

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA**



DISSERTAÇÃO DE MESTRADO

Caracterização da estrutura genética da raça Guzerá (*Bos indicus*) através de genotipagem em escala genômica

ORIENTADO: Pablo Augusto de Souza Fonseca

ORIENTADORA: Profa. Dr^a. Maria Raquel Santos Carvalho

CO-ORIENTADORA: Prof^a. Dr^a. Maria Gabriela Campolina Diniz Peixoto

BELO HORIZONTE

Setembro de 2014

043

Fonseca, Pablo Augusto de Souza.

Caracterização da estrutura genética da raça Guzerá (*Bos indicus*) através de genotipagem em escala genômica [manuscrito] / Pablo Augusto de Souza Fonseca. - 2014.

26 f. : il. ; 29,5 cm.

Orientadora: Profa. Dra. Maria Raquel Santos Carvalho. Co-orientadora: Profa. Dra. Maria Gabriela Campolina Diniz Peixoto.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas.

1. Guzera (Zebu) - Genética. 2. Variação Genética. 3. Bovinos. I. Carvalho, Maria Raquel Santos. II. Peixoto, Maria Gabriela Campolina Diniz. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título

CDU: 575

Ficha catalográfica elaborada pela Biblioteca do Instituto de Ciências Biológicas da UFMG



**"Caracterização da estrutura genética da raça Guzerá (Bos indicus),
através de genotipagem em escala genômica."**

Pablo Augusto de Souza Fonseca

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Dra. Maria Raquel Santos Carvalho - Orientadora
UFMG

Dra. Maria Gabriela Campolina Diniz Peixoto - coorientadora
EMBRAPA/JUIZ DE FORA

Dr. Ricardo Vieira Ventura
GUELPH/CANADÁ

Dr. Fábio Luiz Buranelo Toral
UFMG

Belo Horizonte, 26 de setembro de 2014.

Pablo Augusto de Souza Fonseca

Caracterização da estrutura genética da raça Guzerá (*Bos indicus*) através de genotipagem em escala genômica

Dissertação apresentada ao programa de Pós-Graduação em Genética do Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do grau de Mestre em Genética.

Área de Concentração: Genômica e Bioinformática

Orientadora: *Profa. Dra. Maria Raquel Santos Carvalho*

Co-orientadora: *Profa. Dra. Maria Gabriela Campolina Diniz Peixoto*

BELO HORIZONTE
INSTITUTO DE CIÊNCIAS BIOLÓGICAS - UFMG
2014

À minha família, por todo apoio,
incentivo, alegrias e aprendizado.

Dedico

“No meio da dificuldade encontra-se a oportunidade.”

Albert Einstein

“Agradeço todas as dificuldades que enfrentei; não fosse por elas, eu não teria saído do lugar.”

Chico Xavier

Agradecimentos

À Universidade Federal de Minas Gerais (UFMG), ao Instituto de Ciências Biológicas e ao Programa de Pós-Graduação em Genética, pela oportunidade da realização do curso de Mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

À Associação brasileira de criadores de Guzerá e a Embrapa de Gado de Leite que tornaram possível a realização deste trabalho.

À professora Maria Raquel Carvalho, pela orientação, pela amizade e pelos conselhos fundamentais não somente para realização deste trabalho, como também para meu crescimento profissional e pessoal.

À professora Maria Gabriela Campolina Diniz Peixoto, pela co-orientação, pela amizade e toda a confiança depositada ao longo destes anos. Sem a sua presença, este trabalho não seria possível.

À Vânia Martins Penna, Maria de Fátima Ávila Pires e Ariane Figueirêdo Menicucci pelas conversas e imensa contribuição para entendimento da história da raça Guzerá no Brasil.

A Peter Laspina pelas horas de aula e correções do Inglês dos artigos, resumos e demais trabalhos desenvolvidos ao longo do Mestrado.

À banca por terem aceitado o convite de participarem de um momento tão importante da minha vida.

Aos meus amigos de graduação, em especial, João Locke, Pedro Lamounier, Raffello Di Ponzio, Marco Antônio e Newton Gontijo pelas inúmeras horas de conversa, estudo e diversão. Sem vocês, os quatro anos de graduação teriam sido muito mais difíceis.

Aos meus amigos de LGHM. Os atuais e todos que tive ao longo destes cinco anos laboratório. Contudo, agradeço especialmente, Izinara Rosse e Raphael Steinberg por terem sido os responsáveis pelo meu acolhimento e treinamento no laboratório; se obtive sucesso ao longo destes anos, devo muito a vocês. Além destes, agradeço profundamente, Gabriela Salazar, Aline Martins e Laura Moraes pelas horas e horas de conversa, pela companhia em congressos e pelo auxílio em todo meu processo de aprendizagem.

À minha família. Aos meus irmãos César e Izabela, pela amizade imensurável e inabalável, pelo apoio incondicional, pela motivação e por serem modelos de força de vontade, altruísmo e alegria. Ao meu primo Guilherme pela companhia, amizade e por ser mais um irmão em minha vida. Aos meus tios, principalmente, Marco Aurélio e Júnia; e Paulinho e Fatinha. Agradeço por todos os tipos de apoio possíveis, pelo carinho, confiança plena e conselhos. Ao meu amigo de infância, Guilherme Henrique, por todas

as alegrias e por ter participado de cada etapa da minha vida até o momento. Aos meus pais, Marília e Carlos, por em hipótese alguma terem medido esforços para me auxiliar em cada decisão tomada, em cada passo dado, em cada fracasso. Agradeço por terem me criado da melhor forma possível, incentivando meus estudos e em muitos momentos colocado o meu bem acima de outra coisa.

À minha amiga de graduação, amiga de LGHM, amiga de Mestrado, membro da minha família e namorada nas horas vagas, Fernanda Caroline. Todo seu amor e apoio incondicional tornaram cada objetivo que a priori seria impossível, um sucesso. O meu crescimento pessoal e profissional ao seu lado é imensurável. Agradeço aos meus sogros, Maikson e Soraia e ao meu cunhado, Filipe, por terem me recebido de braços abertos em sua família e todo apoio e confiança depositada nestes anos.

A todos que em algum momento contribuíram, mesmo que da menor maneira possível, me auxiliando, incentivando e inspirando para que eu alcançasse meus objetivos. Sou extremamente grato a cada uma destas pessoas.

Sumário

LISTA DE ABREVIATURAS.....	vii
Resumo	8
Abstract.....	9
Introdução Geral	10
Sobre a produção de leite no Brasil e as raças leiteiras	10
Diversidade genética da raça Guzerá.....	10
CAPITULO 1: Reducing cryptic relatedness in genomic datasets via a central node exclusion algorithm	14
Abstract.....	14
CAPITULO 2: Retelling the recent evolution of genetic diversity for Guzerá: inferences from LD decay, runs of homozygosity and N_e over the generations	15
Abstract.....	15
Discussão Geral	16
Conclusão Geral	19
Referências	20
Anexos.....	23
Produção bibliográfica.....	23
Artigos publicados em periódicos	23
Artigos aceitos para publicação.....	23
Artigos submetidos a periódicos	23
Manuscritos em desenvolvimento.....	23
Resumos publicados em anais de eventos.....	24
Resumos aceitos para publicação	24
Produção técnica	26
Softwares desenvolvidos	26

LISTA DE ABREVIATURAS

MOET	multiple ovulation and embryo transfer
F_{PED}	coeficiente de endogamia baseado em dados de pedigree
ROH	runs of homozygosity
F_{ROH}	coeficiente de endogamia baseado em runs of homozygosity
QTLs	quantitative trait locus
N_e	tamanho efetivo da população; effective population size
SNP	single nucleotide polymorphism
LD	linkage disequilibrium
IBD	identity by descendent
P_n	portion of polymorphic markers
NGS	next generation sequencing
FAO	food and agriculture organization of the United Nations
MAF	minor allele frequency
SVS7	SNP and variation Suite 7
E_{Ho}	expected homozygosity
O_{ho}	observed homozygosity
E_{He}	expected heterozygosity
O_{He}	observed heterozygosity
EM	expectation-maximization
HWE	Hardy-Weinberg Equilibrium
F_{HOM}	Inbreeding coefficient based on Runs of Homozygosity

Resumo

A indústria leiteira brasileira passou por um aumento significativo em sua produção na última década, atingindo uma produção anual de 32,30 bilhões de litros em 2012. Este aumento é explicado em parte pelo melhoramento genético dos rebanhos leiteiros submetidos a programas de seleção genética. Cerca de 80% de todo o rebanho brasileiro é constituído de animais zebuínos ou mestiços. A raça Guzera se destaca entre as outras raças zebuínas devido a algumas importantes características como resistência a parasitas, como carrapatos, habilidade de consumir forragem bruta e resistência à temperatura. Entretanto, as estimativas de diversidade genética da raça Guzera estão próximas do limiar em que é observada perda de variabilidade genética ocasionado pela deriva genética e depressão endogâmica. Até o momento, os estudos que visam estimar a variabilidade genética na raça Guzera foram baseados em dados de pedigree, microssatélites e poucos SNPs, não permitindo uma estimativa em escala genômica da diversidade genética da raça. O desenvolvimento de chips de alta-densidade para genotipagem de SNPs para bovinos viabilizou os estudos de diversidade ou de associação em escala genômica. Estes métodos, em geral, assumem que a população em estudo é uma amostra homogênea. No entanto, as populações naturais estão ligadas em redes, devido ao fluxo gênico. Isto pode resultar em falsas representações da estrutura populacional e/ou associações espúrias, caso isto não seja corretamente verificado antes das análises. Assim, estratégias para reduzir o nível de parentesco em amostras, são fundamentais para evitar associações espúrias e conclusões errôneas. O presente trabalho é composto de dois estudos. No primeiro, foi avaliado o desempenho da plataforma Illumina Bovine SNP50 para a raça Guzera e a aplicação de uma metodologia baseada análise de estatísticas de centralidade para a reamostragem de indivíduos baseado nos valores do coeficiente de Kinship. O segundo é composto pela análise de estrutura genética populacional da raça Guzera através de genotipagem em escala genômica. Para isso, foram avaliados indivíduos provenientes de rebanhos pertencentes ao Programa Nacional de Melhoramento da raça Guzera. A amostra utilizada no presente estudo representa a população mais selecionada e um dos maiores repositórios genéticos da raça Guzera no Brasil. Os resultados destes estudos indicaram uma ótima aplicabilidade do algoritmo de seleção de nós baseado em graus de rede através de estatísticas de centralidade, mostrando-se mais eficiente para a detecção de alguns parâmetros de estimativas de diversidade genética como ROH e um número maior de indivíduos na amostra final quando comparado com outra estratégia de reamostragem. Além disso, o estudo da estrutura genética populacional da raça Guzera indicou uma forte influência dos gargalos de garrafa e aumento da endogamia, sofridos pela raça durante o processo de estabelecimento da mesma no Brasil, na atual diversidade genética da raça. Estes resultados reforçam resultados já presentes na literatura que indicam a necessidade de um monitoramento da diversidade genética da raça através de critérios de diversidade populacional para garantir o sucesso dos programas de melhoramento.

Abstract

The Brazilian dairy industry had a significant increase in its production in the last decade, reaching an annual production of 32.30 billion liters in 2012. This increase is explained, in part, by the genetic improvement of the dairy herds subjected to genetic selection programs. About 80% of all Brazilian cattle heads are Zebu animals or their crossbreeds. The Guzerá breed stands out among the other zebuine breeds due to some important characteristics like resistance to parasites such as ticks, ability to consume gross forage and heat tolerance. However, the genetic estimative for Guzerá breed is close to the threshold in which the loss of genetic variability by genetic drift and inbreeding depression is observed. Until the present moment, the studies developed in the Guzerá breeds, which aim to estimate the genetic variability, are based on pedigree, microsatellites and on few SNPs data, which do not make possible a Genome-Wide estimative. The development of high-density genotyping chips for cattle permitted a more precise approach to estimate the genetic variability. These methods, in general, assume that the population studied is a homogeneous sample. However, natural populations are connected in networks due to gene flow. This can lead to spurious representations of population structure and/or associations, if they are not adjusted correctly before the analyses. Thus, strategies to reduce the relatedness level in samples are important tools to avoid spurious associations and erroneous conclusions. The present work is composed of two studies. In the first, the performance of the platform Illumina Bovine SNP50 for the Guzerá breed and the applicability of a methodology based on centrality statistics analysis to resampling individuals based on the Kinship coefficient was evaluated. The second study is the analysis in Genomic scale of the population genetic structure of Guzerá breed. For this, individuals from herds belonging to the National breeding programs of Guzerá were evaluated. The sample used in the present study represents the most selected and one of the largest genetic repositories of the Guzerá breed in Brazil. The results of these studies indicate a great applicability of the node selection algorithm based on a network's degree of centrality statistic, showing up more efficient for the detection of some estimative parameters of the genetic diversity such as ROH and resulting in a higher number of individuals in the final sample when compared to the other strategy of resampling. Furthermore, the population genetic diversity study of Guzerá breed indicates a strong influence of the bottlenecks and increase of the endogamy, suffered by the breed during the process of settlement in Brazil, in the current genetic diversity of the breed. These findings reinforce the results already presented in other studies, which indicate the need for monitoring the genetic diversity of the breed through parameters of population diversity, to ensure the success of the breeding programs.

Introdução Geral

Sobre a produção de leite no Brasil e as raças leiteiras

A indústria leiteira brasileira passou por um aumento expressivo de sua produção na última década. Em 2000, a produção leiteira era estimada em 19,76 bilhões de litros ao ano, alcançando uma produção anual de 32,30 bilhões de litros em 2012. Dentre os estados brasileiros, Minas Gerais é o maior produtor de leite, sendo responsável por 27,6% da produção nacional [1]. Contudo, a produtividade, ou seja, a produção por vaca/ano praticamente não modificou durante este período (ADD REFS).

O rebanho leiteiro brasileiro é constituído de raças taurinas (*Bos taurus*), raças zebuínas (*Bos indicus*) e de mestiços oriundos de cruzamentos entre essas raças, principalmente pelas raças Holandesa, Gir e Guzerá. Os animais mestiços representam cerca de 70% de todo o rebanho leiteiro nacional [2]. Dentre as raças zebuínas, a Guzerá destaca-se pelo rápido crescimento, caracterizando-se como raça de dupla aptidão e adequada para programas de cruzamento [3].

Na última década, em função do potencial produtivo da raça, deu-se início ao programa de melhoramento do Guzerá para leite baseado na utilização de duas estratégias: o teste de progênie e o esquema de seleção em núcleo MOET (*multiple ovulation and embryo transfer*) [4]. No teste de progênie, os touros em prova são avaliados principalmente pelo desempenho de suas filhas; e no núcleo MOET pelas irmãs completas. O esquema MOET apresenta vantagem de antecipação da avaliação genética dos animais devido a redução do intervalo de gerações.

Diversidade genética da raça Guzerá

O sucesso dos programas de seleção pode ser prejudicado pela baixa acurácia da seleção e pelo risco de aumento da endogamia, uma vez que um pequeno efetivo de animais é selecionado a partir dos índices de família, gerados pela matriz de parentesco, e submetido intensamente ao processo de melhoramento genético [5, 6].

Este pequeno efetivo de animais submetidos aos processos de seleção pode levar aos efeitos deletérios da depressão endogâmica. Esses efeitos, já foram avaliados tanto em rebanhos taurinos quanto em zebuínos, revelando que características como o desenvolvimento embrionário nas duas primeiras semanas pós-fertilização [7] e idade no primeiro parto, produção diária de leite e intervalo de partos têm seu desempenho reduzido [7].

Até o presente momento, estudos que buscaram avaliar a diversidade genética na raça Guzerá baseiam-se em dados genealógicos ou dados moleculares de microssatélites [8, 9]. No caso dos estudos baseados em microssatélites, o baixo número de indivíduos e *loci* avaliados não permitem uma estimativa, em escala genômica, da diversidade genética da raça. Os estudos baseados em dados genealógicos (F_{PED}), geralmente utilizam o conceito desenvolvido por Wright (1922) [10], porém, falham em detectar a influência de uma

população fundadora em relação à população base do estudo, por considerar uma baixíssima relação entre estas populações e se basear em cálculos de probabilidade de semelhança entre indivíduos. Além disso, estes estudos não levam em consideração a natureza estocástica da recombinação, ocasionando às vezes resultados de difícil compreensão [11]. No caso do uso de dados genealógica para a avaliação da diversidade genética de animais envolvidos em processos de seleção genética, existem dois pontos que devem ser ressaltados: 1) Estudos demonstram que erros nos pedigrees devido a má identificação, interpretação e registros são comuns [12, 13], podendo assim, ocasionar associações espúrias; 2) Estudos baseados em F_{PED} assumem que todo o genoma está sob seleção neutra, não abrindo espaço para os efeitos da seleção artificial, podendo ocasionar alguns vieses de avaliação [14].

Keller et al. (2011) [15] concluiu que coeficientes de endogamia derivados de Runs of Homozygosity (ROH) são ótimas medidas para determinar os trechos de autozigosidade ao longo do genoma, assim como os efeitos da endogamia. A definição de um segmento do DNA como sendo um ROH, utilizando-se de dados de genotipagem, pode ser determinada como longos trechos ininterruptos de marcadores homocigotos [16]. Além disso, longos trechos de autozigosidade são compartilhados entre indivíduos aparentados [17]. Espera-se que trechos maiores sejam compartilhados entre indivíduos que possuem ancestrais em comum mais próximos, da mesma forma que trechos compartilhados mais curtos são originados de ancestrais mais distantes. O tamanho esperado de um trecho em ROH segue uma distribuição exponencial com uma média estimada em $\frac{1}{2}g$ Morgans [18], onde g é o número de gerações desde o ancestral comum. Assim, o tamanho esperado para um trecho de ROH na prole derivada de um cruzamento entre irmãos completos ($g=2$, levando-se em consideração os irmãos completos, que se reproduziram, e os pais destes irmãos) é de $\frac{1}{4}$ Morgans, ou seja, 25cM.

Inúmeros trabalhos já foram conduzidos em humanos [11, 19, 20] e bovinos [21, 22] para estimar o coeficiente de endogamia baseado em ROH (F_{ROH}). Inclusive, Ferencakovic et al. (2012) [21], utilizando-se de dados de pedigree e genotípicos das raças Brown Swiss, Fleckvieh, Norwegian Red e Tyrol Grey (todas taurinas), demonstraram que os dados derivados de F_{ROH} apresentam uma correlação elevada com os dados derivados de F_{PED} e alta acurácia para estimar a probabilidade de depressão endogâmica. Desta forma, F_{ROH} pode ser considerada uma ótima alternativa para a estimativa do coeficiente de endogamia nos rebanhos bovinos, tendo em vista os problemas relacionados à estimativa F_{PED} já citados anteriormente. Por este motivo, esta abordagem será a utilizada no presente trabalho.

O padrão de distribuição do desequilíbrio de ligação ao longo do genoma da raça Guzerá ainda não foi avaliado. A ausência destes estudos dificulta o processo de seleção genômica baseada em QTLs (*quantitative trait locus*), método altamente utilizado nos processos de melhoramento genético em bovinos [23, 24]. Esta análise assume que o efeito de um determinado segmento cromossômico é homogêneo para toda população desde que o valor de desequilíbrio de ligação entre o marcador utilizado e o *locus* seja o mesmo em toda a população. Assim, é necessário o desenvolvimento de um complexo

mapa de desequilíbrio de ligação ao longo do genoma para garantir a associação entre o *QTL* e a característica de interesse [25]. Além disso, a comparação de mapas de desequilíbrio de ligação entre diferentes rebanhos, com diferentes atributos de produção, auxilia na identificação de regiões genômicas sujeitas a diferentes tipos de seleção [26].

Outras medidas de acesso à variabilidade genética de uma população são a presença de estruturação e o tamanho efetivo da população (N_e). O N_e e os dados de estruturação populacional, auxiliam nas análises da variabilidade genética observada em uma população sob um ponto de vista retrospectivo, além de possibilitar a predição da perda de variabilidade genética e da sobrevivência de pequenas populações [27]. Estas medidas já foram averiguadas para a raça Guzerá através de dados de pedigrees [9]. Contudo, estudos ainda não foram realizados para avaliar estas medidas utilizando-se de marcadores moleculares em escala genômica, o que nos propusemos a fazer.

Além disso, o grau de parentesco é um dos fatores que podem alterar a confiabilidade de estudos, que visam à estimativa da diversidade genética e aos estudos de associação. Uma série de estudos já demonstrou que SNPs utilizados para a seleção genômica podem, além de detectar os níveis de desequilíbrio de ligação entre SNPs e QTLs, podem identificar relações familiares entre indivíduos [28, 29]. Além disso, Wientjes et al. (2013) [30] demonstraram que a confiabilidade das predições genômicas é mais afetada pelo nível de parentesco genético presente na amostra do que pelos níveis de desequilíbrio de ligação. Desta forma, o desenvolvimento de estratégias para redução do nível de parentesco genético, em amostras que são oriundas de populações endogâmicas, se torna uma ferramenta fundamental para evitar associações espúrias e conclusões errôneas em estudos de associação e estimativas da variabilidade genética.

O desenvolvimento de *chips* de genotipagem em escala genômica para bovinos [31] permitiu uma abordagem mais precisa da variabilidade genética nas raças bovinas. Trabalhos avaliando o padrão de desequilíbrio de ligação, heterozigosidade média, estrutura de blocos haplotípicos e coeficientes de endogamia em rebanhos taurinos já foram desenvolvidos [21, 32-34]. Contudo, os trabalhos de McKay et al. (2007) [26] e Espigolan et al. (2013) [35], foram os únicos a utilizar *chips* de genotipagem em escala genômica em zebuínos, ambos determinando o mapa de desequilíbrio de ligação na raça Nelore. McKay et al. (2007) empregou 2670 marcadores derivados da montagem Btau3.1 e Espigolan et al. (2013), o High Density Bovine SNP BeadChip contendo 777.962 marcadores.

A ausência de trabalhos que utilizem plataformas mais abrangentes para avaliar a estrutura genética e a variabilidade populacional na raça Guzerá torna mais difícil o processo de seleção assistida por marcadores moleculares e a prevenção da endogamia, por exemplo. Além disso, as estimativas da real variabilidade genética são prejudicadas, o que pode ocasionar conclusões e aplicações equivocadas para a definição de estratégias de manutenção da variabilidade genética.

Desta forma, o presente trabalho visou desenvolver um estudo, em escala-genômica, de avaliação da estrutura e da diversidade genética da raça Guzerá. Para isso, foram

avaliados indivíduos oriundos de rebanhos pertencentes do Programa Nacional de Melhoramento do Guzerá, de forma que a amostra representasse população selecionada para leite, que constitui um dos repositórios genéticos da raça Guzerá no Brasil.

Neste contexto, o primeiro capítulo deste trabalho, descreve a comparação de duas estratégias de reamostragem, feitas com o intuito de se identificar a melhor estratégia para diminuição/exclusão do efeito da relação familiar em estudos de avaliação da diversidade genética e estudos de associação. A partir da definição da melhor estratégia, foi realizada a análise da estrutura e diversidade genética da raça Guzerá, que são descritas no segundo capítulo. Na sessão Anexos, são listados todos os produtos gerados durante os dois anos do mestrado, inclusive os oriundos de colaboração em projetos de pesquisa, que não pertencem ao escopo desta dissertação.

CAPITULO 1: Reducing cryptic relatedness in genomic datasets via a central node exclusion algorithm

Pablo A. S. Fonseca¹; Thiago P. Leal¹; Fernanda C. Santos¹; Mateus H. Gouveia¹; Samir Id-Lahoucine²; Izinara C. Rosse¹; Ricardo V. Ventura^{2,3}; Frank A. T. Bruneli⁴; Marco A. Machado⁴; Maria Gabriela C. D. Peixoto⁴; Eduardo Tarazona-Santos¹; Maria Raquel S. Carvalho^{1*}

¹ Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

² Center for Genetic Improvement of Livestock, University of Guelph, Guelph, ON N1G 2W1, Canada

³ Beef Improvement Opportunities, Guelph, ON N1K 1E5, Canada

⁴ Embrapa Dairy Cattle, Juiz de Fora, MG, Brazil

Abstract

Cryptic relatedness is a confounding factor in genetic diversity and genetic association studies. Development of strategies to reduce cryptic relatedness in a sample is a crucial step for downstream genetic analyzes. The present study uses a node selection algorithm, based on network degrees of centrality, to evaluate its applicability and impact on evaluation of genetic diversity and population stratification. 1,036 Guzerá (*Bos indicus*) females were genotyped using Illumina Bovine SNP50 v2 BeadChip. Four strategies were compared. The first and second strategies consists on a iterative exclusion of most related individuals based on PLINK kinship coefficient (ϕ_{ij}) and VanRaden's ϕ_{ij} , respectively. The third and fourth strategies were based on a node selection algorithm. The fourth strategy, Network G matrix, preserved the larger number of individuals with a better diversity and representation from the initial sample. Determining the most probable number of populations was directly affected by the kinship metric. Network G matrix was the better strategy for reducing relatedness due to producing a larger sample, with more distant individuals, a more similar distribution when compared with the full dataset in the MDS plots and keeping a better representation of the population structure. Resampling strategies using VanRaden's ϕ_{ij} as a relationship metric was better to infer the relationships among individuals. Moreover, the resampling strategies directly impact the genomic inflation values in Genome-wide association studies. The use of the node selection algorithm also implies better selection of the most central individuals to be removed, providing a more representative sample.

Keywords: Bovine, Inbreeding, Genetic diversity, Population genetic structure, Cryptic relatedness

Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.12746>

DOI: <https://doi.org/10.1111/1755-0998.12746>

CAPITULO 2: Retelling the recent evolution of genetic diversity for Guzerá: inferences from LD decay, runs of homozygosity and Ne over the generations

Pablo Augusto de Souza Fonseca¹; Fernanda Caroline dos Santos¹; Izinara Cruz Rosse¹; Ricardo Vieira Ventura^{2, 3}; Frank Ângelo Tomita Brunelli⁴; Vânia Maldini Penna⁵; Rui da Silva Verneque⁴; Marco Antônio Machado⁴; Marcos Vinícius Gualberto Barbosa da Silva⁴; Maria Raquel Santos Carvalho¹, and Maria Gabriela Campolina Diniz Peixoto⁴

¹ Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, 31.270-901, Brazil

² Center for Genetic Improvement of Livestock, University of Guelph, Guelph, ON N1G 2W1, Canada

³ Beef Improvement Opportunities, Guelph, ON N1K 1E5, Canada

⁴ Embrapa Gado de Leite, Juiz de Fora, 36038-330, Brazil

⁵ Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, 31.270-901, Brazil

Abstract

Genetic diversity is the one of the most important issues in conservation studies of livestock breeds or endangered species. In the present study, we tested the feasibility of describing the recent evolution in genetic diversity through genome-wide SNP genotyping and estimates of linkage disequilibrium decay patterns, effective population size, inbreeding coefficient based on runs of homozygosity and population structure. We choose the bovine indicine breed Guzerá because it has suffered recent bottlenecks which have been registered historically. A sample of 1,036 females was genotyped using Illumina BovineSNP50. A resampling strategy was applied to correct for sampling biases caused by the population structure in herds, and by the extensive use of some sires for artificial reproduction. A subsample of 210 animals and 32,806 markers with MAF>0.01 was used. Very low linkage disequilibrium was detected for distances greater than 120Kb between two markers. Furthermore, three points of decrease in effective population size between generations were detected, which coincide with the historically registered bottlenecks. The inbreeding coefficient, based on runs of homozygosity, confirmed a strong contribution of the last 20-30 generations to current inbreeding. In the population structure analysis, the most probable number of sub-populations is 2, reflecting selection purpose (beef or dual-purpose). Taken together, these results allow a retelling of the recent evolution of this breed. The strategy described here will be useful for other breeds or even species for which a careful historical registry is not available for conservation proposals.

Keywords: Genetic Diversity, Genome-wide Genotyping, Genetic structure, Bovine, Guzerá

Disponível em: <https://www.sciencedirect.com/science/article/pii/S1871141316302268>

DOI: <https://doi.org/10.1016/j.livsci.2016.10.006>

Discussão Geral

Os resultados obtidos no presente trabalho reforçam a necessidade do monitoramento da diversidade genética da raça Guzerá através de parâmetros populacionais. Além disso, os mesmos destacam a influência do nível de parentesco genético presente na amostra nos resultados obtidos, desta forma, indicando a necessidade da utilização de estratégias para reduzir ou eliminar estes efeitos em estudos de monitoramento da diversidade genética e em estudos de associação.

No presente trabalho, foi testada a aplicabilidade de um algoritmo de seleção de nós baseados em graus de rede através de estatísticas de centralidade utilizando como conector os valores de coeficiente de Kinship entre os indivíduos [36]. Esta metodologia mostrou-se extremamente útil, uma vez que os dados obtidos no presente trabalho indicam uma interferência significativa do grau de parentesco em estimativas de diversidade genética, corroborando os já existentes na literatura [37-39].

Além disso, neste trabalho foi avaliado o desempenho do chip Illumina bovine SNP50 para a raça Guzerá. Os resultados de avaliação desta plataforma indicaram que a raça Guzerá apresenta valores médios de MAF e proporção de loci polimórficos semelhantes às raças zebuínas utilizadas na construção do chip (Nelore e Gir). Contudo, um número elevado de marcadores fixados foi observado, aproximadamente 25% da plataforma. Isto pode ocasionar uma baixa densidade de marcadores em determinadas regiões do genoma e conseqüentemente ocasionar estimativas equivocadas. Apesar disto, esta plataforma apresentou resultados satisfatórios nas análises desenvolvidas no presente trabalho.

Até ao presente momento, não há estudos que avaliaram o padrão LD ao longo do genoma da raça Guzerá. No presente estudo, também foi observado um padrão inconsistente de decréscimo de LD até intervalos intermarcadores de 20Kb, o que pode ser explicado pela baixa densidade de pares de marcadores nestas distâncias (a distância média entre os marcadores após a filtragem de controle de qualidade foi 76.265Kb). Estes resultados aqui obtidos, foram muito semelhantes aos observados por O'Brien et al. (2014) [40]. Além disso, O'Brien et al. (2014) [40] também avaliaram o padrão de queda do LD para outras raças zebuínas (Nelore, Gir e Brahman), observando valores abaixo de 0,1 em distâncias maiores que 200Kb. No presente estudo, foi observado um padrão similar, onde valores de $r^2 < 0,1$ foram atingidos em distâncias entre 120 e 140Kb.

A análise de LD por cromossomo apresentou os cromossomos 6, 7 e 14 como os possuidores das maiores médias entre autossomas para r^2 e $|D'|$. Outros estudos também encontraram estes cromossomos como os possuidores das maiores médias de LD, tanto em raças zebuínas quanto em raças taurinas [35, 41]. Os maiores valores de LD observados nesses cromossomos podem refletir a presença de importantes *loci de características quantitativas* relacionados com importantes traços de produção, tais como, a produção de leite, a resistência a parasitas e características reprodutivas [42-45].

O desenvolvimento de mapas de equilíbrio de ligação podem aumentar o poder e a precisão de estudos de associação e ocasionar uma melhoria dos processos de seleção

genética. Apesar da baixa densidade de marcadores em algumas distâncias, o número total de marcadores obtidos no presente estudo, após a filtragem de controle de qualidade, está de acordo com o número de marcadores sugeridos por McKay et al. (2007) [26] para estudos de desenvolvimento de mapas de LD (cerca de 30 mil).

A estrutura genética da população e a distribuição das ROH ao longo do genoma foram comparadas entre a amostra inicial (constituída pelos 1.036 animais inicialmente coletados) e as duas subamostras obtidas após a filtragem de indivíduos pelos valores do coeficiente de Kinship (“IBD sample” e “Centrality sample”). Os resultados da análise de estrutura populacional, conduzida através dos valores de Cross-validation error no software Admixture 1.23 [46], determinaram $K=2$ como o valor mais provável para “IBD Animals subsample” e “Centrality Animals subsample”, enquanto para a amostra total o valor mais provável de K foi igual a 75. Estes resultados corroboraram os resultados obtidos por Kehdi et al (2014) [47], onde o software Admixture 1.23 foi sensível o suficiente para detectar relações de parentesco existentes na amostra.

Os resultados da análise de distribuição, número e comprimento total das ROH entre os indivíduos salientaram as vantagens da utilização do algoritmo de seleção de nós baseados em graus de rede através de estatísticas de centralidade. A amostra “IBD Animals subsample” apresentou cerca de metade do número de corridas e do comprimento total das ROH quando comparado com a amostra total (1.036 animais) e a subamostra “Centrality Animals subsample”. Este resultado pode ser explicado pelo fato de que a estratégia de re-amostragem baseada seleção de nós através de estatísticas de centralidade não exclui todos os membros de uma rede familiar pertencente à amostra, diferentemente do que ocorre na estratégia utilizada para a formação da amostra “IBD Animals subsample”. Como animais aparentados compartilham longos segmentos autozigotos [17], a exclusão de todos os membros de uma família reduzirá a média geral do número e comprimento das ROH na amostra. Além disso, o maior número de indivíduos presente na amostra “Centrality Animals subsample” em relação à amostra “IBD Animals subsample” (210 animais e 158 animais, respectivamente), também pode ser responsável pelo maior número de ROH observado na primeira. Além disso, a amostra “Centrality Animals subsample” apresentou estimativas como o padrão de desequilíbrio de ligação e proporção de *loci* polimórficos mais semelhantes à amostra total em relação às estimativas obtidas na amostra “IBD Animals subsample”. Desta forma, a amostra obtida através do algoritmo de seleção de nós baseados em graus de rede através de estatísticas de centralidade (“Centrality Animals subsample”), foi selecionada como a amostra alvo para as análises de diversidade genética e estrutura populacional da raça Guzerá que foram conduzidas no presente trabalho.

Na última década, um programa de Múltipla Ovulação e Transferência de Embriões (MOET) foi implementado, no Brasil, para a raça Guzerá. Neste programa são formadas grandes famílias de irmãs e meias-irmãs, assim, permitindo avaliar o potencial produtivo de um touro pelo rendimento de suas irmãs [4]. No entanto, o núcleo MOET exige uma avaliação constante da diversidade genética, para seja evitado o uso de apenas um pequeno número de animais, que são selecionados pelos índices de produção da família

e submetidos a intensivos processos de seleção. A preservação da diversidade genética é crucial, uma vez que o sucesso do melhoramento genético pode ser prejudicado pela baixa acurácia da seleção e pelo risco do aumento da endogamia resultante deste pequeno número efetivo de animais utilizados na reprodução. Os efeitos deletérios da depressão endogamia já foram descritos, tanto em rebanhos taurinos quanto em zebuínos, afetando características como o desenvolvimento embrionário nas duas primeiras semanas pós-fertilização [7] e ocasionando baixa performance na produção diária de leite, idade no primeiro parto e intervalo entre partos [48].

Os resultados obtidos no presente trabalho corroboram outros estudos que previamente indicaram preocupantes estimativas de parâmetros genéticos, como baixo N_e e elevados coeficientes de inbreeding, para a raça Guzerá [8, 9, 48]. Além disso, os resultados aqui obtidos proporcionam uma análise retrospectiva dos eventos de redução populacional e aumento de endogamia da raça Guzerá. Estes resultados permitiram observar, possivelmente, a forte interferência de três períodos durante o processo de estabelecimento da raça no país. O primeiro deles constitui a entrada da raça no país no final do século 19, período em que um pequeno efetivo populacional foi importado da Índia e utilizado para a formação de rebanhos. Esta hipótese é reforçada pela redução do N_e observada entre 24 e 22 gerações passadas e pela estabilidade dos valores de contribuição de cada geração ao F_{ROH} até 20 gerações passadas. Estas gerações correspondem a aproximadamente 120 anos atrás, retomando o período de importação e formação do primeiro rebanho Guzerá no país. Além disso, os outros dois eventos que contribuíram drasticamente para a constituição populacional atual da raça também são indicados pelas reduções no N_e . Estes períodos correspondem a 12 e 8 gerações atrás, o que corresponde às décadas de 1940 e 1970, respectivamente. Estes dois momentos são de grande importância no estabelecimento da população da raça Guzerá. Por volta de 1940, existiu uma forte tendência do uso das fêmeas da raça Guzerá para a formação da raça mestiça Indulbrasil, o que contribuiu para uma forte redução do seu efetivo populacional. Na década de 70 ocorreu a formação dos “herd books” e os registros em “livro fechado”, exigindo assim provas de paternidade do animal a ser registrado como prole do cruzamento entre animais da raça Guzerá (neste período, o exame de DNA não era disponível, assim as provas constituíam-se de comunicação de cruzamento e nascimento, ou tipagens de grupos sanguíneos, dentre outras). Além disso, o exame morfológico e o registro junto à Associação Brasileira de Criadores de Zebu era necessário para que um animal fosse reconhecido como constituinte da raça. Todo este processo reduziu o número de animais mestiços, que anteriormente eram considerados como constituintes puros da raça, reduzindo assim o efetivo de rebanhos puros que poderiam ser utilizados como reservatórios genéticos.

Os tamanho efetivo populacional e F_{ROH} obtidos para as gerações mais recentes (10-2 gerações atrás) foram muito semelhantes aos obtidos por Peixoto et al. (2010)[9]. Esta semelhança de resultados salientam duas importantes observações. A primeira, sobre a metodologia aqui utilizada, corroborando elevadas correlações entre as estimativas genéticas derivadas de metodologias utilizadas neste trabalho e estimativas obtidas

através de dados de pedigree já descritas na literatura [14, 21]. A segunda, indica a necessidade da utilização de estratégias de manutenção da diversidade genética da raça Guzerá através de critérios de monitoramento da estrutura genética populacional. Isto deve ocorrer devido à proximidade das estimativas de diversidade genética obtidas para a raça estarem próximas do limiar onde a perda de diversidade genética ocasionada pela deriva genética e depressão endogâmica é observada [49, 50].

A existência de dois principais componentes populacionais presentes na amostra utilizada no estudo, evidenciada pela análise realizada no software Admixture 1.23, pode ser de grande valia para o desenvolvimento das estratégias de manutenção da diversidade genética da raça. A utilização de diferentes reservatórios genéticos para os processos de cruzamento e formação de rebanhos é uma importante etapa para evitar a ocorrência de processos endogâmicos e posterior redução da diversidade genética populacional.

Conclusão Geral

O presente trabalho foi o primeiro a avaliar o desempenho do chip Illumina Bovine SNP50 para a raça Guzerá, indicando, uma elevada proporção de marcadores fixados na plataforma, assim como observado para outras raças zebuínas. Estes resultados reforçam a necessidade da construção de um chip de genotipagem customizado para raças zebuínas. Além disso, também foi o primeiro a obter estimativas de diversidade genética utilizando abordagens em escala genômica para a raça Guzerá. Os resultados aqui obtidos reforçam a necessidade da utilização de estratégias de aumento da diversidade genética para a raça, devido aos índices relativamente baixos de diversidade genética observados neste e em trabalhos anteriores. Contudo, o presente trabalho destaca-se por ter conseguido inferir a contribuição de gerações passadas ao atual estado de diversidade genética da raça, contribuindo assim, para uma melhor escolha dos passos que devem ser dados para o desenvolvimento das estratégias de manutenção da diversidade genética. Além disso, a aplicação de um algoritmo de seleção de nós baseados em graus de rede através de estatísticas de centralidade utilizando como conector o valor do coeficiente de Kinship, permitiu uma avaliação livre de interferência de estrutura familiar. Desta forma, possibilitou resultados que refletem de forma mais fidedigna, a diversidade genética presente na raça. Adicionalmente, a amostra utilizada no presente trabalho foi constituída de animais oriundos das principais fazendas pertencentes ao Programa Nacional de Melhoramento do Guzerá para leite. Assim, os resultados aqui obtidos refletem as estimativas genéticas do principal reservatório genético da raça no país e consequentemente, da população mais selecionada para a produção leiteira.

Referências

1. Instituto brasileiro de geografia e estatística – IBGE (2012). Banco de dados. Disponível em: ftp://ftp.ibge.gov.br/Producao_Pecuaria/Producao_da_Pecuaria_Municipal/2012/tabelas_pdf/tab06.pdf. Acessado em: 20, Nov. 2012.
2. Alvim MJ, Paciullo DSC, Carvalho MM, Aroeira LJMA, Novaes LP, Gomes AT, Miranda JEC: **Sistema de produção de leite com recria de novilhas em sistemas silvipatoris. Embrapa Gado de Leite – Sistema de Produção nº 7. Versão eletrônica 2005.** Disponível em: <http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Leite/LeiteRecriadeNovilhas/racas.htm>. Acessado em: 20, Agost. 2014.
3. Winkler R, Penna, V.M.: **O Guzerá.** *Inf Agropec* 1992, **Inf. Agropec.**:16:10-14.
4. Teodoro RL, Verneque, R.S., Martinez, M.L., Penna, V.M.: **Programa Nacional de Melhoramento do Guzerá para Leite: resultados do Teste de Progênie, do Arquivo Nacional e do Núcleo Moet.** . *Juiz de Fora : Embrapa Gado de Leite* 2003:20.
5. Nicholas FW, Smith C: **Increased rates of genetic change in dairy cattle by embryo transfer and splitting.** *Animal Production* 2010, **36(03):341-353.**
6. Penna VM, Madalena FE, Alvim MTT: **Open MOET Nucleus of Selection in Guzerá.** Proceedings of the 6th World Congress on Genetics Applied to Livestock Production 1998. Armidale, Australia. CD-ROM.
7. Lazzari G, Colleoni S, Duchi R, Galli A, Houghton FD, Galli C: **Embryonic genotype and inbreeding affect preimplantation development in cattle.** *Reproduction* 2011, **141(5):625-632.**
8. Silva Filho E, dos Santos YM, Margarido YMM, Mendes TJ, Pinheiro LMDL, de Souza PH, Campelo JEG: **Genetic Variability of Guzerat and Senepol Bovine Breeds by Microsatellite Markers.** *Journal of Biotechnology and Biodiversity* 2013, **4(3).**
9. Peixoto MGCD, Poggian CF, Verneque RS, Egito AA, Carvalho MRS, Penna VM, Bergmann JAG, Viccini LF, Machado MA: **Genetic basis and inbreeding in the Brazilian Guzerat (Bos indicus) subpopulation selected for milk production.** *Livestock Science* 2010, **131(2-3):168-174.**
10. Wright, S: **Coefficients of inbreeding and relationship.** *American Naturalist* 1922, **330-338.**
11. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A *et al*: **Runs of homozygosity in European populations.** *American journal of human genetics* 2008, **83(3):359-372.**
12. Carneiro PLS, Euclides RF, Silva MDA, Lopes PS, Torres RDA, Carneiro APS, Torres Filho RDA: **Effect of pedigree errors on the selection.** *Revista Brasileira de Zootecnia* 1999, **28(2), 269-274.**
13. Ron M, Blanc Y, BandM, Ezra E, Weller JI: **Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement.** *Journal of Dairy Science* 1996, **79(4), 676-681.**
14. Curik I, Ferenčaković M, Sölkner J: **Inbreeding and runs of homozygosity: A possible solution to an old problem.** *Livestock Science* 2014, **166:26-34.**
15. Keller MC, Visscher PM, Goddard ME: **Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data.** *Genetics* 2011, **189(1):237-249.**
16. Gibson J, Morton NE, Collins A: **Extended tracts of homozygosity in outbred human populations.** *Human molecular genetics* 2006, **15(5):789-795.**
17. Broman KW, Weber JL: **Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain.** *The American Journal of Human Genetics* 1999, v. 65, n. 6, p. 1493-1500.

18. Howrigan DP, Simonson MA, Keller MC: **Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms.** *BMC genomics* 2011, **12**:460.
19. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF: **Genomic runs of homozygosity record population history and consanguinity.** *PloS one* 2010, **5**(11):e13996.
20. Nothnagel M, Lu TT, Kayser M, Krawczak M: **Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans.** *Human molecular genetics* 2010, **19**(15):2927-2935.
21. Ferencakovic M, Hamzic E, Gredler B, Solberg TR, Klemetsdal G, Curik I, Solkner J: **Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations.** *Journal of animal breeding and genetics = Zeitschrift fur Tierzucht und Zuchtungsbiologie* 2013, **130**(4):286-293.
22. Sölkner J, Filipcic L, Hampshire N: **Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees.** *Animal Science* 2010, **67**(02):249-256.
23. Hayes BJ, Bowman, P. J., Chamberlain, A. J., Goddard, M. E.: **Invited review: Genomic selection in dairy cattle: Progress and challenges.** *Journal of dairy science* 2009, **92**(2):433-443.
24. Chen L, Wu, X., Jiang, H., Huang, Y., Chen, M., Yi, J., & Fang, J. : **Review on QTL for milk production traits and application of non-antibiotic feed additives in dairy cattles.** *Journal of Food, Agriculture & Environment* 2013, **11**(1):511-515.
25. Bohmanova J, Sargolzaei, M., & Schenkel, F. : **Characteristics of linkage disequilibrium in North American Holsteins.** *BMC genomics* 2010, **11**(1): 421.
26. McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C *et al*: **Whole genome linkage disequilibrium maps in cattle.** *BMC genetics* 2007, **8**:74.
27. Wang J: **Estimation of effective population sizes from data on genetic markers.** *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2005, **360**(1459):1395-1409.
28. Habier D, Fernando RL, Dekkers JC: **The impact of genetic relationship information on genome-assisted breeding values.** *Genetics* 2007, **177**(4):2389-2397.
29. Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G: **The impact of genetic relationship information on genomic breeding values in German Holstein cattle.** *Genetics, selection, evolution : GSE* 2010, **42**:5.
30. Wientjes YC, Veerkamp RF, Calus MP: **The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction.** *Genetics* 2013, **193**(2):621-631.
31. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS *et al*: **Development and characterization of a high density SNP genotyping assay for cattle.** *PloS one* 2009, **4**(4):e5350.
32. Flury C, Tapio M, Sonstegard T, Drogemuller C, Leeb T, Simianer H, Hanotte O, Rieder S: **Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium.** *Journal of animal breeding and genetics = Zeitschrift fur Tierzucht und Zuchtungsbiologie* 2010, **127**(5):339-347.
33. Qanbari S, Pimentel EC, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H: **The pattern of linkage disequilibrium in German Holstein cattle.** *Animal genetics* 2010, **41**(4):346-356.
34. Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ: **High-resolution haplotype block structure in the cattle genome.** *BMC genetics* 2009, **10**:19.
35. Espigolan R, Baldi F, Boligon AA, Souza FR, Gordo DG, Tonussi RL, Cardoso DF, Oliveira HN, Tonhati H, Sargolzaei M *et al*: **Study of whole genome linkage disequilibrium in Nellore cattle.** *BMC genomics* 2013, **14**:305.

36. Gouveia MH, Leal TP: **EPIGEN initiative**. Available at: <http://epigen.grude.ufmg.br/>
37. Astle W, Balding DJ: **Population Structure and Cryptic Relatedness in Genetic Association Studies**. *Statistical Science* 2009, **24**(4):451-471.
38. Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA: **The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes**. *Genetics* 2010, **186**(3):983-995.
39. Morrison J: **Characterization and correction of error in genome-wide IBD estimation for samples with population structure**. *Genetic epidemiology* 2013, **37**(6):635-641.
40. Pérez O'Brien AM, Mészáros G, Utsunomiya YT, Sonstegard TS, Garcia JF, Van Tassell CP, Carvalheiro R, da Silva MVB, Sölkner J: **Linkage disequilibrium levels in *Bos indicus* and *Bos taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions**. *Livestock Science* 2014.
41. Lu D, Sargolzaei M, Kelly M, Li C, Vander Voort G, Wang Z, Plastow G, Moore S, Miller SP: **Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle**. *Frontiers in genetics* 2012, **3**:152.
42. Gasparin G, Miyata M, Coutinho LL, Martinez ML, Teodoro RL, Furlong J, Machado MA, Silva MV, Sonstegard TS, Regitano LC: **Mapping of quantitative trait loci controlling tick [*Rhipicephalus (Boophilus) microplus*] resistance on bovine chromosomes 5, 7 and 14**. *Animal genetics* 2007, **38**(5):453-459.
43. Gonda MG, Arias JA, Shook GE, Kirkpatrick BW: **Identification of an ovulation rate QTL in cattle on BTA14 using selective DNA pooling and interval mapping**. *Animal genetics* 2004, **35**(4):298-304.
44. Marques E, Grant JR, Wang Z, Kolbehdari D, Stothard P, Plastow G, Moore SS: **Identification of candidate markers on bovine chromosome 14 (BTA14) under milk production trait quantitative trait loci in Holstein**. *Journal of animal breeding and genetics = Zeitschrift für Tierzucht und Zuchtungsbiologie* 2011, **128**(4):305-313.
45. Weikard R, Widmann P, Buitkamp J, Emmerling R, Kuehn C: **Revisiting the quantitative trait loci for milk production traits on BTA6**. *Animal genetics* 2012, **43**(3):318-323.
46. Alexander DH, Novembre J, Lange K: **Fast model-based estimation of ancestry in unrelated individuals**. *Genome research* 2009, **19**(9):1655-1664.
47. Kehdy FSG, Magalhães WCS, Gouveia MH, Pereira A, Rodrigues A, Eduardo Tarazona-S. **Origin and dynamics of admixture in Brazilians: a population-based fine-scale approach**. 2014. Manuscript in submission process.
48. Panetto JC, Gutierrez JP, Ferraz JB, Cunha DG, Golden BL: **Assessment of inbreeding depression in a Guzerat dairy herd: effects of individual increase in inbreeding coefficients on production and reproduction**. *Journal of dairy science* 2010, **93**(10):4902-4912.
49. FAO: **The Management of Global Animal Genetic Resources**. FAO Anim. Prod. Health Paper. United Nations, Rome. 1992.
50. Frankham R, Ballou, J.D., Briscoe, D.A., : **Introduction to Conservation Genetics**. Cambridge University Press, Cambridge, UK.; 2002.

Anexos

Produção bibliográfica

Artigos publicados em periódicos

1. FONSECA, P.A.S.; ROSSE, I.C.; DEMIRANDA, M.; MACHADO, M.A.; VERNEQUE, R.S.; PEIXOTO, M.G.C.D.; CARVALHO, M.R.S. **A new tetra-primer ARMS-PCR for genotyping bovine kappa-casein polymorphisms**. Genetics and Molecular Research. , v.12, p.6521 - 6526, 2013.

Artigos aceitos para publicação

1. SANTOS, F. C.; FONSECA, P.A.S.; MORO, L.; LIMA-TAVARES, M. **Temas Transversais - Enfoque na abordagem e desenvolvimento de temas com ênfase em drogas em um colégio particular de Belo Horizonte**. Trabalhos completos aceitos publicação em eventos - V encontro de ensino em biologia. 2014.

Artigos submetidos a periódicos

1. ROSSE, I.C; FONSECA, P.A.S.; SANTOS, F. C.; LAMOUNIER, P.F; STEINBERG, R.S; MIRANDA, M.; Pires, M.F.A; PEIXOTO, M. G. C. D.; CARVALHO, M. R. S. **Functional analysis of SNPs in the oxytocin gene: an in silico approach**. Molecular Biology Reports. 2014.

2. PEIXOTO, M. G. C. D.; PANETTO, J. C. C.; EGITO, A. A.; BRUNELI, F. A. T.; SANTOS, G. G.; MACHADO, M. A.; STEINBERG, R. S.; FONSECA, P. A. S.; SANTOS, F. C; ROSSE, I. C; FARIA, F. J. C.; CARVALHO, M. R. S.. **Genetic diversity and structure of a Brazilian Guzerá (*Bos indicus*) metapopulation assessed by microsatellite markers**. Journal of Heredity. 2014

Manuscritos em desenvolvimento

1. MOURA, G.S; FONSECA, P.A.S; SANTOS, F.C; SANTOS, D.J; ROSSE, I.C; OLIVEIRA,G.C; ANDRADE, V. J, VALE-FILHO, V. R; SILVA, M. V. G. B; CARVALHO, M.R.S. **An in family case-control GWAS identifies genes/candidate regions for gonadal hypoplasia and sperm abnormalities on X-chromosome in Dairy-Gir zebu breed**.

2. Fonseca, P.A.S; Santos, F.C; Gouveia, M.H; Leal, T.P; Rosse, I.C; Peixoto, M.G.C.D; Tarazona-Santos, E; Carvalho, M.R.S. **Applicability of a node selection algorithm based on a network's degree of centrality to resample individuals through Kinship coefficient in the Guzerá breed**.

3. FONSECA, P.A.S; SANTOS, F.C; ROSSE, I.C; PENNA, V.M; PIRES, M.F.A; PEIXOTO, M.G.C.D AND CARVALHO, M.R.S. **Characterization of the genetic structure of Guzerá (*Bos indicus*) using genome-scale genotyping**.

Resumos publicados em anais de eventos

1. MIRANDA, M., GRILLO, L. T., TEIXEIRA, T. B., PRADO, A. C. A., PAIVA, A. E., FONSECA, P. A. S., LEAO, L. L., AGUIAR, M. J. B., CARVALHO, M. R. S., HAASE, V. G. **Duplicação do cromossomo 7p21.3-p14.3: Relato de caso e revisão de literatura.** In: XXVI Congresso brasileiro de genética médica, 2014, Fortaleza-CE. XXVI Congresso brasileiro de genética médica. , 2014.

2. ALMEIDA, M. P., FONSECA, P. A. S., SANTOS, F. C., SALAZAR, G. C., MARTINS, A. A. S., MIRANDA, M., CARVALHO, M. R. S. **Fluxograma para investigação da repercussão funcional de CNVs por data mining.** In: XXVI Congresso brasileiro de genética médica, 2014, Fortaleza-CE. XXVI Congresso brasileiro de genética médica. , 2014.

3. ROSSE, I.C, FONSECA, P. A. S., SANTOS, F. C., PEIXOTO, M. G. C. D., Pires, M.F.A, MARTINS, A. A. S., CARVALHO, M. R. S. **Fluxograma para detecção de variantes potencialmente funcionais.** In: XXV Congresso brasileiro de Genética Médica, 2013, Florianópolis-SC. XXV Congresso brasileiro de Genética Médica. , 2013.

4. ROSSE, I.C, ASSIS, J. G., FONSECA, P. A. S., SANTOS, F. C., LAMOUNIER, P.F, STEINBERG, R.S, MIRANDA, M., OLIVEIRA, G. C., Pires, M.F.A, PEIXOTO, M. G. C. D., CARVALHO, M. R. S. **Functional analysis in intronics SNPs.** In: X-meeting/BSB 2013 Conference,, 2013, Recife, PE - Brasil. X-meeting/BSB 2013. , 2013.

5. LAMOUNIER, P.F, ROSSE, I.C, FONSECA, P. A. S., SANTOS, F. C., STEINBERG, R.S, PEIXOTO, M. G. C. D., Pires, M.F.A, FONSECA, C. G., Verneque, RS, MACHADO, M. A., PENNA, V. M., MIRANDA, M., CARVALHO, M. R. S. **Investigação de SNPs descobertos no gene da oxitocina na raça Guzerá: frequências haplotípicas e associação ao fenótipo habilidade materna.** In: XXII SEMANA DE INICIAÇÃO CIENTÍFICA da UFMG, 2013, Belo Horizonte - MG. XXII SEMANA DE INICIAÇÃO CIENTÍFICA da UFMG. , 2013.

Resumos aceitos para publicação

1. FONSECA, P.A.S.; ALMEIDA, M. P.; MOURA, G. S.; SANTOS, F. C.; SANTOS, D. J. A.; OLIVEIRA, G. C.; VALE FILHO, V. R.; SILVA, M. V. G. B.; CARVALHO, M. R. S. **Bovine animal model for spermatid and scrotal alterations: additional clues for an X-chromosome component.** Resumo aceito - 64th Annual Meeting of the American Society of Human Genetics. 2014.

2. ZEBRAL, N. P; FONSECA, P.A.S; SANTOS, F.C; FREITAS, D.R; CERQUEIRA, M.M.O.P; CARVALHO, M.R.S. **The frequency of a polymorphism in the s1-casein gene for 1/2 and 3/4 blood Girolando breed.** Resumo aceito – V Encontro Genético de Minas Gerais. 2014.

3. FONSECA, P.A.S; MOURA, G.S; SANTOS, F.C; SANTOS, D.J; ROSSE, I.C; OLIVEIRA, G.C; ANDRADE, V. J., VALE-FILHO, V. R; SILVA, M. V. G. B; CARVALHO, M.R.S. **Identification of candidate genes and chromosomal regions for spermatic and scrotal alterations in bovines: additional clues for an x-chromosome component.** Resumo aceito - V Encontro Genético de Minas Gerais. 2014.
4. GRILLO, LT; FONSECA, P.A.S; FREITAS, D.R; SANTOS, F.C; ROSSE, I.C; PAIVA, A.E; CERQUEIRA, M.M.O.P; CARVALHO, M.R.S. **Frequency of k232a polymorphism in dcat1 gene for and blood Girolando breed.** Resumo aceito - V Encontro Genético de Minas Gerais. 2014.
5. PAIVA, A.E; FONSECA, P.A.S; SANTOS, F.C; ROSSE, I.C; FREITAS, D.R; CERQUEIRA, M.M.O.P; CARVALHO, M.R.S. **The frequency of polymorphism in the kappa-casein gene in and blood Girolando breed.** Resumo aceito - V Encontro Genético de Minas Gerais. 2014.
6. MARTINS, A.A.S; SALAZAR, G.C; COELHO, F.F; FONSECA, P.A.S; JULIO-COSTA, A; HAASE, V.G; CARVALHO, M.R.S. **Allelic frequency in CGG repeat of the FMR1 gene in individuals with typical development and with math learning disability.** Resumo aceito - V Encontro Genético de Minas Gerais. 2014.
7. ALMEIDA, MP; MIRANDA, M; TEIXEIRA, TB; PRADO, ACA; FONSECA, PA; LEÃO, LL; AGUIAR, MJB; HAASE, VG; CARVALHO, MRS. **Tandem duplication of the short arm of chromosome 7: metastasis risk, cancer susceptibilities and ethical implications.** Resumo aceito - V Encontro Genético de Minas Gerais. 2014.
8. SANTOS, F.C; PEIXOTO, M. G. C. D; FONSECA, P.A.S; PIRES, M. F. Á.; ROSSE, I.C; BRUNELI, F. A. T.; SANTOS, G.G.; CARVALHO, M.R.S. **Two new candidate regions for bovine behavior identified through a genome-wide association study in the Guzerá breed in Brazil.** Resumo aceito - V Encontro Genético de Minas Gerais. 2014.
9. PAIVA, A.E; FONSECA, P.A.S; 2; SANTOS, F.C; 2; MOURA, G.S; SANTOS, D.J; ROSSE, I.C; OLIVEIRA, G.C; ANDRADE, V. J., VALE-FILHO, V. R; SILVA, M. V. G. B; CARVALHO, M.R.S. **Literature data-mining for selecting potential functional candidate genes for gonadal hypoplasia and sperm defects in X-chromosome regions associated with these phenotypes by GWAS.** Resumo aceito - ISCB-Latin America x-Meeting. 2014.
10. FONSECA, P.A.S; SANTOS, F.C; GOUVEIA, M.H; LEAL, T.P; ROSSE, I.C; PEIXOTO, M.G.C.D; TARAZONA-SANTOS, E; CARVALHO, M.R.S. **Resampling individuals through Kinship coefficient by using a node selection algorithm in Guzerá breed.** Resumo aceito - ISCB-Latin America x-Meeting. 2014.

11. SANTOS, F.C; FONSECA, P.A.S; PIRES, M. F. Á.; ROSSE, I.C; BRUNELLI, F. A. T.; SANTOS, G.G.; CARVALHO, M.R.S; PEIXOTO, M. G. C. D. **Three SNPs identified by GWAS contribute significantly to reactivity in the Guzerá breed.**

Resumo aceito - ISCB-Latin America x-Meeting. 2014.

12. PAIVA, A.E; FONSECA, P.A.S; SANTOS, F.C; MOURA, G.S; SANTOS, D.J; ROSSE, I.C; OLIVEIRA, G.C; ANDRADE, V. J., VALE-FILHO, V. R; SILVA, M. V. G. B; CARVALHO, M.R.S.**Data-mining de literatura para a seleção de potenciais genes candidatos funcionais para defeitos de hipoplasia e espermatozoides gonadal em regiões do cromossomo X associados a esses fenótipos por GWAS.**

Resumo aceito - XXIII Semana de Iniciação Científica (UFMG). 2014.

13. ZEBRAL, N. P; FONSECA, P.A.S; SANTOS, F.C; FREITAS, D.R; CERQUEIRA, M.M.O.P; CARVALHO, M.R.S. **Frequência alélica e genotípica de um polimorfismo no gene da alfa-S1-caseína em animais da raça Girolando 1/2 H:Gir e 3/4 H:Gir.** Resumo aceito - XXIII Semana de Iniciação Científica (UFMG). 2014.

Produção técnica

1. SANTOS, F. C., ROSSE, I.C, PEIXOTO, M. G. C. D., CARVALHO, M. R. S., BRUNELLI, F. A. T., SANTOS, G. G., FONSECA, P.A.S. **Resultados dos projetos de pesquisa. Programa Nacional de Melhoramento do Guzerá para Leite -** Documentos 168. , 2014.

Softwares desenvolvidos

1. Durante o desenvolvimento deste projeto, foi desenvolvido, na linguagem C++, um módulo de detecção de ROH em associação com o professor Ricardo Ventura da University of Guelph (Canadá). O módulo faz parte de um software que contém uma série de funções para auxiliar no processo de seleção genômica. Este módulo permite ao usuário detectar ROH ao longo do genoma através dos dados de genotipagem utilizando a critérios pré-definidos. Estes critérios baseiam-se no tamanho da janela de leitura utilizada pelo software; o número mínimo de marcadores presente em cada corrida; a distância mínima entre SNPs consecutivos; número máximo de marcadores heterozigotos; e número máximo de genótipos perdidos. O usuário tem a possibilidade de alterar cada um destes critérios e moldá-los em relação aos seus objetivos. Como output o módulo disponibiliza ao usuário uma tabela por cromossomo contendo todas as ROH por indivíduo e os dados de início e fim de cada trecho. Este output pode ser usado para diversas análises baseadas em dados de ROH, como por exemplo, a identificação de *hotspots* e *coldspots* de homozigosidade ao longo do genoma.