility in the EM breed (Schrack et al., 2017), infertility and the *SYCE3* variant are plausible candidates underlying the HDHR5 signal in our sample of EM dogs.

3.13 HDHR6, HDHR7, HDHR11, HDHR12 and HDHR14 may be linked with homozygous-lethal mutations

Near zero homozygotes were observed in regions HDHR6, HDHR7, HDHR12 and HDHR14, suggesting that these haplotypes may partially tag homozygous-lethal mutations. Since our sequence analysis was limited to a small number of dogs, we were unable to pinpoint putative deleterious mutations for each one of these regions, although some candidates were suggested from our haplotype analysis. These regions should be targets of future in-depth investigation for the search of lethal alleles. Using average litter size data from 64 female EM dogs, we were able to associate HDHR7 with a significant decrease in litter size, estimated at -0.61 ± 0.29 puppies per litter ($p = 0.042$). This finding suggests a homozygous-lethal mutation underlying HDHR7.

4. Discussion

We reported the first genome-wide screening for haplotypes with deficiency of homozygosity segregating at high frequency in domestic dogs. Our analysis revealed 16 homozygous-deficient haplotype regions (HDHRs) across four inbred canine lineages. More than one third of these HDHRs were linked with either hair coat phenotypes or retinal disease. As for the remaining HDHRs, half of them were predicted to tag lethal alleles, whereas the other half were predicted to underlie various unfavorable recessive phenotypes. We were also able to map putative causal nucleotides responsible for some of these HDHR signals, which are now available for further investigation.

The occurrence of HDHR in the dog genome may result from a variety of distinct phenomena. For instance, the same allele may cause a dominant phenotype that is artificially or naturally selected for, and at the same time cause an unfavorable recessive phenotype that would be eliminated by the action of purifying selection otherwise. One example is the *FOXI3* frameshift variant in Chinese Crested dogs (HDHR9), which promotes the dominant hairless phenotype that is a hallmark of the breed, but is also lethal in the homozygous state (Drögemüller et al., 2008). Selective pressure on the dominant phenotype might be so strong that, even if the allele is homozygous-lethal, the mutation is kept at high frequency in the heterozygous state in the population. Similar cases have also been observed in dairy cattle populations, where a homozygous-lethal deletion of four genes was kept at high frequency in Nordic red cattle due to its positive effects on milk production (Kadri et al., 2014).

Similar to the scenario above, mutations causing unfavorable recessive phenotypes may also produce HDHR if they are in linkage disequilibrium with variants under selection, in which case they would hitchhike to high frequency. In the present study we found that the highly selected *AMY2B* locus, which was extremely important for the adaptation of the domestic dog to starch-rich diets (Axelsson et al. 2013; Arendt et al. 2014; 2016), may also contain mutations linked with unfavorable recessive phenotypes in retriever dogs.

Another possible source of HDHR is inbreeding and genetic drift. In fact, this has been the main driver behind the use of haplotypes to detect unfavorable mutations in

livestock species (VanRaden et al. 2011; Sahana et al., 2013; Hoff et al., 2017; Häggman e Uimari, 2017). The heavy use of specific sires and dams in livestock breeding programs may result in the increase of frequency of deleterious mutations, and consequently economic losses. The same rationale could be applied to dogs, since the high incidence of congenital and hereditary diseases in the species is in part explained by inbreeding. The series of HDHR pointing to retinal disease, obesity-related traits and infertility may be an example of disease alleles prevailing in the population thanks to inbreeding.

Finally, other more complex mechanisms might also underlie the sustained presence of HDHRs in the dog genome. For instance, epistasis between the *CBD103* and *MC1R* loci may be responsible for keeping the wild type *CBD103* allele in Golden retrievers at high frequency but preferably at the heterozygous state with the $\Delta G23$ mutation. If our model is correct, heterozygous or $\Delta G23$ -homozygous individuals that are also homozygous for the R306ter *MC1R* mutation might exhibit coat color lightness varying between yellow and dark golden, which is the usual desirable range in the breed. However, R306ter/R306ter dogs that are also homozygous wild type at the *CBD103* locus will present a white or cream coat color, which is less appreciated by breeders. Similar epistatic interactions may be expected for the *TBX15* locus, presumably involved with the rare black-and-tan phenotype in retriever dogs.

5. Conclusion

In conclusion, we showed that screenings for HDHR can reveal haplotypes tagging mutations in the dog genome that are involved with unfavorable recessive phenotypes in the species. Consorting HDHR analyses with whole genome sequence data mining can further contribute to the identification of these mutations and yield sensible predictions of the unobserved phenotypes based on gene function.

6. Funding

Author R.B.P.T. was supported by scholarships from the Coordination for the Improvement of Higher Education Personnel (CAPES) and the Sao Paulo Research

Foundation (FAPESP, process 2017/08373-1). This work was also supported by grants from the European Commission (LUPA-GA.201270).

7. Acknowledgments

We are thankful to Paloma Pegorer from Lapinus Kennel for kindly providing the picture of Chinese Crested puppies used in Figure 3.

8. References

Arendt M, Fall T, Lindblad-Toh K, Axelsson E (2013) Amylase activity is associated with *AMY2B* copy number in dog: implications for dog domestication, diet and diabetes. **Anim Genet** 45:716-722.

Arendt ML, Melin M, et al (2015). Genome-Wide Association Study of Golden Retrievers Identifies Germ-Line Risk Factors Predisposing to Mast Cell Tumours. **PLoS Genetics** 11:e1005647.

Arendt M, Cairns KM, Ballard JW, Savolainen P, Axelsson E (2016) Diet adaption in dog reflects spread of prehistoric agriculture. **Heredity** 117:301-306.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. **Nature** 495:360-364.

Barbato M, Orozco-Terwengel P, Tapio M, Bruford MW (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. **Front Genet** 6:109.

Berglund J, Nevalainen EM, et al (2012) Novel origins of copy number variation in the dog genome. **Genome Biol** 13:R73.

Blake JA, Eppig JT, Kadin JA, Richardson JE, Smith CL, Bult CJ (2017) The Mouse Genome Database Group. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. **Nucleic Acids Res** 45:D723-D729.

Brockerville RM, McGrath MJ, Pilgrim BL, Marshall HD (2013) Sequence analysis of the three pigmentation genes in the Newfoundland population of *Canis latrans* links the Golden Retriever Mc1r variant to white coat color in coyotes. **Mamm Genom** 24:134-141.

Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS (2007) A β -Defensin mutation causes black coat color in domestic dog. **Science** 318:1418-1423.

Candille SI, Van Raamsdonk CD, Chen C, Kuijper S, Chen-Tsai Y, Russ A, Meijlink F, Barsh GS (2004) Dorsoventral patterning of the mouse coat by *TBX15*. **PLoS Biol** 1: E3.

Casper J, Zweig AS, et al (2018) The UCSC Genome Browser database: 2018 update. **Nucleic Acids Res** 46:D762-D769.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. **Gigascience** 4.

Dickinson ME, Flenniken AM, et al (2016) High-throughput discovery of novel developmental phenotypes. **Nature** 537:508-514.

Downs LM, Wallin-Håkansson B, Bergström T, Mellersh CS (2014) A novel mutation in *TTC8* is associated with progressive retinal atrophy in the golden retriever. **Canine Genet Epidemiol** 1.

Downs LM, Wallin-Håkansson B, Boursnell M, Marklund S, Hedhammar A, Truvé K, Hubinette L, Lindblad-Toh K, Bergstrom T, Mellersh CS (2011) A frameshift mutation in golden retriever dogs with progressive retinal atrophy endorses *SLCA3* as a candidate gene for human retinal degenerations. **PLoS One** 6:e21452.

Dreger DL, Schmutz SM (2011) A SINE insertion causes the black-and-tan and saddle tan phenotypes in domestic dogs. **J Hered** 1:S11-18.

Drögemüller C, Karlsson EK, Hytönen MK, Perloski M, Dolf G, Sainio K, Lohi H, Lindblad-Toh K, Leeb T (2008) A mutation in hairless dogs implicates *FOXI3* in ectodermal development. **Science** 321:1462.

Ferenčaković M, Sölkner J, Curik I (2013) Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. **Genet Sel Evol** 45:42.

Forsberg SK, Kierczak M, et al (2015) The Shepherds' Tale: A Genome-Wide Study across 9 Dog Breeds Implicates Two Loci in the Regulation of Fructosamine Serum Concentration in Belgian Shepherds. **PLoS ONE** 10:e0123173.

Häggman J, Uimari P (2016) Novel harmful recessive haplotypes for reproductive traits in pigs. **J Anim Breed Genet** 134:129-135.

He S, Parapuram SK, Hurd TW, Behnam B, Margolis B, Swaroop A, Khanna H (2008) Retinitis pigmentosa GTPase Regulator (RPGR) protein isoforms in mammalian retina insights into X-linked retinitis pigmentosa and associated ciliopathies. **Vision Res** 48:366-376.

Heitmann M, Hamann H, Brahm R, Grußendorf H, Rosenhagen CU, Distl O (2005) Analysis of prevalence of presumed inherited eye diseases in Entlebucher Mountain Dogs. **Veterinary Ophthalmology** 8:145-151.

Hoff JL, Decker JE, Schnabel RD, Taylor JF (2017) Candidate lethal haplotypes and causal mutations in Angus cattle. **BMC Genomics** 18:799.

Howrigan DP, Simonson MA, Keller MC (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. **BMC Genomics** 12:460.

Kadri NK, Sahana G, et al (2014) A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. **PLoS Genet** 10:e1004049.

Kim ES, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, Liu G, Da Y, Sonstegard TS (2013) Effect of artificial selection on runs of homozygosity in u.s. Holstein cattle. **PLoS One** 8:e80813.

Kinsella RJ, Kähäri A, et al (2011) Ensembl BioMart: a hub for data retrieval across taxonomic space. **Database (Oxford)** 23:bar030.

Kohmoto T, Tsuji A, Morita K, Naruto T, Masuda K, Kashimada K, Enomoto K, Morio T, Harada H, Imoto I (2016) A novel *COL11A1* missense mutation in sblings with non-ocular Stickler syndrome. **Hum Genome Var** 3:16003.

Leonard BC, Marks SL, Outerbridge CA, Affolter VK, Kananurak A, Young A, Morre PF, Bannasch DL, Bevins CL (2012) Activity, expression and genetic variation of canine beta-defensin 103: a multifunctional antimicrobial peptide in the skin of domestic dogs. **J Innate Immun** 4:248-259.

Lewis TW, Abhayaratne BM, Blott SC (2015) Trends in genetic diversity for all Kennel Club registered pedigree dog breeds. **Canine Genet Epidemiol** 2:13.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics** 25:1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. **Bioinformatics** 25: 2078-2079.

Lu J, Gu Y, Feng J, Zhou W, Yang X, Shen Y (2014) Structural insight into the central element assembly of the synaptonemal complex. **Sci Rep** 4:7059.

Marsden CD, Ortega-Del Vecchyo D, et al (2016) Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. **Proc Natl Acad Sci U S A** 113:152-157.

Mclaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F (2016) The Ensembl Variant Effect Predictor. **Genome Biol** 17:122.

Mcquillan R, Leutenegger AL, et al (2008) Runs of homozygosity in European populations. **Am J Hum Genet** 83:359-372.

Mills MG, Patterson LB (2009) Not just black and white: pigment pattern development and evolution in vertebrates. **Semin Cell Dev Biol** 20:72-81.

Molin AM, Berglund J, Webster MT, Lindblad-Toh K (2015) Genome-wide copy number variant discovery in dogs using the CanineHD genotyping array. **BMC Genomics** 15.

Mouse Genome Informatics (MGI) **The Jackson Laboratory**, Bar Harbor, Maine. Disponível em: <<http://www.informatics.jax.org>, 2018>. Acesso em: 28 nov 2018.

Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R (2016) BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. **Bioinformatics** 32:1749-1751.

Newton JM, Wilkie AL, He L, Jordan SA, Metallinos DL, Holmes MN, Jacksin HI, Barsh GS (2000) Melanocortin 1 receptor variation in the domestic dog. **Mamm Genome** 11:24-30.

Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM (2009) The genomic architecture of segmental duplications and associated copy number variants in dogs. **Genome Res** 19:491-499.

O'Connell J, Gurdasani D, et al (2014) A general approach for haplotype phasing across the full spectrum of relatedness. **PLoS Genet** 10:e1004234.

Oguno-Okano M, Honda M, Yamazaki K, Okano K (2011) Mutations in the melanocortin 1 receptor, β -defensin103 and agouti signaling protein genes, and their association with coat color phenotypes in Akita-inu dogs. **J Vet Med Sci** 73:853-858.

Olsson M, Tengvall K, et al (2015) Genome-Wide Analyses Suggest Mechanisms Involving Early B-Cell Development in Canine IgA Deficiency. **PLoS ONE** 10:e0133844.

Online Mendelian Inheritance in Animals (OMIA) **Faculty of Veterinary Science**, University of Sydney. Disponível em: <<http://omia.angis.org.au/>>. Acesso em: 28 nov 2018.

Online Mendelian Inheritance in Man (OMIM) **McKusick-Nathans Institute of Genetic Medicine**, Johns Hopkins University (Baltimore, MD). Disponível em: <<https://omim.org/>>, Acesso em: 28 nov de 2018.

Ostrander EA, Wayne RK, Freedman AH, Davis BW (2017) Demographic history, selection and functional diversity of the canine genome. **Nat Rev Genet** 18:705-720.

Purcell S, Neale B, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. **Am J Hum Genet** 81:559-575.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. **Bioinformatics** 26:841-842.

Ramirez O, Olalde I, et al (2014) Analysis of structural diversity in wolf-like canids reveals post-domestication variants. **BMC Genomics** 15.

Rose AM, Shah AZ, Venturini G, Rivolta C, Rose GE, Bhattacharya SS (2014) Dominant PRPF31 mutations are hypostatic to a recessive *CNOT3* polymorphism in retinitis pigmentosa: a novel phenomenon of “linked trans-acting epistasis”. **Ann Hum Genet** 78: 62-71.

Sahana G, Nielsen US, Aamand GP, Lund MS, Guldbrandtsen B (2013) Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. **PLoS One** 8: e82909.

Schmutz SM, Berryere TG (2007) Genes affecting coat colour and pattern in domestic dogs: a review. **Anim Genet** 28:539-549.

Schrack J, Dolf G Reichler IM, Schelling C (2017) Factors influencing litter size and puppy losses in the Entlebucher Mountain dog. **Theriogenology** 95:163-170.

Schramm S, Fraune J, Naumann R, Hernandez-Hernandez A, Höög C, Cooke HJ, Alsheimer M, Benavente R (2011) A novel mouse synaptonemal complex protein is essential for loading of central element proteins, recombination, and fertility. **PLoS Genet** 7:e1002088.

Shinwari JM, Khan A, Awad S, Shinwari Z, Alaiya A, Alanazi M, Tahir A, Poizat C, Al Tassan N (2015) Recessive mutations in *COL25A1* are a cause of congenital cranial dysinnervation disorder. **Am J Hum Genet** 96:147-152.

Song Y, Altarejos J, et al (2010) *CRTC3* links catecholamine signaling to energy balance. **Nature** 468:933-939.

Sullivan LS, Bowne SJ, et al (2017) A novel dominant mutation in *SAG*, the Arrestin-1 gene, is a common cause of retinitis pigmentosa in Hispanic families in the Southwestern United States. **Invest Ophthalmol Vis Sci** 58:2774-2784.

Switonski M, Mankowska M (2013) Dog obesity- the need for identifying predisposing genetic markers. **Res Vet Sci** 95:831-836.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. **Brief Bioinform** 14:178-192.

Tong Y, Xu Y, Scearce-Levie K, Ptáček LJ, Fu YH (2010) *COL25A1* triggers and promotes Alzheimer's disease-like pathology in vivo. **Neurogenetics** 11:41-52.

Tonomura N, Elvers I, et al (2015) Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers. **PLoS Genetics** 11:e1004922.

Utsunomiya YT, Milanese M, Utsunomiya AT, Ajmone-Marsan P, Garcia JF (2016) GHap: an R package for genome-wide haplotyping. **Bioinformatics** 32:2861-2862.

VanRaden PM, Olson KM, Null DJ, Hutchison JL (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. **J Dairy Sci** 94:6153-6161.

Venturini G, Rose AM, Shah AZ, Bhattacharya SS, Rivolta C (2012) *CNOT3* is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. **PLoS Genet** 8:e1003040.

APÊNDICES

APÊNDICE A - Capítulo 2 Supplementary Figures

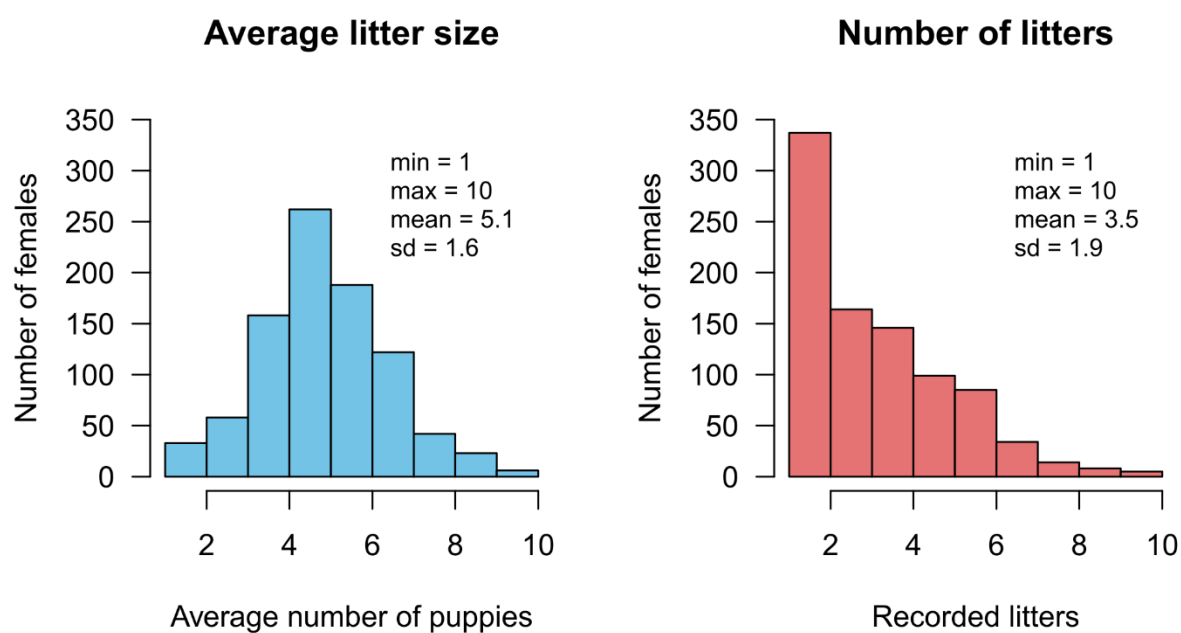


Figure 1A. Histograms and summary statistics of average litter size and number of observed litters for 892 females of the Entlebucher Mountain dog breed.

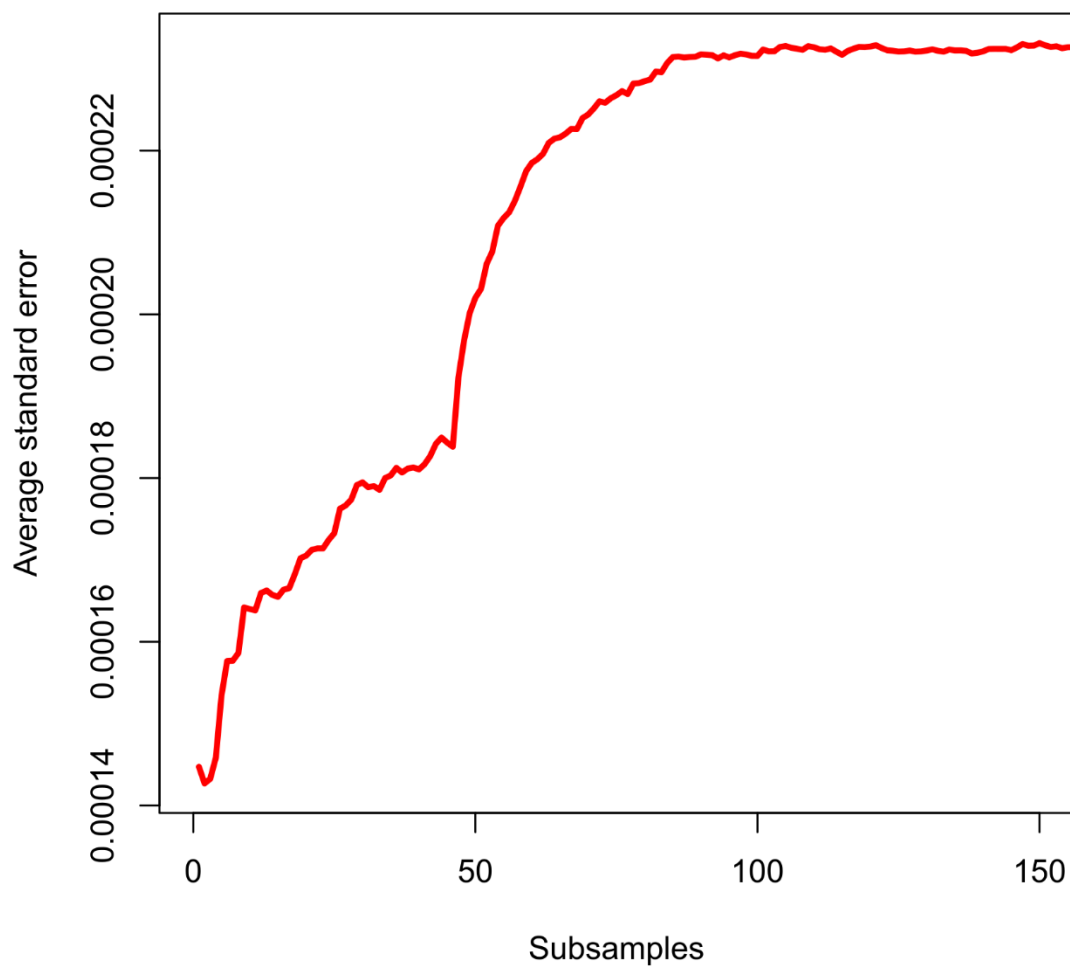


Figure 2A. Convergence plot of standard errors averaged across 87855 SNV markers.

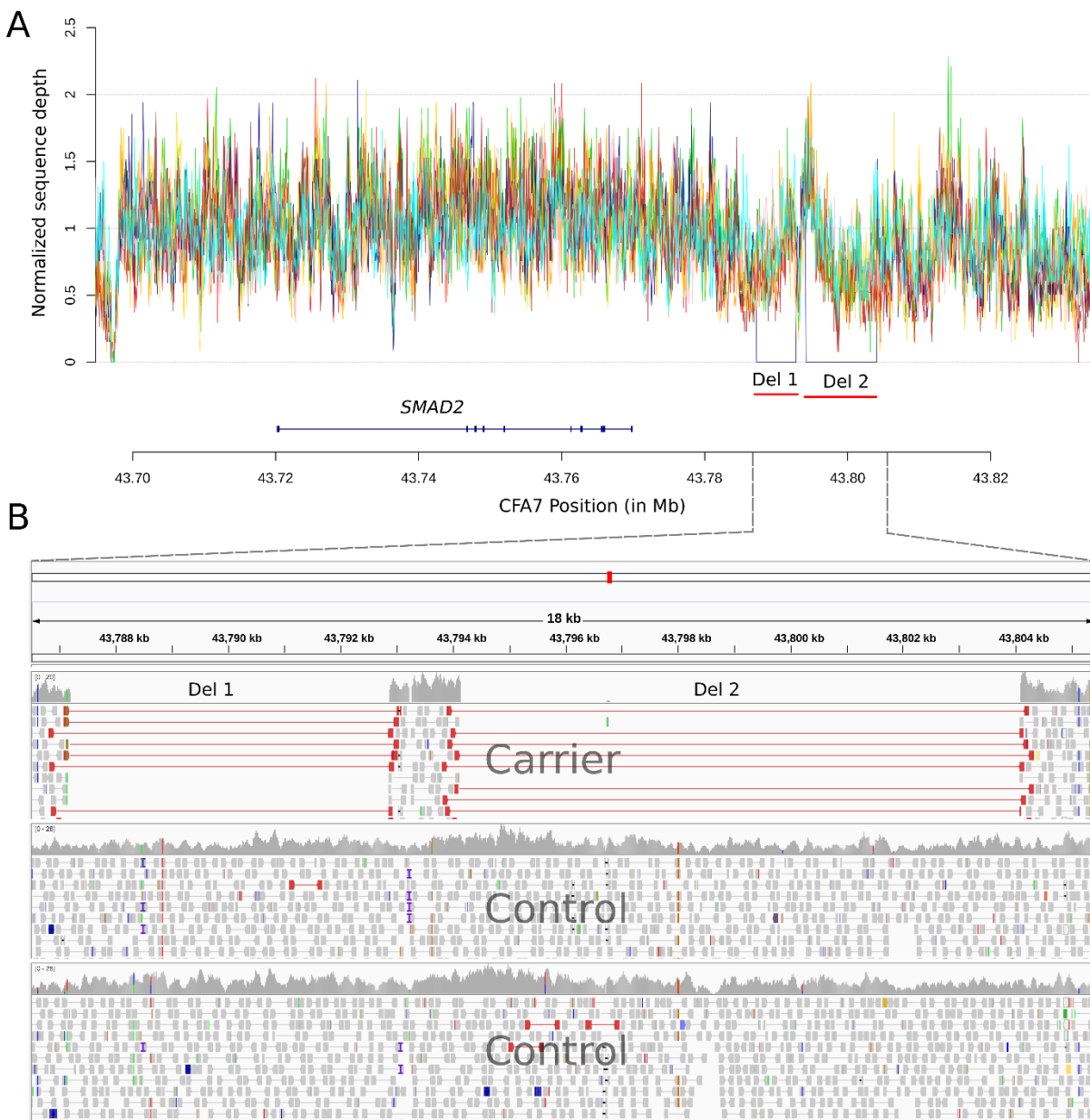


Figure 3A. Genotype status at deletions near *SMAD2* in eight Entlebucher Mountain dogs. **(A)** Sequence depth normalized over 100 bp reveals a single dog carrying the two deletions (Del 1 and Del 2) associated with decreased body size described by Rimbault et al. (2013). **(B)** IGV snapshot of read alignments for the carrier dog, confirming occurrence of the deletions through higher than average insert sizes for paired-end reads (in red) at the deletion position as compared with two control dogs.

APÊNDICE B - Capítulo 3 Supplementary Figures

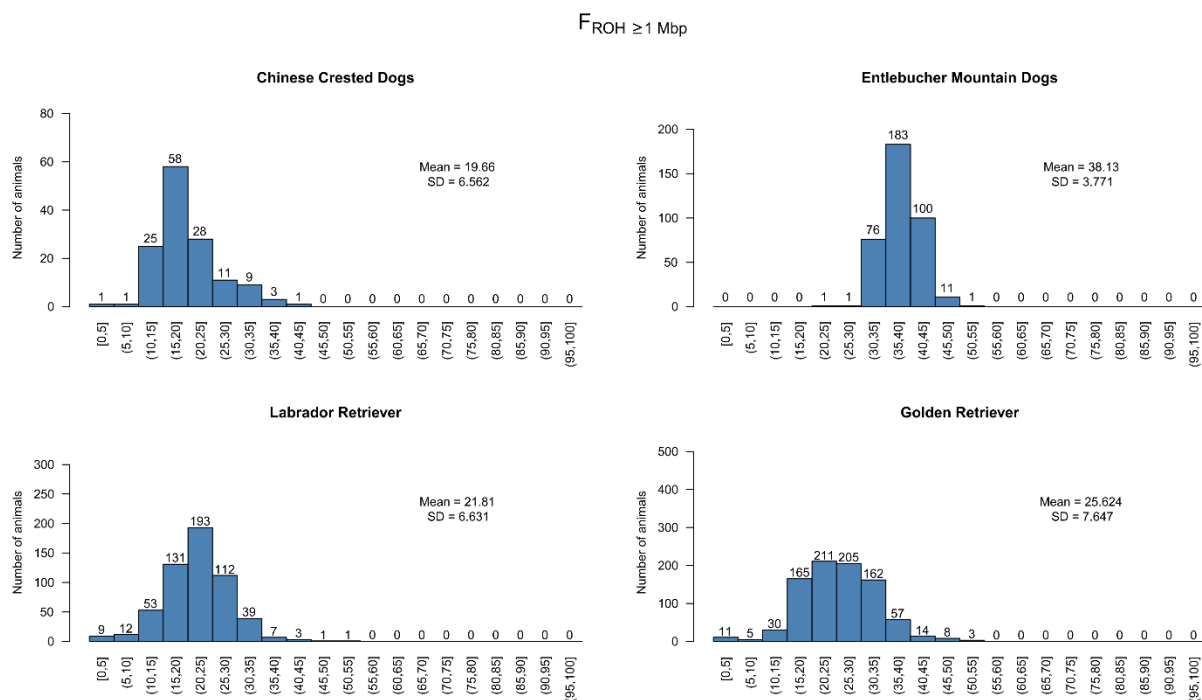


Figure 1B. Genomic autozygosity estimated with runs of homozygosity larger than 1 Mbp.

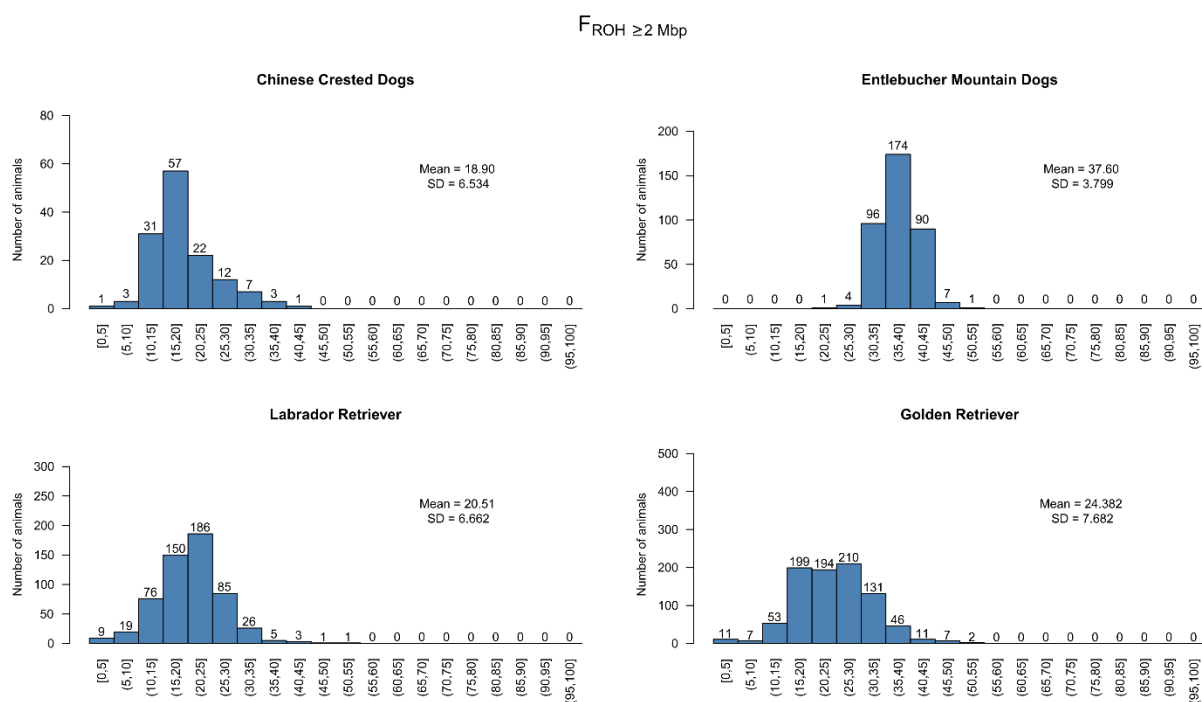


Figure 2B. Genomic autozygosity estimated with runs of homozygosity larger than 2 Mbp.

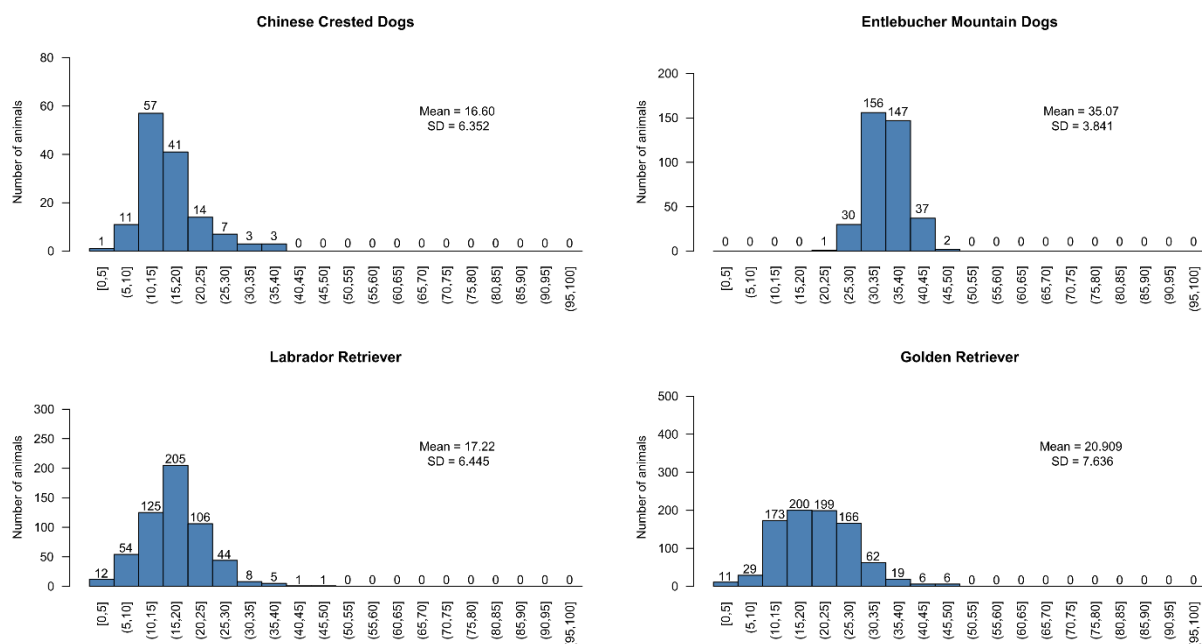
$F_{ROH} \geq 4 \text{ Mbp}$ 

Figure 3B. Genomic autozygosity estimated with runs of homozygosity larger than 4 Mbp.

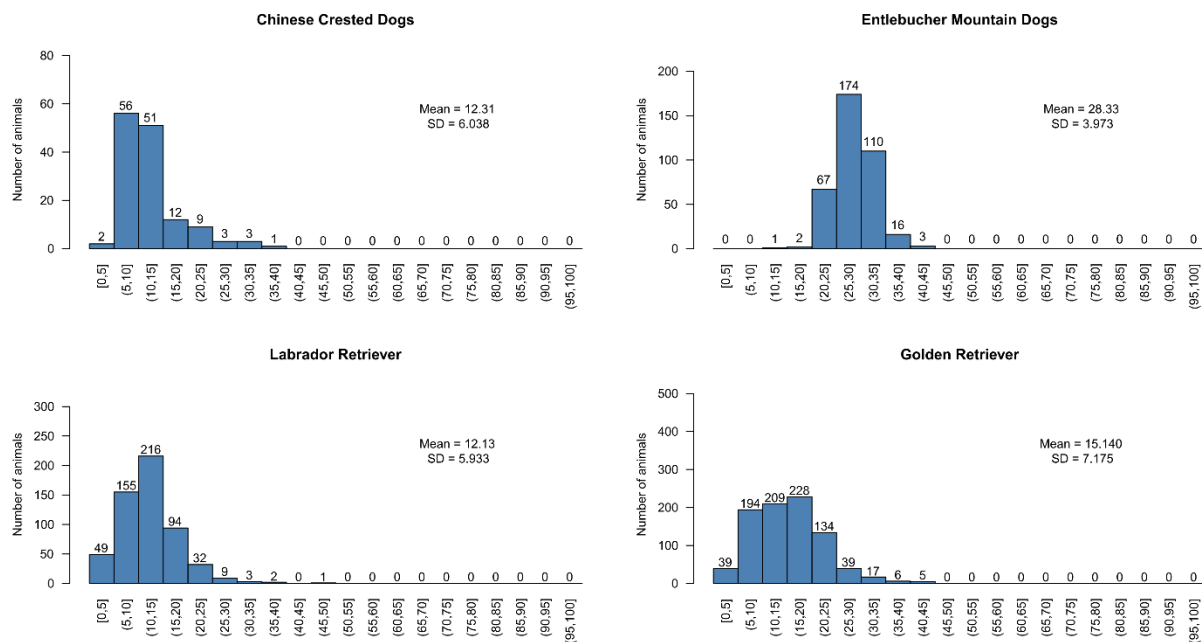
 $F_{ROH} \geq 8 \text{ Mbp}$ 

Figure 4B. Genomic autozygosity estimated with runs of homozygosity larger than 8 Mbp.

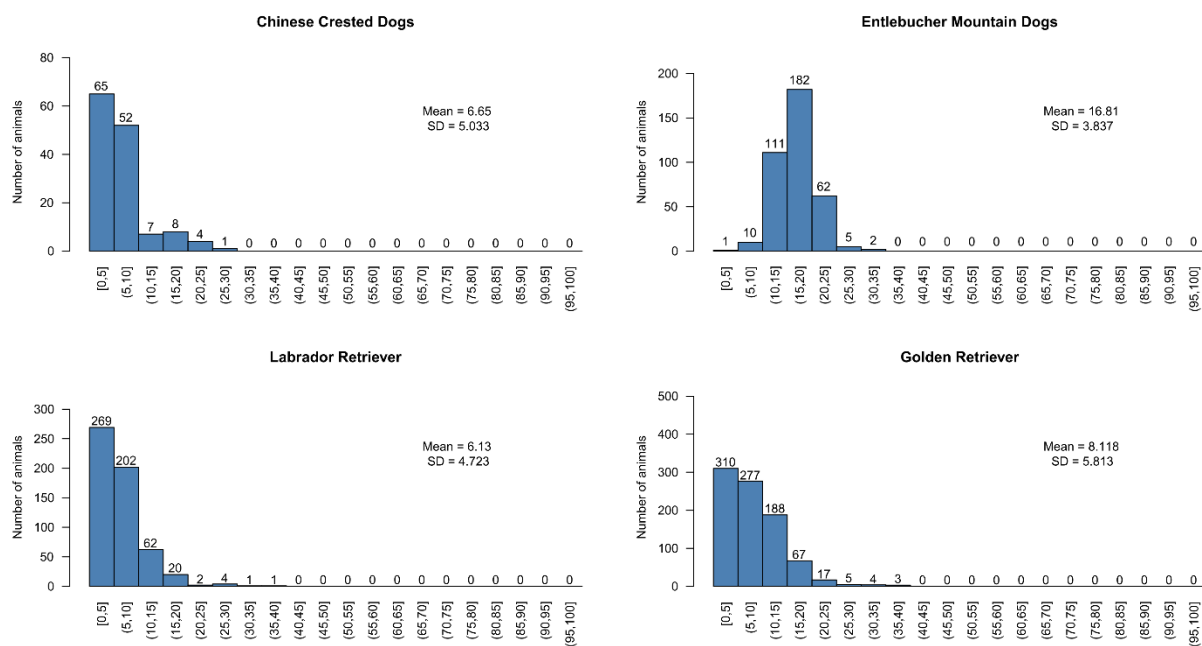
$F_{ROH} \geq 16 \text{ Mbp}$


Figure 5B. Genomic autozygosity estimated with runs of homozygosity larger than 16 Mbp.

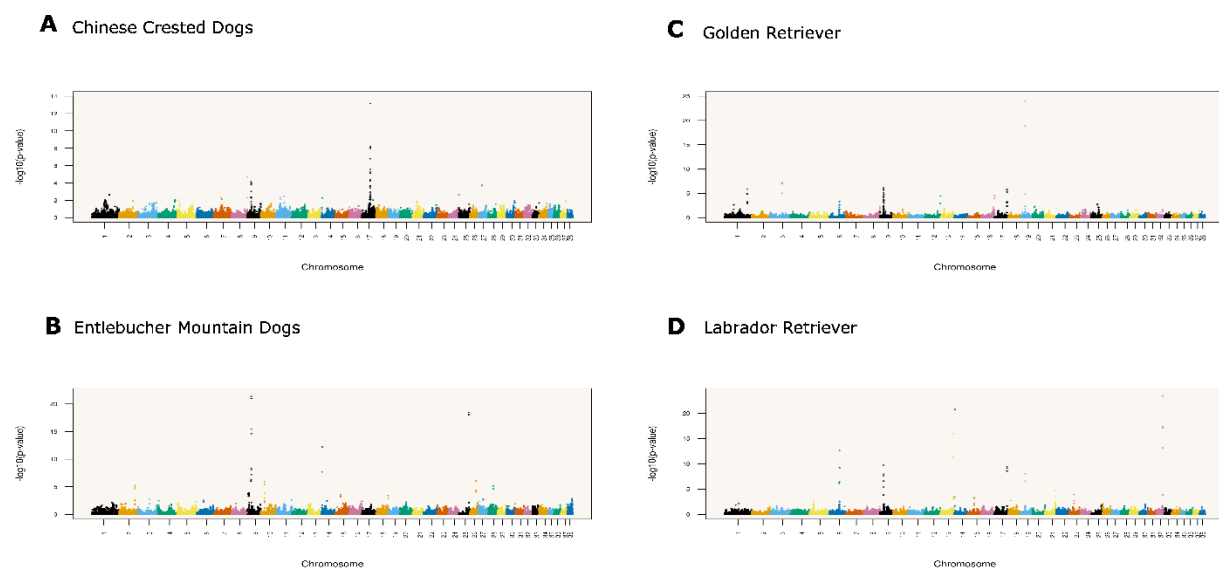


Figure 6B. Manhattan plot of genome-wide screenings for homozygous-deficient haplotypes with window sizes of 500 kbp.

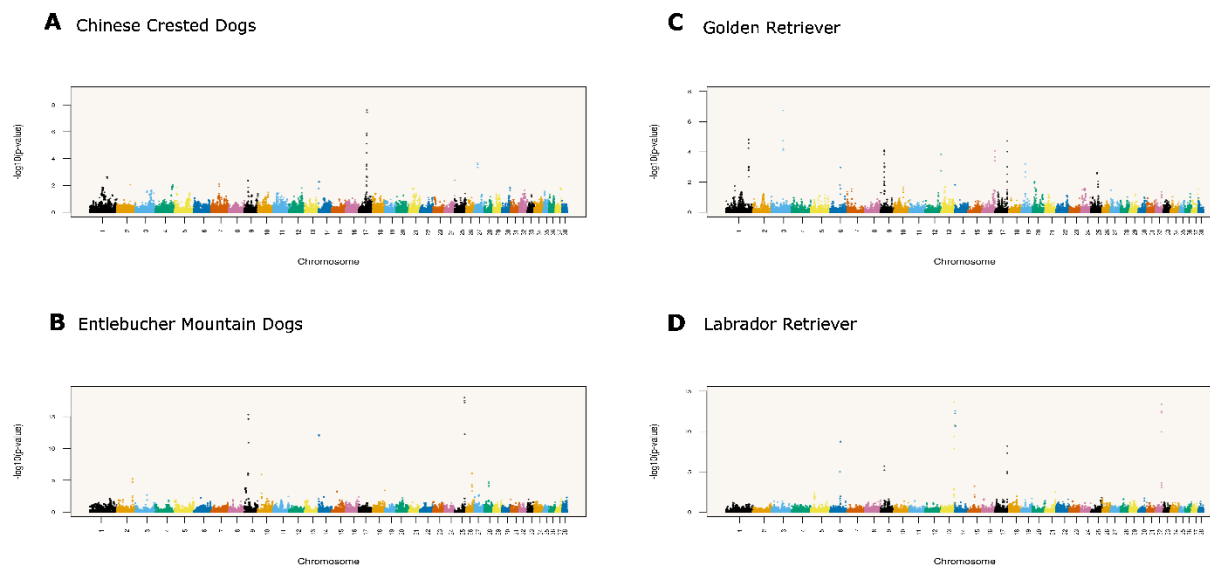


Figure 7B. Manhattan plot of genome-wide screenings for homozygous-deficient haplotypes with window sizes of 1 Mbp.

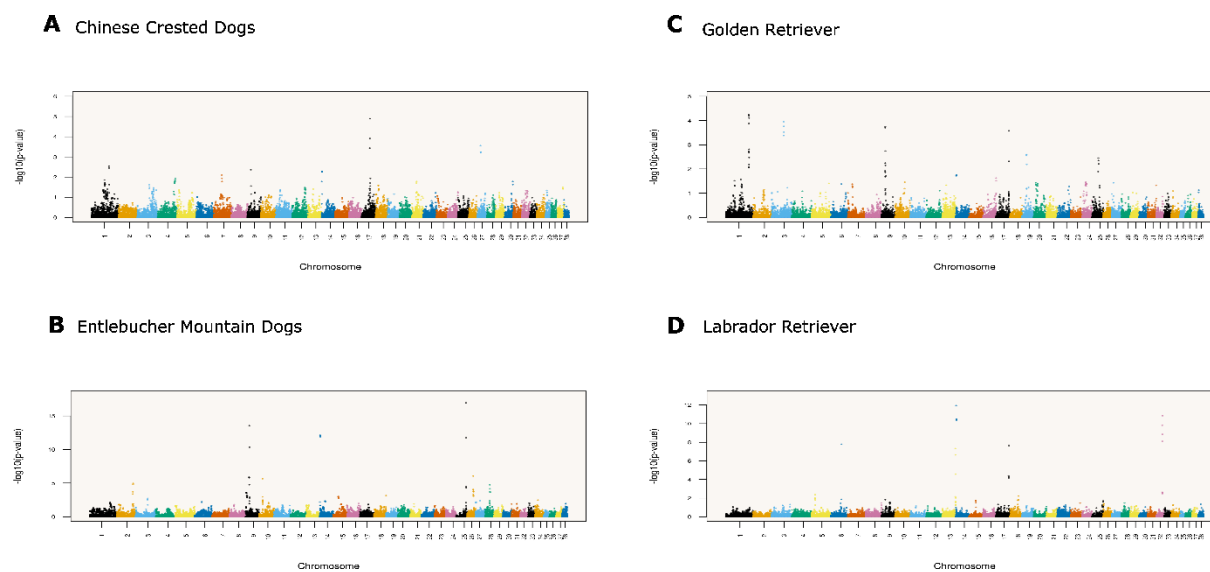


Figure 8B. Manhattan plot of genome-wide screenings for homozygous-deficient haplotypes with window sizes of 2 Mbp.

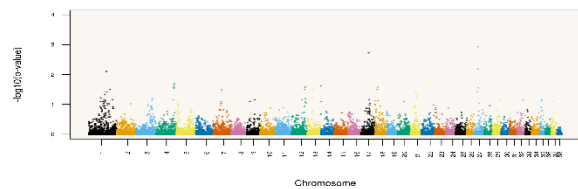
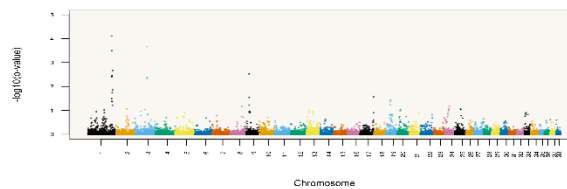
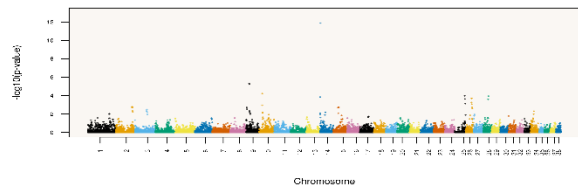
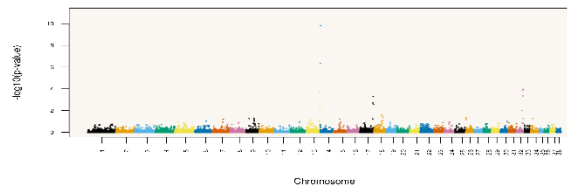
A Chinese Crested Dogs**C** Golden Retriever**B** Entlebucher Mountain Dogs**D** Labrador Retriever

Figure 9B. Manhattan plot of genome-wide screenings for homozygous-deficient haplotypes with window sizes of 5 kbp.