UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ZOOTECNIA E ENGENHARIA DE ALIMENTOS

MATEUS CASTELANI FREUA

Use of genetic variance in dynamic mechanistic models of growth to predict cattle performance and carcass composition under feedlot conditions

Uso da variância genética em modelos mecanicistas dinâmicos de crescimento para predizer o desempenho e a composição da carcaça de bovinos confinados

Pirassununga
2015

MATEUS CASTELANI FREUA

Use of genetic variance in dynamic mechanistic models of growth to predict cattle performance and carcass composition under feedlot conditions

Uso da variância genética em modelos mecanicistas dinâmicos de crescimento para predizer o desempenho e a composição da carcaça de bovinos confinados

A dissertation submitted to the Faculdade de Zootecnia e Engenharia de Alimentos at Universidade de São Paulo in partial satisfaction of the requirements for the degree of Master of Science.

Concentration: Animal Bioscience

Chair: Prof. Dr. José Bento Sterman Ferraz

Pirassununga

2015

Dados Internacionais de Catalogação na Publicação

Serviço de Biblioteca e Informação da Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo

F889u

Freua, Mateus Castelani
Uso da variância genética em modelos mecanicistas dinâmicos de crescimento para predizer o desempenho e a composição da carcaça de bovinos confinados / Mateus Castelani Freua. -- Pirassununga, 2015.
87 f.
Dissertação (Mestrado) -- Faculdade de Zootecnia e Engenharia de Alimentos – Universidade de São Paulo.
Departamento de Medicina Veterinária.
Área de Concentração: Biociência Animal.

Orientador: Prof. Dr. José Bento Sterman Ferraz.

1. Crescimento de bovinos de corte 2. Fisiologia genômica computacional 3. Mapeamento genótipo-fenótipo 4. Predição genômica 5. Modelagem mecanicista. I. Título.

Use of genetic variance in dynamic mechanistic models of growth to predict cattle performance and carcass composition under feedlot conditions

Mateus Castelani Freua

A dissertation submitted in partial satisfaction of the requirements for the degree of Master of Science

in

Animal Bioscience

in the

Faculdade de Zootecnia e Engenharia de Alimentos
Universidade de São Paulo

Committee in charge:

Dr. _____; Affiliation: _____
Decision:_____; Signature:_____

Dr. _____; Affiliation: _____
Decision:_____; Signature:_____

Dr. _____; Affiliation: _____
Decision:_____; Signature:_____

Approved in _____, Pirassununga, 2015.

Para

Minha Avó, filha de Oxum,

Minha Mãe, filha de Iansã.

À esplêndida genealogia que me gerou.

Acknowledgments

Bárbara Laís Unglaube Schmidt, Elisângela Chicaroni de Mattos Oliveira, Dr. José Bento Sterman Ferraz, Dr. Miguel Henrique de Almeida Santana, Dr. Ricardo Vieira Ventura, Thaís Brochado, and the São Paulo Research Foundation (2013/26902-0).

Resumo

FREUA, M. C. **Uso da variância genética em modelos mecanicistas dinâmicos de crescimento para predizer o desempenho e a composição da carcaça de bovinos confinados**. 2015. 87 f. Dissertação (Mestrado) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2015.

A predição da variância fenotípica é de grande importância para que os sistemas de produção de bovinos de corte consigam aumentar a rentabilidade otimizando o uso de recursos. Modelos mecanicistas dinâmicos do crescimento bovino vêm sendo utilizados como ferramentas de suporte à tomada de decisão em sistemas de manejo individual do gado. Entretanto, a aplicação desses modelos ainda fundamenta-se em parâmetros populacionais, sem qualquer abordagem para que se consiga capturar a variabilidade entre sujeitos nas simulações. Assumindo que modelos mecanicistas sejam capazes de simular o componente de desvio ambiental da variância fenotípica e considerando que marcadores SNPs possam predizer o componente genético dessa variância, esse projeto objetivou evoluir em direção a um modelo matemático que considere a variabilidade entre animais em seu nível genético. Seguindo conceitos de fisiologia genômica computacional, nós assumimos que a variância genética da característica complexa (i.e. produto do comportamento do modelo) surge de características componentes (i.e. parâmetros dos modelos) em níveis hierárquicos mais baixos do sistema biológico. Esse estudo considerou dois modelos mecanicistas do crescimento de bovinos – Cornell Cattle Value Discovery System (CVDS) e Davis Growth Model (DGM) – e ao questionar se os parâmetros de tais modelos mapeariam regiões genômicas que englobam QTLs já descritos para a característica complexa, verificou as suas interpretações biológicas esperadas. Tal constatação forneceu uma prova de conceito de que os parâmetros do CVDS e do DGM são de fato fenótipos cuja interpretação pode ser confirmada através das regiões genômicas mapeadas. Um método de predição genômica foi então utilizado para computar os parâmetros do CVDS e do DGM. Os efeitos dos marcadores SNPs foram estimados tanto para os parâmetros quanto para os fenótipos observados. Nós buscamos qual o melhor cenário de predição – simulações dos modelos com parâmetros computados a partir das informações genômicas ou predição genômica conduzida diretamente nos fenótipos complexos. Nós encontramos que enquanto a predição genômica dos fenótipos complexos pode ser uma melhor opção em relação aos modelos de crescimento, simulações conduzidas com parâmetros obtidos a partir de dados genômicos estão condizentes com simulações geradas

com parâmetros obtidos a partir de métodos regulares. Esse é o principal argumento para chamar atenção da comunidade científica de que a abordagem apresentada nesse projeto representa um caminho para o desenvolvimento de uma nova geração de modelos nutricoinais aplicados capazes de capturar a variabilidade genética entre bovinos de corte confinados e produzir simulações com variáveis de entrada específicas de cada genótipo. Esse projeto é a primeira abordagem no Brasil conhecida dos autores a usar genótipos *Bos indicus* para o estudo da aplicação de genômica integrada à modelos mecanicistas para o manejo e comercialização de animais na pecuária.

Palavras-chave: crescimento de bovinos de corte, fisiologia genômica computacional, mapeamento genótipo-fenótipo, predição genômica, modelagem mecanicista.

Abstract

FREUA, M.C. **Use of genetic variance in dynamic mechanistic models of growth to predict cattle performance and carcass composition under feedlot conditions**. 2015. 87 f. M.Sc. Dissertation – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2015.

The prediction of phenotypic variance is important for beef cattle operations to increase profitability by optimizing resource use. Dynamic mechanistic models of cattle growth have been used as decision support tools for individual cattle management systems. However, the application of such models is still based on population parameters, with no further approach to capture between-subject variability. By assuming that mechanistic models are able to simulate environmental deviations components of phenotypic variance and considering that SNPs markers may predict the genetic component of this variance, this project aimed at evolving towards a mathematical model that takes between-animal variance to its genetic level. Following the concepts of computational physiological genomics, we assumed that genetic variance of the complex trait (i.e. outcome of model behavior) arises from component traits (i.e. model parameters) in lower hierarchical levels of biological systems. This study considered two mechanistic models of cattle growth – Cornell Cattle Value Discovery System (CVDS) and Davis Growth Model (DGM) – and verified their expected biological interpretation by asking whether model parameters would map genomic regions that harbors QTLs already described for the complex trait. This provided a proof of concept that CVDS and DGM parameters are indeed phenotypes whose expected interpretations may be stated by means of their mapped genomic regions. A method of genomic prediction to compute parameters for CVDS and DGM was then used. SNP marker effects were estimated both for their parameters and observed phenotypes. We looked for the best prediction scenario – model simulation with parameters computed from genomic data or genomic prediction on complex phenotypes directly. We found that while genomic prediction on complex phenotypes may still be a better option than predictions from growth models, simulations conducted with genomically computed parameters are in accordance with those performed with parameters obtained from regular methods. This is the main argument to call attention from the research community that this approach may pave the way for the development of a new generation of applied nutritional models capable of representing genetic variability among beef cattle under feedlot conditions and performing simulations

with inputs from individual's genotypes. To our knowledge, this project is the first of this kind in Brazil and the first using *Bos indicus* genotypes to study the application of genomics integrated with mechanistic models for the management and marketing of commercial livestock.

Keywords: beef cattle growth, computational physiological genomics, genotype-to-phenotype mapping, genomic prediction, mechanistic modeling.

Contents

1 General introduction

Mathematical modeling has been a very active research area in Animal Sciences. The first scientific conference on the subject, the *International Workshop on Modeling Nutrient Utilization in Farm Animals*, happened in Hurley, England, in 1979. Today, models are recognized as promising tools for decision support in the industry and for designing scenarios to guide field experiments. The main example is the discipline of Animal Nutrition, which is no longer conceived without modeling. The standard nutritional systems published by the National Research Council (NRC) which direct many research agendas on animal nutrition worldwide are essentially comprised of mathematical models.

One of the main components of model development is parameterization: the process of fitting models to the data in an attempt to derive parameter values from measurements (GUNAWARDENA, 2010). It is then very clear to see that models are somewhat dependent on the range of values its variables assume in the dataset used for computing their parameters. This becomes a limitation for modeling. Although the possibility to "foresee" an individual animal characteristic through simulation is what makes modeling attractive for the livestock industry, phenotypic measurements are still needed for parameterization. In addition, between-animal variability is only possible to be captured by model simulation if there is a way to compute subject-specific parameters for entering the model when simulations are performed.

Quantitative genetics is another discipline in Animal Sciences that is concerned with phenotypic prediction. The animal model of Henderson (1984), which is widely applied for animal breeding, is used to estimate breeding values (EBV) that guide selection of animals with greater genetic merit in an attempt to improve the phenotypic mean of a population over generations.

With the advent of genomic technologies new methods for phenotypic prediction have been developed. These novel approaches rely on the genomic best linear unbiased prediction (GBLUP) concept (MEUWISSEN; HAYES; GODDARD, 2001). They have initially been applied as a further development of the Henderson model. Following the basic animal model, if the relationship matrix from pedigree is replaced by a genomic relationship matrix, the model outcome is called GEBV: genomic estimated breeding value. On the other hand, if no relationship matrix is used in the statistical model, but instead, the matrix of random effects is set to be the genotype of a genomic marker such as single nucleotide polymorphisms (SNPs), then the model outcome is the estimated marker effects (ENDELMAN, 2011). Both

approaches are used for genomic selection, which tries to improve accuracy of the more traditional EBVs at the same time it increases genetic gain by also managing the generation interval as these technologies allow estimating genetic merit for younger bulls.

Given what is said, a straightforward conclusion is that the interplay between nutritional modeling and genomics may be designed from the parameterization problem. It seems reasonable to assume that genomic prediction could assist modeling by allowing prediction of input variables or even parameters without the need of measuring these variables experimentally. In fact, this was recognized as soon as genomic prediction methods were first proposed. Looking at patent repositories it is possible to find full systems designed to integrate nutritional models with genomics for commercial application in beef cattle feedlots which dates from the beginning of the 2000's (BATES et al., 2005; BUENFIL et al., 2009; DENISE et al., 2004; FLUHARTY; JACKWOOD, 2001; TAYLOR et al., 2002).

These first attempts comprised linear empirical models whose predictors were molecular genetic values. Typical response variables were average daily gain (ADG), dry matter intake (DMI), or variables associated with meat quality traits. The main concept is that at feedlot entry individuals would be genotyped and inputs (i.e. model predictors) would be obtained to enter the models for subject-specific simulation. From this analysis, animals could be better grouped within pens and targeted to specific markets considering their growth potential and carcass traits – a strategy recognized to enhance profitability of feedlot operations. Another approach was typically targeting animals to specific markets by inferring their growth potential from any sort of genotypic line useful to classify animals among different groups.

It appears that these first attempts had not evolved to something really applicable until recently. GeneMax™[1] has developed programs that resemble those systems proposed by Taylor et al. (2002) and are designed for feedlot operations with Angus cattle. But their range of application is still very narrow in terms of breeds, costs, production scenarios, and variables considered.

From these approaches one thing is certain: the need to develop strategies to manage cattle individually. Individual cattle management systems (ICMS) may improve overall profitability of the beef industry, minimize excess fat produced, increase consistency of product, and depending on the contract between cattle owners and feedlot operations, or between the cattle owner and the slaughter plant, ICMS may help to identify and reward

[1] http://www.cabpartners.com/genemax/

individual owners for superior performance (TEDESCHI; FOX; GUIROY, 2004). In fact, such strategic alliance is now consolidated in the United States, particularly after the concepts of feed efficiency were better defined (ØSTERGAARD, 1990). Recently, ICMS has gained more attention due to the intricate association between livestock production and climate change (GERBER et al., 2013). At the same time feed efficiency enhances profitability, it serves as a strategy to mitigate greenhouse gases emission by unit of production (e.g. meat or milk). Moreover, ICMS is the unique management strategy to deal with "individual efficiency".

Few nutritional models directed to beef cattle have incorporated the concept of ICMS into their structure. The Cornell Cattle Value Discovery System (CVDS) is one example that was first developed by Fox and Black (1984) and then improved by Tedeschi, Fox and Guiroy (2004). The CVDS model is comprised of sequential equations based on the NRC (2000) system. Animal characteristics, diet and environmental information are used to simulate daily individual DMI, body weight (BW) and carcass composition. The CVDS also computes a reverse simulation. Observed ADG and diet characteristics may enter the model to predict DMI using a backward calculation technique. In this case, DMI is conveniently named dry matter required (DMR), in a reference to the expected intake that would be necessary for the animal to meet observed performance. The basic model underlying CVDS is the same one adopted by the NRC (2000), the only difference is on how CVDS assumes the dynamics of the conversion efficiency of net energy for gain (k_g). In CVDS, the daily net energy available for gain (NEg) is computed from dietary metabolizable energy converted to NEg by k_g which is obtained from an exponential decay equation that considers protein deposition; therefore, k_g is not a fixed coefficient for CVDS.

The Decision Evaluator for the Cattle Industry (DECI) has also evolved towards ICMS. This model was developed at the US Meat Animal Research Center, Clay Center, Nebraska. DECI is based on the growth model of Keele, Williams and Bennett (1992), and was later incorporated into a net energy system by Williams and Jenkins (1998, 2003ab). As CVDS, DECI uses rate of BW gain to predict the composition of empty BW gain and feed requirements for maintenance and gain. Different from CVDS, DECI is based on a system of thirteen ordinary differential equations. DECI also computes a reverse simulation which yields a DECI-predicted average daily DMI. In this case, the observed ADG is used as the main input in the model and two parameters that determine body composition are iterated until the predicted ending BW and composition are the same as that observed.

ICMS may not only help allocate individuals to pens but also these models may assist animal breeding and selection for feed efficiency. From reverse simulation, both models yield predictions of daily DMI which may replace observations of individual intake, often an expensive and laborious measurement for the cattle industry, mainly when we consider genetics evaluations such as progeny tests and others that do need of many animals to compose an EBV with acceptable accuracy. In the case of CVDS, the difference between observed DMI and DMR is named predicted intake difference (PID), a variable that may be used as a proxy for feed efficiency and represent an alternative to the traditional residual feed intake (RFI).

Tedeschi et al. (2006) and Williams (2010) have assessed the viability of using CVDS and DECI predictions to replace observed individual DMI or measurements of feed efficiency from Angus (*Bos taurus*) cattle. DMR:ADG was able to explain 84% of the variation in the actual feed:gain ratio. Hence, CVDS may be used to identify differences in the feed:gain or gain:feed ratios by computing DMR for individual growing cattle fed in groups. Moreover, alternatives for the usual RFI such as PID are better able to identify animals with low DMI and small ADG as inefficient compared with actual RFI. Thus biological simulation models could provide RFI estimates that are closer to the true biological efficiency of animals. This is because CVDS and DECI predict performance on an individual-animal basis in contrast to empirical linear models that estimate a fixed relationship between DMI and ADG for all animals.

It is important to take into account the stage of development of the current models. Even in the case of CVDS and DECI which have been extensively tested and updated, mathematical descriptions of the biological hierarchies underlying a phenotype are still very fundamental and represent only the most basic nutritional concepts. If the models were "more mechanistic", certainly other outcomes would bring a different set of conclusions. Thus, an interesting aspect of the possibility to use model predictions in genetic analysis is to assess genetic parameters of these predictions when compared to actual measurements.

Kirschten (2007), Kirschten, Pollak and Fox (2007), and Williams et al. (2006) have submitted CVDS and DECI predictions to methods of quantitative genetics and found that predicted and observed DMI show high genetic correlation. For CVDS, genetic correlation between observed DMI and feed required for maintenance and gain ranged from 0.63 to 0.89. For DECI, this correlation was 0.77. Heritability of predicted measures of DMI ranged from 0.33 to 0.41. These strong genetic relationships indicate that these predicted measurements of DMI may be used in genetic evaluations.

Genetics is in its most basic concepts entirely defined towards individuals and thus ICMS. Some researchers have hypothesized that combining traditional genomic approaches and mechanistic descriptions of phenotypes may support the discovery of candidate genes (FU et al., 2011; WANG et al., 2012). In this context a new vocabulary emerges to differentiate levels of aggregation embedded within selection criteria. Complex traits, which are polygenic and may be understood as outcomes of model behavior, are referred as higher-phenotypes, in a reference to where it stands throughout the biological hierarchy that mechanistically describes a given phenotype. Model core parameters, which are understood as entities that define the system structure, are referred as lower-level phenotypes or component traits of higher-level phenotypes. Thus, parameters are important components for modeling between-animal variability.

A very basic but still valid approach in the interplay of biological models and genetics is that where EBVs are obtained for parameters of highly empirical equations such as nonlinear growth functions like Gompertz, von Bertalanffy, Richards, and many others (FORNI et al., 2008). Nonetheless, these approaches do not contribute to that basic concept of Taylor et al. (2002) because growth functions do not describe a given phenotype as arising from the intersection between biological entities, which greatly lowers the potential to translate the feedlot reality into mathematical descriptions. However, as demonstrated by Lusk (2007), the use of such equations could be used as a criteria for grouping animals into homogeneous pens as they capture between-leptin genotypes variability in growth trajectories.

The functional genome-wide association analysis (*fGWAS*) method proposed by Das et al. (2011) is also within the same context of highly empirical models, however, it estimates SNPs effects that comprise the whole-genome. In an attempt to better capture the dynamics of longitudinal traits such as BW, the expectation of a phenotype is described in the statistical model by a previously defined nonlinear function such as Gompertz or even Legendre polynomials. This statistical approach allows for estimating both additive and dominance effects of genes under a nonlinear mixed effect framework (NLME). The main advantage of *fGWAS* is that it analyzes the patterns of genetic control over development, the duration of genetic effects, and what causes developmental trajectories to change or stop changing. *fGWAS* is focused on estimating fixed and random effects of model parameters, as is true for any NLME model.

Causally-cohesive phenotype-to-genotype modeling (cGP) is a very recent approach that attempts to integrate mechanistic models with genomics. The cGP method was proposed by Vik et al. (2011) under the umbrella of the new discipline *computational physiological*

genomics. Different from *fGWAS*, cGP models are indeed developed for subject-specific simulation. For now, its main application is to the development of predictive modeling as a tool for personalized human medicine, where the genotype is used as an input to compute model parameters (NORDBØ et al., 2015).

As pointed earlier, integration of system dynamics with genomics by focusing on model parameters is not only pragmatic and intuitive but also brings a biological explanation. Genotype-to-phenotype mapping is not an easy task because their direct association is difficult. This fact explains why the rise of genomic science is still deterred by the so called missing heritability, i.e. genotyping individuals has not been enough to explain full heritable variance (BOGARDUS, 2009; MAHER et al., 2008; MANOLIO et al., 2009). The map from the genotype space to the phenotype space is still largely unknown because of a reaction space that bridges the gap. It is reasonable to hypothesize that dynamic mathematical models that have some description of the biological hierarchy underlying a phenotype bring in their structure a more stable behavior throughout this reaction space. In other words, genetic variance penetrates parameters more directly and linearly (VIK et al., 2011; WANG et al., 2012). The biological model structure is reflected in its core parameters, which are responsible for the description of individual variability within the model.

The main goal of the research described herein is to contribute to the discussions about the integration between system dynamics with genomics as applied to the livestock industry. The framework adopted is that of integration through model parameters. Specifically, we depart from the first attempts of Taylor et al. (2002), but update the approach with more mechanistic models and using recent genomic technologies and methods such as SNP markers, GWAS and genome-wide prediction.

Many growth models may serve as a standard for exploring the possibility of obtaining nutritional model predictions with parameters computed from genomic information. Two of those available on the scientific literature offer an interesting contrast when it comes to the aggregation level: the CVDS, and the Davis Growth Model (DGM), which refers to the system of ordinary differential equations developed by Oltjen et al. (1986) describing DNA accretion, protein synthesis and degradation, and fat synthesis. The reason why we chose DGM and not DECI is because differently from CVDS and DECI, the DGM first simulates protein turnover, fat deposition and whole-body cell proliferation to later compute BW gain from body composition.

To our knowledge this is the first work to present to the Animal Science research community the concepts and potentials of computational physiological genomics. For a

broader audience, our work may serve as evidence that the concept of genetic variation penetrating parameters is indeed promising, once we demonstrate, by contrasting genomic regions mapped for model parameters (i.e. genes and quantitative trait loci) to mathematical formalisms, their biological validity. This is only possible because a real dataset comprised of phenotypic and genotypic information of Nellore (*Bos indicus*) cattle is used in the analyses.

In addition, we present and discuss the potential of alternative prediction scenarios for complex phenotypes that are important for the livestock industry – i.e. predicting them at the beginning of the production system may enhance profitability. We show that parameters computed by means of genomic information may yield very reasonable model outcomes when compared to simulations that are done with parameters computed regularly.

This dissertation is organized as follows. First, we describe our initial attempts to integrate CVDS and DGM with genomic studies under what we call the NLME-Mechanistic framework. Due to inconsistencies of parameter estimation, this approach was then replaced by a two-step strategy adopted to develop the subsequent analyses. Chapters 3 and 4 report GWAS on parameters of the CVDS and DGM and aim at assessing the validity of the biological interpretation of these parameters despite what is imposed by the model formalism. Finally, Chapter 5 considers a method of genomic prediction to estimate SNP marker effects in a reference population and compute CVDS and DGM parameters for a training dataset. These parameters enter the models for predictive purposes. Comparisons between simulations and actual measurements are discussed as well as simulations performed with parameters computed from genomic information versus those conducted with parameters obtained from regular methods. A general discussion with final remarks is presented.

References

BATES, S. et al. **Methods and systems for inferring traits to manage non-beef livestock**. [s.l.] Google Patents, 2005.

BOGARDUS, C. Missing Heritability and GWAS Utility. **Obesity**, v. 17, n. 2, p. 209, 2009.

BUENFIL, M. A. G. et al. **Animal management**. [s.l.] Google Patents, 2009.

DAS, K. et al. A dynamic model for genome-wide association studies. **Human Genetics**, v. 129, n. 6, p. 629, 2011.

DENISE, S. K. et al. **Compositions, methods, and systems for inferring bovine breed**. [s.l.] Google Patents, 2004.

ENDELMAN, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. **The Plant Genome Journal**, v. 4, n. 3, p. 250, 2011.

FLUHARTY, F.; JACKWOOD, D. J. **Methods for assessing the beef characteristics of live cattle**. [s.l.] Google Patents, 2001.

FORNI, S. et al. Comparison of different nonlinear functions to describe Nelore cattle growth. **Journal of Animal Science**, v. 87, n. 2, p. 496-506, 2008.

FOX, D. G.; BLACK, J. R. A system for predicting body composition and performance of growing cattle. **Journal of Animal Science**, v. 58, n. 3, p. 725-739, 1984.

FU, G. et al. A Dynamic Model for Functional Mapping of Biological Rhythms. **Journal of Biological Dynamics**, v. 5, n. 1, p. 84, 2011.

GERBER, P. J. et al. **Tacking climate change through livestock - a global assessment of emissions and mitigation**. Rome: Food and Agriculture Organization of the United Nations (FAO), 2013.

GUNAWARDENA, J. Models in Systems Biology: The Parameter Problem and the Meanings of Robustness. In: Lodhi, H. M. (Ed.). **Elements of Computational Systems Biology**. John Wiley & Sons, Inc., 2010. p. 19--47.

HENDERSON, C. R. **Applications of Linear Models in Animal Breeding**. Ontario, Canada: University of Ghelph, 1984.

KEELE, J. W.; WILLIAMS, C. B.; BENNETT, G. L. A computer model to predict the effects of level of nutrition on composition of empty body gain in beef cattle: I. Theory and development. **Journal of Animal Science**, v. 70, p. 841-857, 1992.

KIRSCHTEN, D. P. **Genetic Evaluation of Cornell Value Discovery System Model Predicted Feed Intake in Beef Cattle: Genetic Parameters Between Actual and Model-Predicted Feed Intake and Correlations with Other Traits**. 2007. Tese (Doutorado) - Cornell University, Ithaca, NY.

KIRSCHTEN, D. P.; POLLAK, E. J.; FOX, D. G. Use of a mathematical computer model to predict intake in Angus cattle: Genetic parameters between observed and predicted values, and relationships with other traits. **Anais**...San Antonio, TX: Federation of Animal Science Societies, 2007

LUSK, J. L. Association of single nucleotide polymorphisms in the leptin gene with body weight and backfat growth curve parameters for beef cattle. **Journal of Animal Science**, v. 85, n. 8, p. 1865-1972, 2007.

MAHER, B. et al. The case of the missing heritability. **Nature**, v. 456, p. 18-21, 2008.

MANOLIO, T. A. et al. Finding the missing heritability of complex diseases. **Nature**, v. 461, n. 7265, p. 747-753, 2009.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, v. 157, p. 1819-1829, 2001.

NATIONAL RESEARCH COUNCIL. **Nutrient Requirements of Beef Cattle**. 7th ed. Washington, D.C.: National Academy Press, 2000.

NORDBØ, Ø. Y. et al. Towards causally cohesive genotype–phenotype modelling for characterization of the soft-tissue mechanics of the heart in normal and pathological geometries. **Journal of The Royal Society Interface**, v. 12, n. 106, abr. 2015.

OLTJEN, J. W. et al. Development of a Dynamic Model of Beef Cattle Growth and Composition. **Journal of Animal Science**, v. 62, p. 86--97, 1986.

ØSTERGAARD, V. et al. Main report — E.A.A.P. working group on: Efficiency in the dairy cow. **Livestock Production Science**, v. 24, n. 4, 1990.

TAYLOR, J. F. et al. **Method of managing and marketing livestock based on genetic profiles**. [s.l.] Google Patents, 2002.

TEDESCHI, L. O. et al. Identifying differences in feed efficiency among group-fed cattle. **Journal of Animal Science**, v. 84, p. 767-776, 2006.

TEDESCHI, L. O.; FOX, D. G.; GUIROY, P. J. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. **Agricultural Systems**, v. 79, n. 2, p. 171, 2004.

VIK, J. O. et al. Genotype–phenotype map characteristics of an in silico heart cell. **Frontiers in Physiology**, v. 2, dez. 2011.

WANG, Y. Et al. Parameters in Dynamic Models of Complex Traits are Containers of Missing Heritability. **PLoS Computational Biology**, v. 8, n. 4, p. e1002459, 2012.

WILLIAMS, C. B.; JENKINS, T. G. A computer model to predict composition of empty body weight changes in cattle at all stages of maturity. **Journal of Animal Science**, v. 76, p. 980-987, 1998.

WILLIAMS, C. B.; JENKINS, T. G. A dynamic model of metabolizable energy utilization in growing and mature cattle. I. Metabolizable energy utilization for maintenance and support metabolism. **Journal of Animal Science**, v. 81, p. 1371-1381, 2003a.

WILLIAMS, C. B.; JENKINS, T. G. A dynamic model of metabolizable energy utilization in growing and mature cattle. II. Metabolizable energy utilization for gain. **Journal of Animal Science**, v. 81, p. 1382-1389, 2003b.

WILLIAMS, C. B. et al. Using simulation models to predict feed intake: Phenotypic and genetic relationships between observed and predicted values in cattle. **Journal of Animal Science**, v. 84, p. 1310-1316, 2006.

WILLIAMS, C. B. Application of biological simulation models in estimating feed efficiency of finishing steers. **Journal of Animal Science**, v. 88, n. 7, p. 2523, 2010.

2 The NLME-Mechanistic framework: describing the response of a nonlinear mixed effect model with dynamic models of cattle growth

Following the approach to integrate nutritional models and genomic science through model parameters, our initial attempt was mainly based on the previous work of Das et al. (2011), where the authors report a functional genome-wide association method whose the expectation of the response in a nonlinear mixed effect model (NLME) is defined using a biological function. For instance,

$$y_{ij} = f(t_{ij}, \varphi_{ij}) + \varepsilon_{ij}$$
$$\varphi_i = h(\mu, c_i, \eta_i) \qquad \text{eq.(1)}$$

Where y_{ij} is the jth observation of subject i, t_{ij} is the time regression variable, φ_i is the vector of individual parameters for subject i, c_i is a known vector of covariates, $h(.)$ is any transformation function, μ is an unknown vector of fixed effects, η_i is an unknown vector of normally distributed random effects with mean zero and covariance Ω, ε_{ij} is the residual error, a random variable with mean zero and variance 1.

However, for this dissertation we had to update eq.(1) by modeling y_{ij} not with a single function $f(.)$ but with a mechanistic model comprised of a system of equations, as is the case for the Cornell Cattle Value Discovery System (CVDS, TEDESCHI; FOX; GUIROY, 2004) and the Davis Growth Model (DGM, OLTJEN et al., 1986). For CVDS, solving eq.(1) was more pragmatic than the solution for DGM because CVDS is comprised of equations in closed-form. The DGM requires an algorithm to solve initial value ordinary differential equations (ODE). Therefore, instead of just describing y_{ij} with DGM, an implementation of an ODE solver was necessary.

We chose a NLME framework because it allows decomposing model parameters into fixed and random effects. Fixed terms are source of explained variation in y_{ij} (e.g. contemporary group, genotype of a given genetic marker, etc.) and the random term models residual between-animal variation. Despite of obtaining distributions of CVDS and DGM parameters for a given dataset, NLME would yield statistics to test whether fixed effects, and then genotypes, had a significant contribution to explain between-animal variation. In a second step, these fixed effects would have been used to differentiate genotypes in terms of the growth model parameters and provide the necessary information to evolve towards

individual genotypic simulations. Specifically, we adopted genotypes for a small set of single nucleotide polymorphism (SNP) previously identified as being associated with average daily gain and dry matter intake in Nellore (*Bos indicus*) cattle under feedlot conditions (SANTANA et al., 2014ab).

For solving eq.(1) within a NLME framework a proper algorithm is necessary. Many solutions for NLME are available on the literature, most of them based on the standard Expectation-Maximization (EM) algorithm (PINHEIRO; BATES, 2000; LITTELL et al., 2006), where the likelihood function is approximated (e.g. linearization, quadrature approximation, etc.). The Stochastic Approximation Expectation Maximization algorithm (SAEM) as proposed by Kuhn and Lavielle (2005) has the advantage to converge very quickly to a neighborhood of the Maximum Likelihood Estimate and could also be used in an empirical Bayesian context for estimating the prior distribution of model parameters. In addition, with SAEM there is no need to perform any sort of approximation of *f(.)* in eq.(1), once the usual E-step of EM is replaced by a stochastic procedure. We then implemented in the R statistical environment an intersection between SAEM and the growth models using the packages saemix (COMETS; LAVENU; LAVIELLE, 2011) and deSolve (SOETAERT; PETZOLDT; SETZER, 2010). As suggested by Emanuelle Comets (personal communication), a strategy for such intersection was passing inputs required by the growth models as arguments of the *saemix(.)* function, i.e. as independent variables of eq.(1). It happens that as inputs of CVDS and DGM are correlated to each other, such correlation spreads over the SAEM algorithm as colinearity, and thus, hindering computation of a matrix inverse. There is no practical solution to this problem.

To overcome this situation, the adoption of a different method was a possibility. Fu et al. (2011) has proposed a statistical model similar to that o Das et al. (2011), but implementing a system of ODE's as *f(.)* in eq.(1). However, adopting Fu et al. (2011) would require a multivariate likelihood function to solve NLME. Thus, for the sake of simplicity and also to improve the genomic background of our approach a different solution was designed.

Instead of performing the analysis with a single-step approach, a two-step method where model parameters are treated as regular phenotypes would allow not only the discussion about between-animal variation of their values, but also application a broad set of the main genomic methods currently available in the literature: genome-wide association study and genomic prediction. In addition, analysis would be conducted not only considering previously identified SNPs, but now with SNPs covering the whole-genome. Moving away

from NLME also relaxed the restriction that repeated measures of y_{ij} were needed for each subject, and thus the available phenotypic and genotypic dataset was increased.

It is important to note that parameters of CVDS and DGM have a very different interpretation from that of regular nonlinear growth functions such as Gompertz and Richards. CVDS and DGM are dynamic mechanistic models of cattle growth and their parameters are closer to the *causally-cohesive phenotype-to-genotype* concept of Wang et al. (2012) than that of the already disseminated approach of Forni et al. (2008, 2009) and Lusk (2007). Hence, the following chapters were conceived after the considerations exposed above.

References

COMETS, E.; LAVENU, A.; LAVIELLE, M. SAEMIX, an R version of the SAEM algorithm. In**:** 20th MEETING OF THE POPULATION APPROACH GROUP IN EUROPE, 2011. Athens, Greece. Abstr 2173

DAS, K. et al. A dynamic model for genome-wide association studies. **Human Genetics**, v. 129, n. 6, p. 629, 2011.

FORNI, S. et al. Comparison of different nonlinear functions to describe Nelore cattle growth. **Journal of Animal Science**, v. 87, n. 2, p. 496-506, 2008.

FORNI, S. et al. A dynamic linear model for genetic analysis of longitudinal traits. **Journal of Animal Science**, v. 87, n. 12, p. 3845-3853, 2009.

FU, G. et al. A Dynamic Model for Functional Mapping of Biological Rhythms. **Journal of Biological Dynamics**, v. 5, n. 1, p. 84, 2011.

KUHN, E.; LAVIELLE, M. Maximum likelihood estimation in nonlinear mixed effects models. **Computational Statistics & Data Analysis**, v. 49, n. 4, 2005.

LITTELL, R. C. et al. **SAS System for Mixed Models**. 2nd. ed. Cary, N.C.: SAS Inst. Inc., 2006.

LUSK, J. L. Association of single nucleotide polymorphisms in the leptin gene with body weight and backfat growth curve parameters for beef cattle. **Journal of Animal Science**, v. 85, n. 8, p. 1865-1972, 2007.

OLTJEN, J. W. et al. Development of a Dynamic Model of Beef Cattle Growth and Composition. **Journal of Animal Science**, v. 62, p. 86--97, 1986.

PINHEIRO, J.; BATES, D. **Mixed effects models in S and S-PLUS**. Traducao. New York, NY: Statistics and Computing, Springer, 2000.

SANTANA, M. H. et al. Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. **BMC Genetics**, v. 15, n. 1, p. 21, 2014a.

SANTANA, M. H. A. et al. Genome-wide association study for feedlot average daily gain in Nellore cattle (Bos indicus. **Journal of Animal Breeding and Genetics**, v. 131, n. 3, p. 210--216, 2014b.

SOETAERT, K.; PETZOLDT, T.; SETZER, R. W. Package deSolve : Solving Initial Value Differential Equations in R. **Journal of Statistical Software**, v. 33, n. 9, p. 1--25, 2010.

TEDESCHI, L. O.; FOX, D. G.; GUIROY, P. J. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. **Agricultural Systems**, v. 79, n. 2, p. 171, 2004.

WANG, Y. et al. Parameters in Dynamic Models of Complex Traits are Containers of Missing Heritability. **PLoS Computational Biology**, v. 8, n. 4, p. e1002459, 2012.

3 Parameters of a dynamic mechanistic model of cattle growth retain enough biological interpretation for genotype-to-phenotype mapping

Abstract

A broader application of animal genomics is deterred by the fact that genotyping individuals cannot fully explain the heritable variance yet. Similarly, the application of mathematical models to livestock production is still low due to the necessity to measure many predictor variables. The new scientific field of computational physiological genomics, which states that model core parameters would be closer to the genetic variance of outcome variables, emerges from the combination of dynamic mechanistic models of production traits with single nucleotide polymorphism (SNP) models. It is hypothesized that this framework could assist in the predictability of a phenotype. We investigate the suitability of this concept when using a dynamic model of cattle growth. A dataset comprised of genotypic and phenotypic information of Nellore (*Bos indicus*) cattle was used for a genome-wide association analysis in order to contrast the biological interpretation of core parameters as imposed by the model formalism to their associated genomic regions and nearby quantitative trait loci (QTLs). SNP markers were used to develop prediction equations for the core parameters, which enter the model for simulative prediction purposes. QTLs and genes surrounding genomic regions mapped for two parameters of a dynamic model, one related to mature body weight and another to growth efficiency, are in accordance with what is stated by the model equations. Significantly associated SNPs were used to compute parameters, which yielded reasonable model outcomes when compared to regular parameter computations. Our results showed strong evidence of the biological validity of using parameters as component traits of higher phenotypes and the possibility of using genomic data for genotype-to-parameter mapping. The integration of genomics and system dynamics opens up innumerous opportunities to advance towards not only the discovery of candidate genes for important phenotypes in the Animal Sciences, but also on the development theory of predictive modeling for decision support tools.

3.1 Introduction

Quantitative genetics has traditionally explained the genetic variance of a phenotype as a decomposition of a random term of a linear statistical model into the covariance between individuals (HENDERSON, 1984). In animal genetics, the description of the expectation of a given phenotype is usually defined by a linear model, regardless the complex biological hierarchy underlying the trait. With the advancement of genomic tools and despite many complex models of gene action that has already been implemented (DUTHIE et al., 2010; KARISA et al., 2013; SANTANA et al., 2014), statistical models still fall in the category where the response of a continuous phenotype, assumed to be polygenic, is linearly related to the gene effect. Studies of nonlinearity between phenotypes and gene action have been reported (FRANCESCONI; LEHNER, 2014; GJUVSLAND; PLAHTE; OMHOLT, 2007; PENG et al., 2015), but in such a complex set of variables interacting to yield the phenotype, different approaches might provide some additional insights.

Genotype to phenotype mapping is not an easy task because their direct association is difficult. This fact explains why the rise of genomic science is still deterred by the so called missing heritability, i.e. genotyping individuals has not been enough to explain full heritable variance (BOGARDUS, 2009; MAHER, 2008; MANOLIO et al., 2009). The map from the genotype space to the phenotype space is still largely unknown because of a reaction space that bridges the gap. It is reasonable to hypothesize that dynamic mathematical models that have some description of the biological hierarchy underlying a phenotype bring in their structure a more stable behavior throughout this reaction space. In other words, genetic variance penetrates parameters more directly and linearly (VIK et al., 2011; WANG et al., 2012). The structure of the biological model is reflected in its core parameters, which are responsible for capturing between-subject variability within the model.

Mathematical modeling is a very solid discipline in the field of Animal Science. Especially in animal nutrition, models have long been used to describe nutrient requirements by the animal, nutrient supplied by the feed, and the growth and development of various livestock species. Assuming that such models correctly represent the environmental term of phenotypic variance, it is then possible that dynamic mechanistic models theory combined with single nucleotide polymorphism (SNP) models could improve phenotype predictability and gain insight on genes and metabolic pathways underlying phenotypic variation. Approaches toward this end have come to promising conclusions. Model parameters, or lower-level phenotypes, bring the potential to find more causative SNPs than association

studies performed on higher and complex phenotypes directly, i.e. outcomes of model behavior (WANG et al., 2012).

As a case study, we used the Cornell Cattle Value Discovery System (CVDS, TEDESCHI; FOX; GUIROY, 2004) a dynamic mechanistic model of cattle growth that is able to simulate individuals. Our hypothesis was whether the model core parameters would retain enough biological interpretation to aid the genotype-to-phenotype mapping of complex phenotypes. First, we contrasted the biological interpretation of such parameters as imposed by the model structure to their associated genomic regions and nearby QTL using a genome-wide association analysis (GWAS). Then, we evaluated the possibility of using genotypes of SNP markers to predict parameter values by comparing model outcomes from simulations with either regular or genetically estimated parameters. Our goal was twofold: to provide a case study with a real dataset and to stimulate discussions on what opportunities may arise from the application of the new discipline of computational physiological genomics to the Animal Sciences filed.

3.2 Material and methods

3.2.1 Ethics statement

No statement by the local animal ethics committee was necessary for this research. The study described herein used a dataset obtained from experiments published elsewhere (ALEXANDRE et al., 2014; GOMES et al., 2013; SANTANA et al., 2012, 2014), whose authors report phenotype records and DNA samples from tests approved by their respective institutional animal ethics committee and by the cattle owners. All animals' procedures were undertaken within common management practices in beef cattle feedlot operations and in accordance with standard veterinary protocols.

3.2.2 The CVDS model

The dynamic, mechanistic model of cattle growth described by Tedeschi, Fox and Guiroy (2004) was adopted for this study. The CVDS model is comprised of sequential equations that simulate cattle growth and development and have been commercially applied to beef feedlots in the United States. The model is based on the National Research Council (NRC), (2000) energy system to predict growth rate and body weight (BW) on a daily basis from the net energy available for maintenance (NEm) and growth (NEg). To account for the fact that the composition of gain at a particular weight is a function of mature BW,

simulations are based on the equivalent shrunk BW, which is the standard reference weight (SRW), an overall mean BW at the same degree of maturity, multiplied by the ratio between the actual shrunk BW (SBW) and a body weight value at 28% empty body fat (EBF). The CVDS calculates dry matter intake (DMI) and NEm requirement to obtain energy available for growth from the remaining energy and the conversion efficiency of metabolizable energy to net energy for gain (k_g). See appendix for further details regarding the CVDS model.

The CVDS model was written in Visual Basic 6 and it is available at http://nutritionmodels.com/cvds.html. All the simulations described herein were performed without the environmental submodel and using the exponential decay adjustment for k_g, as recommended by Tedeschi, Fox and Guiroy (2004). Computations for statistical analysis were conducted in R v.3.1.3 (R CORE TEAM, 2015)

3.2.3 Polymorphic parameters

The SBW adjusted to 28% EBF (adjusted final shrunk body weight, AFSBW) is the core parameter responsible to represent phenotypic variability when comparing CVDS predictions among contemporary individuals fed the same diet in the same feedlot. Because the dataset used herein is comprised of *Bos indicus* animals, we performed simulations to assess what would be the optimal SRW. Values of SRW close to 310 kg of empty BW (EBW) at 22% EBF yielded the best prediction, agreeing with values reported by Tedeschi et al. (2002). For computing AFSBW, an equation from Guiroy et al. (2001) for carcass traits and modified by Baker et al. (2006) for ultrasound information was used.

$$AFSBW = \frac{EBW + 14.26 \times (22 - pEBF)}{0.895} \qquad \text{eq.(1)}$$

Where, EBW is empty body weight (kg) and pEBF is the predicted empty body fat percent from ultrasound information. The constant 14.26 as reported by Guiroy et al. (2001) is the relationship between pEBF and EBW. The 0.895 is assumed by Marcondes et al. (2010) to be a better conversion factor between EBW and SBW for *Bos indicus* animals than that of 0.891 assumed by the NRC (2000).

Equations from Chizzotti, Tedeschi and Valadares Filho (2008) to compute EBW and Chizzotti, Paulino and Valadares Filho (2008) to predict body composition from ultrasound information were used to obtain pEBF.

$$EBW = -15.6 + 0.928 \times SBW \quad \text{eq.(2)}$$

$$pEBF = \frac{-47.26 + 2.82 \times BFT + 0.2993 \times EBW}{EBW} \times 100 \quad \text{eq.(3)}$$

Where, EBW is empty body weight (kg), and BFT is body fat thickness (mm).

The k_g was adopted as a second parameter for this study. Although being simulated by CVDS, k_g requires an initial condition obtained from diet information and it will represent individual variability as the exponential decay function calculates retained energy as protein, which is used to update k_g, and so daily gain, for each time step. For the statistical analysis, a mean value of k_g obtained from simulations over the entire days on feed was used.

3.2.4 Global sensitivity analysis

Change in model parameters means change in outcome variables, usually phenotypes of interest. To assess the parameter to phenotype map we conducted a global sensitivity analysis by sampling AFSBW and k_g values with the Latin-Hypercube Sampling (LHS) method and then running the CVDS model to obtain predicted variables. The impact of parameter on mean DMI, average daily gain (ADG), and final EBF was quantified by linearly regressing these variables on parameter values. This procedure was repeated a hundred times using a Monte Carlo simulation approach. The coefficient of determination was kept for each iteration to obtain distributions. For the LHS, AFSBW and k_g were assumed to be correlated ($r = 0.58$, $P < 0.001$). For adequate simulations, initial body weight was also sampled with LHS, and correlations were assumed for both AFSBW ($r = 0.87$, $P < 0.001$) and k_g ($r = 0.88$, $P < 0.001$).

3.2.5 Animals and phenotypes

Individual measurements (n = 1,435) of BW and backfat thickness by ultrasound were obtained for Nellore (*Bos indicus*) young bulls and steers from fifteen feed efficiency trials conducted in feedlot operations in South (SANTANA et al., 2012), Southeast (ALEXANDRE et al., 2014; GOMES et al., 2013), and Central-West Brazil (SANTANA et al., 2014). Data was used to compute lower-level phenotypes, i.e. individual polymorphic parameters, and perform simulations with CVDS. When diet information was not available, an average value of diet metabolizable energy (ME) was assumed to be 3.0 Mcal/kg. In accordance with Marcondes et al. (2010), initial conditions for NEm and NEg were set at 67% and 44% of

ME, respectively. Phenotypic data was tested for normality with the Shapiro-Wilk test, and individuals that were more than twice the interquartile range above the third quartile or below the first quartile were excluded.

3.2.6 Genomic DNA extraction, genotyping and imputation

Genomic DNA of 893 animals was extracted from either blood samples or hairs pulled from the skin. Samples were obtained from Santana et al. (2014), Gomes et al. (2013), and Alexandre et al. (2014), and DNA was prepared as described by these authors. Genotyping was performed with Illumina BovineHD® BeadChip (777,962 SNPs), Affymetrix Axiom® Genome-Wide BOS1 Array (648,874 SNPs), GGP Indicus Neogen HD® (84,379 SNPs) and Illumina BovineSNP50® version 2 BeadChip (54,609 SNPs) according to manufacturer's protocol. Only genotype calls (standard cluster quality) greater than 0.70 and samples with a call rate higher than 90% were used.

Imputation enables the inference of missing genotypes for animals genotyped with a panel containing fewer markers, which likely allows further gains in accuracy of genomic prediction by adopting denser panels in the model (VENTURA et al., 2014). Imputation from these panels to super-dense panel (SDP) was performed in two steps, and using genotypic information of 3,776 Nellore cattle from a database of the Animal Breeding and Biotechnology group at FZEA/USP (http://usp.br/gmab/). First, the super-dense panel (1,261,128 SNPs) was created combining genotypic information from Illumina BovineHD and Affymetrix BOS1 of 2,604 and 279 animals, respectively. All animals genotyped with Affymetrix BOS1 (279) were also genotyped with Illumina BovineHD. Next, the 893 animals with genotypic and phenotypic records were imputed to SDP level. Imputation accuracy from different panels to SDP was determined by cross-validation analysis within each panel in a parallel investigation and the concordance rate between the imputed and true genotypes across all scenarios was higher than 97.51%. All imputation procedures were performed in FImpute 2.2 program (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014).

Genotypic data was filtered and only autosomal SNPs with a minor allele frequency > 2% and deviation from Hardy-Weinberg equilibrium (χ^2-test, 1 *df*, $P > 1 \times 10^{-5}$) were kept for the association test.

3.2.7 Association analysis

A two-step method, GRAMMAR-Gamma, which accounts for relatedness and population substructure and uses mixed effect modeling and regression to estimate the SNPs

effects (SVISHCHEVA et al., 2012) was used for performing GWAS on the polymorphic parameters. Briefly, in the first step, the fixed effects for contemporary group and age were estimated for the phenotypic data by a mixed model. A Gamma correction factor was calculated from the computed variance-covariance matrix to correct for population substructure (i.e. genomic relationship matrix). In the second step, the transformed phenotypic data was regressed to genotypic information and the estimated SNP effect was corrected by the Gamma factor. These computations were done using the GenABEL v1.7-6 package for R (AULCHENKO et al., 2007). The genome-wide threshold to declare the significant SNPs was a modified version of a Bonferroni correction for multiple testing with 5% significant level (GAO; STARMER; MARTIN, 2008).

The 1-Mb regions surrounding significant SNPs were investigated to determine whether they mapped against any previously described QTL deposited in the cattle QTLdb database (HU et al., 2013). The Gene Ontology Annotation Database (HUNTLEY et al., 2014), the Kyoto encyclopedia of Genes and Genomes (KEGG), (KANEHISA et al., 2014) and Reactome (CROFT et al., 2014) were used to identify potential functions and biological pathways of annotated genes.

3.2.8 Genotype to parameter simulation

After GWAS, significant SNP markers related to AFSBW and k_g were assumed as true causative SNPs of the population. To draw a genotype-to-parameter map, the dataset was divided into two different subsets using a cross-validation approach. The training group was used to obtain predictive equations of parameter values given SNP genotypes. The AFSBW and k_g of individuals from the second subset (i.e., the testing group) were then predicted and entered into the CVDS for simulative prediction purposes. At this stage, genetically predicted k_g was not considered as an input in order to guarantee simulations equal to those performed with the computed parameters, i.e. with initial k_g from diet information and the exponential decay adjustment to compute k_g. However, we kept k_g in the study to assess the possibility of prediction with genomic data.

This procedure was performed using Monte Carlo simulation for 100 iterations to assess the possibility of running the model with parameter values obtained from genomic information. At each iteration, DMI, ADG and final EBF were compared to their values calculated from simulations with computed parameters by regressing them on the newly simulated outcomes to obtain distributions of the coefficient of determination.

Prediction equations for AFSBW and k_g were obtained as a non-weighted allelic profile approach (AULCHENKO et al., 2009), and thus, the allele substitution effect estimated with GWAS was not included in the model. The sum of the major alleles in the genotype of an individual is considered in a joint estimation additive linear model, as follows:

$$y = \beta_i iBW + \beta_j age + \beta_{Bk}(PAB_k + 2PBB_k) \quad \text{eq.(4)}$$

Where *y* is either AFBSW or k_g, *age* is age at the beginning of the feeding period (days), *PAB* and *PBB* are the estimated probabilities of the AB and BB genotypes of the kth SNP marker (k from 1 to kth most significant SNP), β_n are regression coefficients. Initial body weight and age at the beginning of the feedlot period was included in the model because they are usually available at a commercial plant, and thus, may be good predictor variables.

3.3 Results and discussion

Our approach is similar to causally-cohesive genotype-phenotype (cGP) models (WANG et al., 2012) for computation physiological genomics, but the current stage of development of the model adopted here still doesn't allow direct measurement of parameters as phenotypes. So, they were estimated properly. Polymorphic parameters were calculated for the 1,435 animals in the dataset (Table 1), and mean and standard deviation estimates were used for the LHS in the sensitivity analysis (Table 2).

The determination of BW at a given body composition, as represented by AFSBW, is regarded as a key parameter that shape growth models and influence accuracy and precision of model estimates (ARNOLD; BENNETTT, 1991). The AFSBW is responsible for distinguishing individuals with respect to DMI and ADG, as shown by our sensitivity analysis. On average, AFSBW explain 46% and 64% of the phenotypic variation in DMI and ADG, respectively. One of the main sources of error in the CVDS model is the computation of AFSBW when carcass traits or ultrasound measurements are not available, or when it has to be computed from biometric measurements (FOX; SNIFFEN; CONNOR, 1988) in order to obtain growth projections at the beginning of the feeding period. That is why the approach proposed herein of mapping genotype-to-phenotype through parameters is appealing not just to support the discovery of important genomic regions associated with higher level phenotypes but also because using genomic information as predictor is ideal for application in

feedlots, particularly to sort out individuals into more homogenous pens, which typically leads to an increased profitability.

Table 1. Descriptive statistics of the database (n = 1,435)

Phenotype[a]	Minimum	Mean	Maximum	SD[b]
AFSBW	349.5	469.2	679	50.83
k_g	0.45	0.53	0.58	0.02
DOF	35	73.2	90	10.64
iBW	216	357	545.5	60.34
fBW	245	450.3	705.5	78.58
iBF	0.0	1.53	6.35	1.36
fBF	1.0	4.23	15.6	3.9
DMI	5.98	9.03	18.3	1.37
ADG	1.3	1.88	2.56	0.27
fEBF	9.4	14.77	17.89	1.66

[a]AFSBW, adjusted final shrunk body weight (kg); k_g, efficiency of conversion of dietary metabolizable energy to net energy for growth (dimensionless); DOF, days on feed (days); iBW, initial body weight (kg); fBW, final body weight (kg); iBF, initial backfat thickness (mm); fBT, final backfat thickness (mm); DMI, mean dry matter intake for the DOF as simulated by CVDS (kg); ADG, average daily gain for the DOF as simulated by CVDS (kg); fEBF, final empty body fat as simulated by CVDS (% of empty body weight).
[b]SD, standard deviation.

Table 2. Statistics of the coefficient of determination (R^2) from the global sensitivity analysis (n=100)

Phenotypes[a]	AFSBW[b]			k_g[c]		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%
DMI	0.455	0.4	0.504	0.109	0.079	0.14
ADG	0.642	0.592	0.684	0.016	0.004	0.03
EBF	0.003	2.34×10^{-5}	1.13×10^{-2}	0.039	0.012	0.071

[a]*DMI* (dry matter intake), *ADG* (average daily gain), and *EBF* (empty body fat), R^2 refers to the linear equation *y ~ parameter*.
[b]*AFSBW* (adjusted final shrunk body weight). Mean and quartiles for 100 Monte Carlo runs.
[c]k_g (efficiency of conversion of dietary metabolizable energy to net energy for growth). Mean and quartiles for 100 Monte Carlo runs.

The k_g parameter is also responsible for some of the variation in DMI and ADG, and this may be due to the moderate correlation between AFSBW and k_g. However, change in k_g also implies differences in body composition (i.e. final EBF), although rather small (only 7% as revealed by the sensitivity analysis). In fact, higher levels of fat deposition are energetically more efficient than protein deposition, which means that animals that were fatter at slaughter are expected to have a greater k_g as imposed by the CVDS structure. Animals that are able to convert ME from the diet to NEg are expected to be more energetically efficient because for a given energy concentration that is fed, a higher conversion would be expressed as greater ADG. It is important to consider the composition of gain, though, because greater ADG does not necessarily equate to greater protein deposition, but fat deposition. So, the fact that k_g is explaining only a few fraction of EBF indicates that k_g relates to energy efficiency, but very few with fat. Therefore, k_g may be an interesting phenotype to further explore as selection criteria for feed efficiency. The variability of EBF reported by the CVDS model is not fully related to AFSBW and k_g, which means that other biological processes still not modeled contribute to EBF.

We expected a higher influence of AFSBW on body composition due to its potential to influence growth trajectory. Animals with higher AFSBW should have a longer fattening period to reach a targeted body composition when compared to individuals with lower AFBSW values. However, this theory develops at the same initial body weight and days on feeding, and the variation encountered in the dataset might have contributed to decrease the association between AFSBW and simulated final EBF.

3.3.1 SNP associations with AFSBW and k_g

We reported the 10 most significant SNPs found to be associated with AFSBW and k_g (Tables 3 and 4) from a whole-genome analysis performed with 893 animals (Figs. 1 and 2). Many of these SNPs are intron variants of specific genes, and the regions surrounding them harbors QTLs for several production traits (Tables 5 and 6). As expected, SNPs significant for AFSBW are found in regions previously associated with BW at different growth stages (birth, weaning, yearling, slaughter, and mature). Residual feed intake (RFI) and post-weaning ADG, QTLs that are related to feed efficiency and growth trajectory, also confirm the biological interpretation expected for AFSBW. This is also the case for k_g, where QTLs for BW and feed conversion ratio were found. Milk production traits such as milk, protein and fat yield found for AFSBW and k_g, may reveal some relation to energy efficiency, expected especially for k_g.

It is interesting that reproductive traits were also significant QTLs found for both parameters. This may reveal some relationship between reproduction and the animal potential to grow.

The biological interpretation of AFSBW and k_g is also supported by the annotated genes. For instance, for k_g, the *UNC13C* gene is related to the synaptic vesicle cycle (GO:0007268). Although a relation between energy balance and the nervous system is not straightforward, feedback looping between plasticity of adipose tissue, which at some instance may modulate feed intake and the nervous system has already been reported (PÉNICAUD et al., 2000; SUTTON et al., 2014). The CVDS model seems to be in accordance with this metabolic pathway as k_g influences both EBF and DMI (Table 2).

MDH2 gene encodes a malate dehydrogenase localized to the mitochondria, and it is an enzyme acting within the tricarboxylic acid cycle, an aerobic pathway that yields energy to the cell (GO:0030060). Kim et al. (2013) have shown that ectopic expression of *MHD2* induced the acceleration of adipogenic differentiation, indicating that the acetylation of *MDH2* is very important for adipogenesis. Zhou et al. (2012) reported that leaner swine breeds show lower expression abundance of *MDH2*, and that, regardless the breed, female exhibits higher expression of this gene than males. This suggest that the fat metabolism may be a process underlying the conversion efficiency of dietary energy to energy available for body weight gain, as is expected for k_g. In addition, expression of *MDH2* presents a different pattern depending on the tissue (Zhou et al., 2012). Subcutaneous adipose tissue had higher expression abundance of *MDH2* than visceral adipose tissue. In fact, the calculation of k_g by the CVDS model considers not only the energy retained by the animal, but mainly, the composition of gain. Higher levels of fat deposition mean higher k_g values, which match the association between k_g and *MDH2*. This suggestive relation between k_g and fat metabolism as shown by the genomic analysis is interesting because its concept is adopted by the standard energy system of the NRC (2000), which drives many model developments and experiments in ruminant nutrition.

Available information on genes *POU6F2* and *SCLT1*, which were also significant for k_g, still not allow establishing any kind of association with cattle growth traits. The *POU6F2* and *SCLT1* are still uncharacterized for *Bos taurus*, however, in *Equus caballus* and *Ovis aries*, *PU6F2* has DNA-binding functions that regulate transcription (GO:0003700).

The *TMEM154* gene encodes a transmembrane protein that was found to be associated with AFSBW. The function of this protein has not been reported in any livestock species yet, however, in humans, variations in *TMEM154* has been found to have a negative role on the beta cell functioning and thus, on the level of insulin resistance and glucose. As reported in

Harder et al. (2015) *TMEM154* shows high mRNA expression in B lymphocytes and immunohistochemical stains of the human gastrointestinal tract show strong positivity of this protein in glandular cells of the digestive tract. As AFSBW is a parameter calculated from outcomes of animal performance, such as body composition at slaughter, it can be suggested that *TMEM154* may have some relationship to nutrient partitioning which in turn would result in AFSBW variation. Following the pathways of the antagonizing effect of the growth hormone on insulin action known for cattle (LUCY, 2008), depression of beta cells, as depicted by *TMEM154*, would benefit the development of lean tissue.

Table 3. Statistics and parameters (n = 893) estimated for the 10 most significant single-nucleotide polymorphisms (SNPs) for AFBSW (adjusted final shrunk body weight).

SNP ID	Chromosome	Position (bp)	β[a]	SE	p-value	Gene[b]
rs207966751	4	35125282	11.433	2.272	4.85×10^{-7}	-
rs135020999	4	35086804	14.2	2.938	1.34×10^{-6}	-
rs521816871	17	4982198	14.677	3.18	3.92×10^{-6}	TMEM154
rs109782726	16	34462941	-8.126	1.765	4.13×10^{-6}	SDCCAG8
rs42735715	16	34460742	7.737	1.691	4.76×10^{-6}	SDCCAG8
rs136058533	24	1124251	-14.404	3.155	4.97×10^{-6}	ATP9B
rs132812243	18	18597882	9.05	1.987	5.23×10^{-6}	HEATR3
rs137449097	24	1147611	-13.866	3.061	5.92×10^{-6}	ATP9B
rs137299855	21	70307103	-7.282	1.618	6.75×10^{-6}	-
rs517269657	4	77569249	9.018	2.014	7.57×10^{-6}	-

[a] β, allele substitution effect estimated with GWAS.
[b] Gene where SNP is located. All SNPs are *intron* variants. *TMEM154*, *Bos taurus* transmembrane protein 154; *SDCCAG8*, serologically defined colon cancer antigen 8; *ATP9B*, ATPase, class II, type 9B; *HEATR3*, HEAT repeat containing 3.

The *SDCCAG8* gene encodes a centrosome associated protein, involved in organizing the centrosome during interphase and mitosis (Reactome:5834576). Pathways within the mitotic cell cycle are expected to be associated with AFSBW. In humans, *SDCCAG8* seems to be relevant for BW regulation with high transcript abundance on the hypothalamus, pituitary and adrenals (SCHERAG et al., 2012), a hormonal axis that also plays an important role in cattle growth traits (PERKINS et al., 2014). Intronic variance in *SDCCAG8* has been associated with lower weight loss in overweight children and adolescents (Scherag et al.,

2012). If this pattern were sustained in beef cattle, AFSBW could be related to the animal's ability for BW regulation, which is in accordance with its interpretation as stated by the CVDS model.

Table 4. Statistics and parameters (n=893) estimated for the 10 most significant single-nucleotide polymorphisms (SNPs) for k_g (efficiency of dietary metabolizable energy to net energy for gain).

SNP ID	Chromosome	Position (bp)	β[a]	SE	p-value	Gene[b]
rs132652367	10	56197171	0.011	0.002	1.32×10^{-7}	UNC13C
rs719659528	7	59706678	-0.005	0.001	1.06×10^{-6}	-
rs134411628	24	32067866	-0.004	0.001	4.23×10^{-6}	-
rs109334860	4	8389419	0.003	0.001	5.63×10^{-6}	-
rs109346688	25	34770404	0.004	0.001	8.24×10^{-6}	MDH2
rs137152396	4	82460519	0.003	0.001	8.64×10^{-6}	POU6F2
rs133459104	17	29212172	0.009	0.002	8.93×10^{-6}	SCLT1
rs137006070	10	34301459	0.003	0.001	1.04×10^{-5}	-
rs132862266	13	59129338	0.005	0.001	1.35×10^{-5}	-
rs43634421	10	55970723	-0.004	0.001	1.38×10^{-5}	UNC13C

[a]β, allele substitution effect estimated with GWAS.
[b]Gene where SNP is located. All SNPs are *intron* variants. *UNC13C*, unc-13 homolog; *MDH2*, malate dehydrogenase 2, NAD (mitochondrial); *POU6F2*, POU class 6 homebox 2; *SCLT1*, sodium channel and clathrin linker 1.

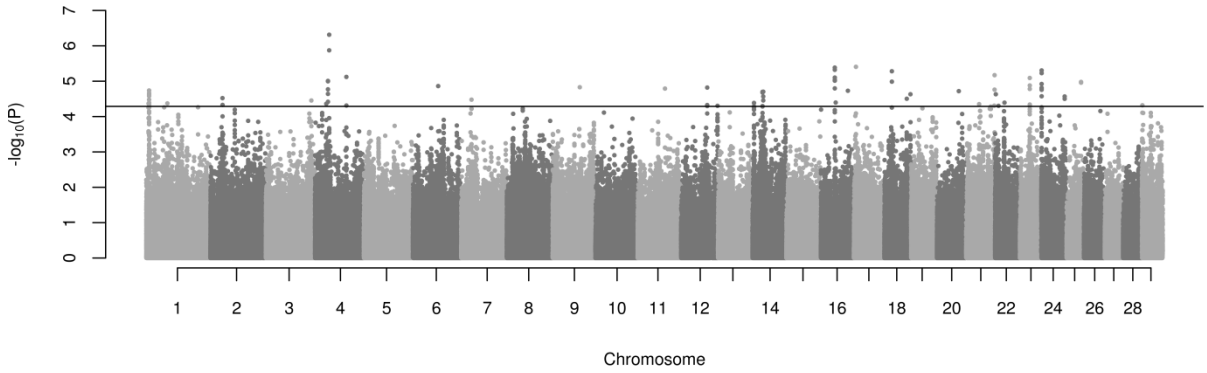
The *ATP9B* encodes a P4-ATPase that is important for translocating phospolipids from the exoplasmic to the cytoplasmic leaflet of lipid bilayers (TAKATSU et al., 2011). Information on *ATP9B* and *HEATR3*, which were also associated with AFSBW, still does not allow for an interpretation of their effects on cattle growth traits.

3.3.2 CVDS performance with genetically predicted AFSBW and k_g

To draw the genotype-to-parameter map and develop equations to predict AFSBW and k_g from genotypic data we used the 10 most significant SNPs, however, SNPs close to each other (within the 1-Mb range) were treated as one to account for the possibility of linkage disequilibrium between them. Prediction of AFSBW and k_g were of good precision. The joint model was able to explain 76% of the variation in AFSBW and 73% of k_g (Table 7). This

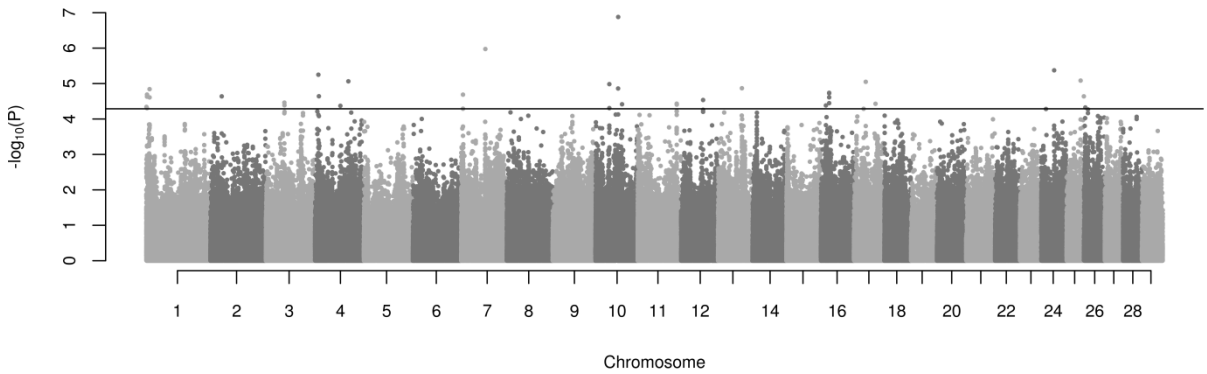
result suggests that it may be possible to derive values of mechanistic model parameters by using genomic information. Interestingly, the 30% lack of ability of the predictive model to compute AFSBW did not spread over the simulations. CVDS runs with AFSBW predicted by genotypes yield good predictions of the outcome variables DMI, ADG, and final EBF (Table 7).

Figure 1. Manhattan plot of $-\log_{10}$ (P-values) test for genome-wide association of AFSBW (adjusted final shrunk body weight) in *Bos indicus* cattle.



The horizontal line represents the significant threshold (5.15×10^{-5}). Source: by the authors.

Figure 2. Manhattan plot of $-\log_{10}$ (P-values) test for genome-wide association of k_g (of conversion of dietary metabolizable energy to net energy for growth) in *Bos indicus* cattle.



The horizontal line represents the significant threshold (5.15×10^{-5}). Source: by the authors.

Table 5. Quantitative trait loci (QTLs) associated with production traits within the 1-Mb region surrounding the most significant single-nucleotide polymorphisms (SNPs) for AFSBW (adjusted final shrunk body weight)

SNP ID	QTLs[a]
rs207966751	BW slaughter (15723), BW weaning (10708), milk fat % (5055), milk protein yield (10277), calving ease (4655)
rs521816871	rump angle (3448), BW birth (11038), BW yearling (11039), RFI (4448), post-weaning ADG (4484), milk yield (4743)
rs109782726	BW birth (11025), ADG (7101), BW weaning (4482), pre-weaning ADG (4486), BW weaning (11026), calving ease (1700)
rs136058533	RFI (5307), BW yearling (11184), calving ease (11181)
rs132812243	RFI (5293, 5294), BW mature (11061)

[a]BW, body weight; RFI, residual feed intake; ADG, average daily gain; number in parenthesis refers to the QTL ID at the Cattle QTL Database (HU et al., 2013).

Although high-level phenotypes when compared to parameters usually adopted in cGP models, AFSBW and k_g can be understood as component traits of more complex phenotypes. The biological interpretation of AFSBW and k_g as imposed by the CVDS model may be confirmed genetically, as the GWAS study and the surrounding QTLs have shown. SNPs found to be significant for the parameters revealed many associated QTLs and annotated genes that could be further explored to better describe biological pathways responsible for cattle growth and energetic efficiency. Specifically, our analyses have pointed that k_g may be related to fat metabolism. To this matter, we also suggest that CVDS could be improved by modeling in more detail the intersection among fat pools, energy balance and growth dynamics. As expected, AFSBW, a parameter close to the concept of BW at maturity, seems to be associated with mechanisms of cell cycle and mitosis.

Table 6. Quantitative trait loci (QTLs) associated with production traits within the 1-Mb region surrounding the most significant single-nucleotide polymorphisms (SNPs) for k_g (efficiency of dietary metabolizable energy to net energy for gain).

SNP ID	QTLs[a]
rs132652367	BW birth (10878)
rs719659528	BW birth (10810), BW weaning (10808), BW yearling (10807), milk yield (10291 / 2448), milk protein % (2532), milk protein yield (2449), calving ease (10811)
rs134411628	Feed conversion ratio (5308), BW weaning (11198), BW yearling (11197), calving ease (11196)
rs109346688	BW weaning (11213), height mature (11216), rum angle (1715), weaning weight-maternal milk (11215), milk protein yield (2610 / 10351), milk yield (2591 / 1538 / 10345), milk fat percentage (10342), calving ease (15224)
rs133459104	BW slaughter (11708), fatty acid
rs137006070	BW mature (10873), ADG (22836)

[a]BW, body weight; ADG, average daily gain; number in parenthesis refers to the QTL ID at the Cattle QTL Database (HU et al., 2013).

Table 7. Statistics of the coefficient of determination (R^2) from the genotype-to-parameter simulations (n=100)

Phenotypes[a]	Mean[b]	Quartiles[b]	
		2.5%	97.5%
AFSBW	0.757	0.727	0.775
k_g	0.734	0.709	0.760
DMI	0.973	0.962	0.980
ADG	0.850	0.783	0.887
EBF	0.984	0.972	0.989

[a]*AFSBW* (adjusted final shrunk body weight), and k_g (efficiency of conversion of dietary metabolizable energy to net energy for growth), R^2 refers to the prediction equation *y ~ genotypes*; *DMI* (dry matter intake), *ADG* (average daily gain), and *EBF* (empty body fat), R^2 refers to the comparison between CVDS runs with regular or genomically computed AFSBW.
[b]Mean and quartiles for 100 Monte Carlo runs.

3.4 Conclusion

Our study shows that the integration of genetics with nutritional models of cattle growth may have an application when it comes to the predictability of a phenotype. It is possible to draw a genotype-to-parameter map, and then simulate the phenotype by using genomic information to compute model core parameters. The integration of these two disciplines, genomics and system dynamics, opens up innumerous opportunities to advance towards not only the discovery of candidate genes for important phenotypes in the Animal Sciences, but also on the development theory of predictive modeling for decision support tools.

References

ALEXANDRE, P. A. et al. Bovine NR1I3 gene polymorphisms and its association with feed efficiency traits in Nellore cattle. **Meta Gene**, v. 2, n. 1, p. 206–217, 2014.

ARNOLD, R. N.; BENNETTT, G. L. Evaluation of Four Simulation Models of Cattle Growth and Body Composition : Part I Comparison and Characterization of the Models. **Agricultural Systems**, v. 35, p. 401–432, 1991.

AULCHENKO, Y. S. et al. GenABEL: An R library for genome-wide association analysis. **Bioinformatics**, v. 23, n. 10, p. 1294–1296, 2007.

AULCHENKO, Y. S. et al. Predicting human height by Victorian and genomic methods. **European journal of human genetics**, v. 17, n. 8, p. 1070–1075, 2009.

BAKER, M. J. et al. Using ultrasound measurements to predict body composition of yearling bulls. **Journal of Animal Science**, v. 84, n. 10, p. 2666–2672, 2006.

BOGARDUS, C. Missing heritability and GWAS utility. **Obesity**, v. 17, n. 2, p. 209–210, 2009.

CHIZZOTTI, M. L.; TEDESCHI, L. O.; VALADARES FILHO, S. C. A meta-analysis of energy and protein requirements for maintenance and growth of Nellore cattle. **Journal of Animal Science**, v. 86, n. 7, p. 1588–1597, 2008.

CHIZZOTTI, M.; PAULINO, P.; VALADARES FILHO, S. **Possibilidade do uso da ultrasonografia para estimaçao da composiçao corporal de bovinos**. Reunião Anual da Sociedade Brasileira de Zootecnia. **Anais**...Lavras, MG: 2008

CROFT, D. et al. The Reactome pathway knowledgebase. **Nucleic Acids Research**, v. 42, n. D1, p. 472–477, 2014.

DUTHIE, C. et al. Epistatic analysis of carcass characteristics in pigs reveals genomic interactions between quantitative trait loci attributable to additive and dominance genetic effects. **Journal of Animal Science**, v. 88, n. 7, p. 2219–2234, 2010.

FOX, D. G.; SNIFFEN, C. J.; CONNOR, J. D. O. Adjusting nutrient requeriments of beef cattle for animal and environemental variations. **Journal of Animal Science**, v. 66, n. 6, p. 1475–1495, 1988.

FRANCESCONI, M.; LEHNER, B. The effects of genetic variation on gene expression dynamics during development. **Nature**, v. 505, n. 7482, p. 208–11, 2014.

GAO, X.; STARMER, J.; MARTIN, E. R. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. **Genetic Epidemiology**, v. 32, n. 4, p. 361–369, 2008.

GJUVSLAND, A. B.; PLAHTE, E.; OMHOLT, S. W. Threshold-dominated regulation hides genetic variation in gene expression networks. **BMC systems biology**, v. 1, p. 57, 2007.

GOMES, R. C. et al. Protein synthesis and degradation gene SNPs related to feed intake, feed efficiency, growth, and ultrasound carcass traits in Nellore cattle. **Genetics and Molecular Research**, v. 12, n. 3, p. 2923–2936, 2013.

GUIROY, P. J. et al. Predicting individual feed requirements of cattle fed in groups. **Journal of Animal Science**, v. 79, n. 8, p. 1983–1995, 2001.

HARDER, M. N. et al. The Type 2 Diabetes Risk Allele of TMEM154-rs6813195 Associates with Decreased Beta Cell Function in a Study of 6,486 Danes. **Plos One**, v. 10, n. 3, p. e0120890, 2015.

HENDERSON, C. **Applications of linear models in animal breeding**. 3. ed. Guelph: University of Guelph, 1984.

HU, Z. L. et al. Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, v. 41, n. D1, p. 871–879, 2013.

HUNTLEY, R. P. et al. The GOA database: Gene Ontology annotation updates for 2015. **Nucleic Acids Research**, v. 43, n. D1, p. D1057–D1063, 2014.

KANEHISA, M. et al. Data, information, knowledge and principle: Back to metabolism in KEGG. **Nucleic Acids Research**, v. 42, n. D1, p. 199–205, 2014.

KARISA, B. et al. Candidate genes and single nucleotide polymorphisms associated with variation in residual feed intake in beef cattle. **Journal of Animal Science**, n. 2009, p. 3502–3513, 2013.

KIM, E. Y. et al. Acceleration of adipogenic differentiation via acetylation of malate dehydrogenase 2. **Biochemical and Biophysical Research Communications**, v. 441, n. 1, p. 77–82, 2013.

LUCY, M. C. Functional Differences in the Growth Hormone and Insulin-like Growth Factor Axis in Cattle and Pigs: Implications for Post-partum Nutrition and Reproduction. **Reproduction in Domestic Animals**, v. 43, n. SUPPL.2, p. 31–39, 2008.

MAHER, B. Personal genomes: The case of the missing heritability. **Nature**, v. 456, p. 18–21, 2008.

MANOLIO, T. A et al. Finding the missing heritability of complex diseases. **Nature**, v. 461, n. 7265, p. 747–753, 2009.

MARCONDES, M. et al. Energy requirements of Zebu beef cattle. In: VALADARES FILHO, S. et al. (Eds.). . **Nutrient Requirements of Zebu Beef Cattle BR-CORTE**. 2. ed. Viçosa, MG: Universidade Federal de Viçosa, 2010. p. 81–106.

NRC. **Nutrient Requirements of Beef Cattle**. 7. ed. Washington, D.C.: National Academy Press, 2000.

PENG, H. et al. MeDiA: Mean Distance Association and Its Applications in Nonlinear Gene Set Analysis. **Plos One**, v. 10, n. 4, p. e0124620, 2015.

PÉNICAUD, L. et al. The autonomic nervous system, adipose tissue plasticity, and energy balance. **Nutrition**, v. 16, n. 10, p. 903–908, 2000.

PERKINS, S. D. et al. Residual feed intake studies in Angus-sired cattle reveal a potential role for hypothalamic gene expression in regulating feed efficiency. **Journal of Animal Science**, v. 92, n. 2, p. 549–560, 2014.

R CORE TEAM. **R: A language and environment for statistical computing**Vienna, AustriaR Foundation for Statistical Computing, , 2015. Disponível em: <http://www.r-project.org>

SANTANA, M. H. A et al. Feed efficiency and its correlations with carcass traits measured by ultrasound in Nellore bulls. **Livestock Science**, v. 145, n. 1-3, p. 252–257, 2012.

SANTANA, M. H. A et al. Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. **BMC genetics**, v. 15, p. 21, 2014.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC genomics**, v. 15, n. 1, p. 478, 2014.

SCHERAG, A. et al. SDCCAG8 Obesity Alleles and Reduced Weight Loss After a Lifestyle Intervention in Overweight Children and Adolescents. **Obesity**, v. 20, n. 2, p. 466–470, 2012.

SUTTON, A. K. et al. Control of Food Intake and Energy Expenditure by Nos1 Neurons of the Paraventricular Hypothalamus. **Journal of Neuroscience**, v. 34, n. 46, p. 15306–15318, 2014.

SVISHCHEVA, G. R. et al. Rapid variance components–based method for whole-genome association analysis. **Nature Genetics**, v. 44, n. 10, p. 1166–1170, 2012.

TAKATSU, H. et al. ATP9B, a P4-ATPase (a putative aminophospholipid translocase), localizes to the trans-Golgi network in a CDC50 protein-independent manner. **Journal of Biological Chemistry**, v. 286, n. 44, p. 38159–38167, 2011.

TEDESCHI, L. O. et al. Energy requirement for maintenance and growth of Nellore bulls and steers fed high-forage diets. **Journal of Animal Science**, v. 80, n. 6, p. 1671–1682, 2002.

TEDESCHI, L. O.; FOX, D. G.; GUIROY, P. J. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. **Agricultural Systems**, v. 79, n. 2, p. 171–204, fev. 2004.

VENTURA, R. V. et al. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. **Journal of Animal Science**, v. 92, n. 4, p. 1433–1444, 2014.

VIK, J. O. et al. Genotype-phenotype map characteristics of an in silico heart cell. **Frontiers in Physiology**, v. 2 DEC, n. December, p. 1–12, 2011.

WANG, Y. et al. Parameters in dynamic models of complex traits are containers of missing heritability. **PLoS Computational Biology**, v. 8, n. 4, 2012.

ZHOU, S. L. et al. Differential expression analysis of porcine MDH1, MDH2 and ME1 genes in adipose tissues. **Genetics and Molecular Research**, v. 11, n. 2, p. 1254–1259, 2012.

Appendix – The Cornell Cattle Value Discovery System

List of equations used in the growth model[1]

$$EqSBW_j = \frac{(SRW \times SBW_j)}{AFSBW}$$

$$BFAF_j = 0.7714 + 0.00196 \times EqSBW_j - 0.00000371 \times EqSBW_j^2, \qquad EqSBW_j \geq 350\ kg$$

$$BFAF_j = 1, \qquad EqSBW_j < 350\ kg$$

$$DMI_{(j+1)} = \frac{SBW_j^{0.75} \times (0.2435 \times NEm_j - 0.0466 \times NEm_j^2 - 0.1128)}{NEm_j} \times BFAF_j,$$

$$age \leq 12\ mo$$

$$DMI_{(j+1)} = \frac{SBW_j^{0.75} \times (0.2435 \times NEm_j - 0.0466 \times NEm_j^2 - 0.0869)}{NEm_j} \times BFAF_j,$$

$$age > 12\ mo$$

$$PN_j = 0.8 + (BCS_j - 1) \times 0.05$$

$$NEmr_j = SBW_j^{0.75} \times a_1 \times PA_j \times PN_j$$

$$FFM_{(j+1)} = \frac{NEmr_j}{NEm_j}$$

$$RE_{(j+1)} = (DMI_{(j+1)} - FFM_{(j+1)}) \times NEg_j$$

$$FFG_{(j+1)} = DMI_{(j+1)} - FFM_{(j+1)}$$

$$EFG_{(j+1)} = FFG_{(j+1)} - NEg_{(j)}$$

$$SWG_{(j+1)} = 13.91 \times EqSBW_j^{-0.6837} \times EFG_{(j+1)}^{0.9116}, \qquad EFG_{(j+1)} > 0$$

$$SWG_{(j+1)} = 0, \qquad EFG_{(j+1)} \leq 0$$

$$SBW_{(j+1)} = SBW_j + SWG_j$$

$$EWG_{(j+1)} = SWG_{(j+1)} \times 0.956$$

$$REp_{(j+1)} = 0.0554 + 1.6939 \times Exp(-0.5573 \times \frac{RE_{(j+1)}}{EWG_{(j+1)}})$$

$$PIG_{(j+1)} = \frac{REp_{(j+1)} \times RE_{(j+1)}}{5.686}$$

$$k_{g(j)} = \frac{3}{4 + 11 \times REp_{(j+1)}}$$

$$AdjNEg_j = k_{g(j)} \times ME_{(j+1)}$$

[1] jth day, $j = 1, 2, \ldots, n$; n is the duration of the feeding trial.

List of variables used in the growth model

a_1 (Mcal $kg^{-0.75}$ day^{-1}), fasting heat production coefficient (0.072 for beef)

AdjNEg (Mcal kg^{-1}), diet NEg adjusted for body composition

AFSBW (kg), adjusted final shrunk body weight; is the SBW at 22% empty body fat (for *Bos indicus*)

BCS (dimensionless), body condition score

BFAF (dimensionless), body fat adjustment factor adjusts DMI for empty body fat content

DMI (kg day^{-1}), predicted dry matter intake

EBW (kg), empty body weight

EFG (Mcal day^{-1}), amount of net energy available for growth

EqSBW (kg), equivalent shrunk body weight

EWG (kg day^{-1}), empty weight gain

FFG (kg day^{-1}), feed intake available to support animal growth

FFM (kg day^{-1}), feed intake available to support animal requirement for maintenance

k_g (dimensionless), conversion efficiency of ME to NEg

ME (Mcal kg^{-1}), dietary content of metabolizable energy

NEg (Mcal kg^{-1}), dietary content of net energy for growth

NEm (Mcal kg^{-1}), dietary content of net energy for maintenance

NEmr (Mcal day^{-1}), animal requirement for net energy for maintenance

PA (dimensionless), physical activity factor

PIG (dimensionless), protein in gain

PN (dimensionless), previous nutrition; is an adjustment for NEm for this effect

RE (Mcal day^{-1}), retained energy

REp (dimensionless), retained energy as protein

SBW (kg), shrunk body weight

SRW (kg), shrunk reference weight

SWG (kg day^{-1}), shrunk weight gain; corresponds to average daily gain (ADG)

4 Using a system of differential equations that models cattle growth to uncover the genetic basis of complex traits

Abstract

The interplay between dynamic models of biological systems and genomics is based on the assumption that genetic variation of the complex trait (i.e. outcome of model behavior) arises from component traits (i.e. model parameters) in lower hierarchical levels. In order to provide a proof of concept of this statement for a cattle growth model we ask whether model parameters map genomic regions that harbors QTLs already described for the complex trait. We conducted a genome-wide association analysis (GWAS) with a Bayesian hierarchical lasso method in two parameters of the Davis Growth Model, a system of three ordinary differential equations describing DNA accretion, protein synthesis and degradation, and fat synthesis. Phenotypic and genotypic data were available for 893 Nellore (*Bos indicus*) cattle. Computed values for parameter k_1 (DNA accretion rate) ranged from 0.005 ± 0.003, and for α (constant for energy for maintenance requirement) 0.134 ± 0.024. The expected biological interpretation of the parameters is confirmed by QTLs mapped for k_1 and α. QTLs within genomic regions mapped for k_1 are expected to be correlated with the DNA pool: body size and weight. SNPs which were significant for α mapped QTLs that had already been associated with residual feed intake, feed conversion ratio, average daily gain, body weight, and also dry matter intake. SNPs identified for k_1 were able to additionally explain 2.2% of the phenotypic variability of the complex average daily again (ADG), even when SNPs for k_1 did not match to genomic regions associated with ADG. Although improvements are needed, our findings suggest that genomic analysis on component traits may help to uncover the genetic basis of more complex traits, particularly when lower biological hierarchies are mechanistically described by mathematical simulation models.

4.1 Introduction

The interplay between dynamic models for phenotypic prediction and genomics is based on the assumption that parameters of mechanistic models describing biological processes underlying the trait formation are themselves important phenotypes of the system. Vik et al. (2011) has coined the term *computational physiological genomics* for this discipline, which is emerging for human physiology (GJUVSLAND et al., 2013) but has been a very active research area for plant biologists (HAMMER et al., 2006; YIN; STRUIK; KROPFF, 2004). Theories of quantitative genetics are combined with mathematical descriptions of traits behavior for genotype-to-phenotype modeling. This approach is promising not only for uncovering the genetic basis of important traits but also for improving phenotypic prediction.

Results of *in silico* experiments of heart cell physiology (NORDBØ et al., 2015; VIK et al., 2011; WANG et al., 2012) and also findings from plant physiology have shown that model parameters may be understood as component traits of more complex phenotypes (i.e. outcomes of model behavior), and that they may show higher heritability than the analysis conducted on the complex phenotype directly. Under these assumptions, parameters should be measured experimentally instead of estimated from optimization routines.

In the Animal Sciences, nutritional models have been traditionally used by the academia and the industry for optimizing animal growth, balancing diets, and also for experimental design. Specifically for cattle growth models, many authors have suggested dynamic mechanistic descriptions of processes underlying weight gain and protein and fat deposition (HOCH; AGABRIEL, 2004; OLTJEN et al., 1986; TEDESCHI; FOX; GUIROY, 2004; WILLIAMS; JENKINS, 2003).

Here, we argue that computational physiological genomics may represent a potential new approach for individual-based modeling in animal nutrition. Current phenotyping tools and experimental protocols in ruminant nutrition still do not allow measuring parameters of dynamic mechanistic models of cattle growth directly, which means that they still need to be estimated by regular optimization procedures. However, in order to provide a proof of concept that this approach would be suitable for ruminant nutritional modeling we ask: is the mathematical interpretation of model parameters – a feature imposed by mathematical formalisms – in accordance with their associated genomic regions? As one of the fundamental concepts of the interplay between system dynamics and genomics is genetic variation in model core parameters, it is reasonable to think that such assumption implies that genomic

regions mapped for model parameters should match to those with quantitative trait loci (QTLs) already described for the complex trait.

Freua (unpublished)[1] has shown that for two parameters of the Cornell Cattle Value Discovery System (CVDS, TEDESCHI; FOX; GUIROY, 2004) this assumption holds. Both the adjusted final shrunk body weight, a parameter whose interpretation is close to the concept of mature body weight, and the efficiency of conversion of dietary metabolizable energy to energy for growth (k_g), map genomic regions whose nearby QTLs and genes resemble to their expected interpretation. In the case of k_g this is even more relevant, because it is adopted by the National Research Council (NRC) energy system, (2000) which is widely used for model development and research agendas on ruminant nutrition.

It happens that the CVDS is comprised of sequential equations based on the NRC system, which depicts a model with a high level of aggregation in terms of its modeled biological hierarchies. It is interesting to ask whether the biological validity of parameters, i.e. model parameters resemble complex traits, would also be evident in entities of lower level systems. Among cattle growth models available in the scientific literature, the Davis Growth Model (DGM) as described by Oltjen et al. (1986) provides the lowest aggregation level: a system of three ordinary differential equations describing DNA accretion, protein synthesis and degradation, and fat synthesis.

Therefore, we conducted a genome-wide association analysis (GWAS) with a Bayesian hierarchical lasso method in two parameters of the DGM for assessing QTLs nearby mapped genomic regions. Also, motivated by the expectation that SNPs significant for model parameters would explain additional phenotypic variation when compared to those mapped for the complex phenotype, we define a linear model whose predictors are genotypic probabilities and test what set of markers yield the best explained variance.

4.2 Material and methods

4.2.1 Ethics statement

No statement by the local animal ethics committee was necessary for this research. The study described herein used a dataset obtained from experiments published elsewhere (ALEXANDRE et al., 2014; GOMES et al., 2013; SANTANA et al., 2012, 2014), whose authors report phenotype records and DNA samples from tests approved by their respective

[1]FREUA, M.C. et al. Parameters of a dynamic mechanistic model of cattle growth retain enough biological interpretation for genotype to phenotype mapping. (unpublished).

institutional animal ethics committee and by the cattle owners. All animals' procedures were undertaken within common management practices in beef cattle feedlot operations and in accordance with standard veterinary protocols.

4.2.2 The DGM model

The DGM is based on the system of three ordinary differential equations developed by Oltjen et al. (1986) and has been widely used in beef cattle management. It was adapted for cattle from a previous growth model of male Sherman rats designed for theoretical applications by Baldwin and Black (1979). The main concept is that of DNA accretion over time, and the relation between whole body DNA and protein, which underlies the dynamics of protein synthesis and degradation. The DNA pool is structured with the DNA accretion rate (k_1), which essentially defines the rate of approximation of current amounts of DNA up to a maximum. This maximum has a correlation with mature body weight. The protein pool relies on protein synthesis, and this dynamics is dependent on DNA accretion and protein degradation.

In order to make DGM suitable as a decision-support tool for cattle feedlot operations, nutritional constraints that are obtained empirically were included in the model. This update placed the DGM within the net energy system adopted by the NRC (2000). Nutritional factors basically depend on the level of dietary metabolizable energy. To compute the daily maintenance energy requirement (EMr) the equation proposed by the NRC (2000) is used

$$EMr = \alpha \times EBW^{B} \qquad \text{eq.(1)}$$

Where α and *B* are constants, and EBW is empty body weight (kg). Fat deposition, the third state variable of the DGM, is defined from the residual dietary energy after maintenance and protein deposition. See Appendix for further details regarding the DGM.

4.2.3 Phenotypic and genotypic data

Phenotypic data for computing the DGM parameters was obtained from fourteen feeding trials conducted in feedlot operations in South (SANTANA et al., 2012), Southeast (ALEXANDRE et al., 2014; GOMES et al., 2013) and Central-West (SANTANA et al., 2014) Brazil. Animals were confined for a period of 73 ± 10.64 days. At feedlot entry animals averaged 550 ± 115 days of age and 357 ± 60.34 kg of live weight. Individual records of daily feed intake were taken by either GrowSafe or Calan Gates.

Genotypic data was available for 893 Nellore (*Bos indicus*) cattle. Genotyping was done as described in Freua (unpublished)[1] and Santana et al. (2014) using four different panels: Illumina BovineHD® BeadChip (777,962 SNPs), Affymetrix Axiom® Genome-Wide BOS1 Array (648,874 SNPs), GGP Indicus Neogen HD® (84,379 SNPs) and Illumina BovineSNP50® version 2 BeadChip (54,609 SNPs). Imputation of missing genotypes was performed with FImpute v2.2 (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014) using genotypic information of 3,776 Nellore cattle from a database of the Animal Breeding and Biotechnology group at FZEA/USP (http://usp.br/gmab/) following recommendations of Ventura et al. (2014). As standard practice, genotypic data was submitted to quality control and only autosomal SNPs with minor allele frequency > 2% and deviation from Hardy-Weinberg equilibrium (χ^2-test, 1 *df*, $P > 1 \times 10^{-5}$) were kept for subsequent analysis. After quality control, 941,033 SNPs remained in the dataset.

4.2.4 Computation of DGM parameters

For assessing the validity of the biological interpretation of DGM parameters, k_1 and α are used for genomic analysis. Their choice was not arbitrary and was dependent on the model structure and the available dataset. DGM outcomes are quite sensible to both parameters as demonstrated in Oltjen et al. (1986), and they are not correlated to the point to impair parameterization due to issues concerning parameter identifiability. k_1 and α are obtained with a Nelder-Mead optimization method. The FME package (SOETAERT; PETZOLDT, 2010) developed for the R statistical environment was used for this procedure. As observations for final protein and fat were not available, whole body composition at slaughter was predicted for all animals using ultrasound measurements and equations developed by Chizzotti, Paulino and Valadares Filho et al. (2008). These predictions were then treated as dependent variables.

4.2.5 Association analysis

We adopted a Bayesian hierarchical lasso method as described by Li et al. (2015) for carrying out the association test. Their statistical model is able to accommodate discrete or continuous covariates. In theory, k_1 and α should not need of any further adjustment due to its conceptual meaning – core parameters of a mechanistic model. When parameters are correlated to any external variable, then the model should be further developed to accommodate such relations endogenously. This type of investigation was beyond our goals for this manuscript, nonetheless no significant relationships between parameter values and potential fixed effects were observed when analyzing regression equations. For average daily

gain (ADG), however, the animal's age at slaughter was treated as a covariate in the model. The response y_i for subject i is decomposed as follows

$$y_i = \mu + \beta_i^T + X_i^T a + \varepsilon_i$$
$$\varepsilon_i \sim N(0;\ \sigma^2) \qquad \text{eq.(2)}$$

Where μ is the overall mean, β is the d_1-dimensional vector of continuous covariates for subject i, $c = (c_{1,\ldots,}c_{d1})^T$ is the vector of regression coefficients for continuous covariates, $a = (a_{1,\ldots,}a_p)^T$ is the p-dimensional vectors of the additive effect of SNPs. X_i is the indicator vector of the additive (1 = *AA*; 0 = *Aa*; -1 = *aa*) effect. ε_i is the residual error.

Solving eq.(2) takes several computational tasks. The difficulty of any GWAS procedure is the situation when the number of predictors far exceeds the sample size, and thus filtering out SNPs is recommended, otherwise selection power may be lowered as false positive rate increases. For taking this challenges into account, only SNPs with p value ≤ 0.1 for the trait were selected from the dataset by first performing GWAS with a single-SNP analysis described by Das et al. (2011). In a second step, pre-selected markers were then submitted to the Bayesian hierarchical lasso procedure defined for eq.(2). The lasso penalty is placed on the size of gene effects and encourages sparse solutions, which means that coefficients of lower importance are shrunk into zero. Additive gene effects were estimated by sampling from their conditional posterior distributions using 5000 Markov Chain Monte Caro iterations with a Gibbs sampler algorithm after convergence. For claiming significance of gene effects, the 95% posterior credible interval should not have contained zero.

As a final step, it was necessary to refit parameters by running the linear regression model (eq.(2)) again but without the penalty term. Only the selected SNPs were included in this final model. This approach has the potential to improve the bias of the parameter estimates introduced by lasso penalties. The posterior samples of the gene effects and the observed genotypes can be used to compute the proportion of the phenotypic variance explained by a particular SNP (m^2). Thus, it is reasonable to assume m^2 as a parameter to identify those significant SNPs that have the largest influence on the phenotype. For discussion, we report only SNPs claimed significant after the refit procedure. In addition, the 1-Mb region surrounding significant SNPs was investigated to determine whether they mapped against any previously described QTL or trait association deposited in the cattle QTLdb database (HU et al., 2013). All computations for the genomic association test were

done with the *gwas.lasso* package for R, which is available at http://www.psu.edu/dept/statgen/software.html.

4.2.6 Explained phenotypic variance of the complex trait

For assessing how much of the phenotypic variance of the complex phenotype would be explained by genomic regions mapped for the component trait we defined a linear model (AULCHENKO et al., 2009) whose response is ADG (a complex phenotype) adjusted for contemporary group and age at slaughter, and predictors are genotypic probabilities of significant SNPs from GWAS conducted directly on the ADG or the parameter k_1 (the DNA accretion rate influences ADG as depicted by the DGM equations). The coefficient of determination (R^2) from the linear regression was assumed to be a measure of the fraction of phenotypic variance explained by the genomic regions. From the SNPs obtained after the refit procedure, we only considered for the linear model those with $m^2 \geq 0.5$. Also, for avoiding the inclusion of potentially correlated SNPs into the model, we only considered for this analysis those that distance at least 1-Mb from one other.

4.3 Results and discussion

The discipline of computational physiological genomics integrates into the same framework concepts of quantitative genetics with mathematical descriptions of biological processes underlying the trait formation. Model parameters are recognized as important phenotypes of the biological system, and they may be understood as component traits of the more complex phenotype. The main assumption is that genetic variation of complex phenotypes may be captured by conducting analysis on the model parameters directly. This approach turns into a method for genotype-to-phenotype modeling. Provided that mechanistic models filter out environmental effects by simulating their influence on the phenotype, component traits are expected to demonstrate higher heritability (WANG et al., 2012).

What becomes difficult for applying this approach is that once parameters are important traits of the system, it becomes necessary to measure them experimentally, instead of estimating them. In plant physiology, this approach has led to the development of new phenotyping tools, such as new image recognition technologies (VAN DER HEIJDEN et al., 2012). However, in the Animal Sciences, and particularly for models of cattle growth dynamics, current phenotyping tools and experimental protocols still do not accommodate practices that allow observing parameters of mechanistic models directly. Measuring

parameters of growth models generally requires agendas such as those of comparative slaughters, which are very expensive and laborious. Moreover, genomic science requires a great number of samples for adequate inference of genomic regions associated with a phenotype. Thus, these requirements are barriers to the application of computational physiological genomics for livestock production.

For providing a proof of concept of the suitability of integrating dynamic mechanistic models of cattle growth with genomics, we estimated parameters of the DGM with a regular optimization routine, and assumed that if genetic variation of complex phenotypes penetrates model parameters, then genomic regions mapped for the parameters should match to those with QTLs already described for the complex trait. This is basically equivalent to asking whether mathematical interpretations of model parameters are in accordance with what their associated genomic regions can tell.

Methods that combine within the same statistical framework the theories of genomic analysis with systems of differential equations have already been proposed (FU et al., 2012). However, for the sake of simplicity, and considering that our goal was to contrast QTLs surrounding mapped genomic regions to the model formalism, we assumed the estimated DGM parameters as observed phenotypes. Therefore, parameters k_1 and α were response variables in the Bayesian hierarchical lasso method for GWAS reported by Li et al. (2015).

Computed values for parameter k_1 ranged from 0.005 ± 0.003, and for α 0.134 ± 0.024. These values are within the range reported for *Bos taurus* by Oltjen et al. (1986). Tables 1-3 describe parameters estimated after refitting the linear model (eq.(2)) with SNPs selected by the lasso procedure. The interpretation of the m^2 estimate should be done with caution because only a small set of SNPs remains for inclusion in the final analysis (HOTI; SILLANPÄÄ, 2006). Thus, the m^2 computation as defined by Li et al. (2011) is used only for the purpose of sorting significant SNPs by their relative importance to the phenotype.

The expected biological interpretation of k_1 and α is confirmed when searching for QTLs within their mapped genomic regions, as shown in Tables 4-5. k_1 is the DNA accretion rate, which basically defines the rate of change of the DNA pool, or as is considered by the DGM, k_1 defines the rate of whole-body cell maturation. QTLs described within the 1-Mb region of significant SNPs for k_1 represent traits that are in accordance with k_1 and the DNA pool: ADG, body weight (BW) at slaughter, and body size. k_1 clearly influences ADG, as implied by the DGM. It is interesting that feed efficiency traits, i.e. feed conversion ratio (FCR) and residual feed intake (RFI), were also associated with regions mapped for k_1. This suggests the interplay between protein turnover and energetic efficiency, which was

investigated by Gomes et al. (2013). The DGM attempts to capture this biological pathway. By taking FCR as an example, and assuming individuals under the same nutritional regime, variance in k_1 implies variation in protein synthesis and degradation, which in turn impact fat deposition and yield divergent ADG. In regard to the association between k_1 and average feed intake, the DGM still does not considers the DNA pool feedback on the animal's intake, once dry matter intake is not computed endogenously.

SNPs that were significant for α, a parameter that defines the energy for maintenance requirement, mapped QTLs associated with RFI, FCR, ADG, BW, and also dry matter intake. The relation between α and efficiency traits such as RFI and FCR was expected. Many authors have recognized that underlying between-animal variation in feed efficiency is variability in maintenance requirements (ARTHUR; HERD, 2008; GOMES et al., 2013; KENNEDY; WERF; MEUWISSEN, 1993; SOBRINHO et al., 2011).

Also, analyses of α computed for *Bos indicus* cattle over many generations have shown that selective breeding may have facilitated the correlation of this parameter with BW (Albertini, personal communication). Our results may confirm this correlation, once many of the SNPs that demonstrated higher m^2 for α have been mapped for QTLs associated with BW. This was expected considering the formalism of eq.(1). It also suggests that body size and weight may underlie variation in energy for maintenance requirement. Thus, interpreting α as an efficiency parameter should be done with caution and further analyses are needed to decompose correlations among α, measurements of feed efficiency and BW.

Motivated by the expectation that component traits would contribute to uncover the genetic basis of complex traits and that their genetic variation would reveal some of the unexplained phenotypic variance of the complex trait we defined a linear regression model in an attempt to predict the complex ADG from genotypic probabilities of SNPs obtained from GWAS with ADG directly or with parameter k_1. SNPs for ADG were able to explain 4.5% of the phenotypic variance and those for k_1 explained 2.2%. This suggests that analysis directly on the ADG would still yield better results than considering its component trait. However, k_1 is not expected to have a direct association with ADG. The DGM describes ADG from fat and protein deposition, and there may be an intermediate biological process between k_1 and the formation of ADG.

Despite the fact that SNPs of the component trait explained some variance of ADG, its mapped genomic regions did not match completely to those found by GWAS on ADG. Significant SNPs for ADG and k_1 were close only in chromosome 4. However, as discussed above, QTLs within regions for the component traits are in accordance with complex traits

expected to arise from them. This suggest that GWAS on the component traits may reveal additional genomic regions underlying the complex phenotype, which otherwise would not be captured by studies conducted on the complex phenotype alone. Thus, there may exists many behaviors happening in lower biological levels that may help animal geneticists to uncover the genetic basis of complex traits

Finally, it is important to highlight that some features of our dataset may have introduced bias in the DGM parameter estimates. This is mainly because actual observations of individual body composition at slaughter were not available and had to be predicted from ultrasound measurements. Thus, in order to carry out genomic analysis with parameters of dynamic models, nutritional experiments need to incorporate into their agenda accurate measurements of the component trait or of variables that are required for an adequate model parameterization. In this context, interdisciplinary research that increases collaboration between modelers and subject matter scientists is necessary to further development the discipline of computational physiological genomics.

4.4 Conclusion

The mathematical interpretation of k_1 and α parameters of the DGM is in accordance with what genomic information can tell, once both parameters map genomic regions that harbors QTLs for traits expected to arise from them. This suggests that genetic variance of complex traits penetrates DGM parameters. This concept should be further explored for enhancing the impact of genomic analysis, both in terms of the discovery of novel candidate genes for desirable traits and genomic prediction.

Table 1. Information about significant SNPs for k_1[1]

SNP ID	Chromosome	Position(bp)	*μ(a)*	*Min(a)*	*Max(a)*	m^2	MAF	L.Ratio	p-value
rs110433300	5	61747525	-0.0006	-0.0007	-0.0006	10.2927	0.02244	10.9138	0.00427
rs42992334	4	28217662	0.00066	0.00062	0.0007	10.2262	0.0212	14.2689	0.0008
rs133411985	5	61752090	0.00064	0.00059	0.00071	9.76792	0.02182	11.3186	0.00348
rs43382298	4	31838723	-0.0006	-0.0006	-0.0006	7.72292	0.0212	10.8094	0.0045
rs43263042	1	122809492	0.00042	0.00038	0.00046	4.83472	0.02494	13.4831	0.00118
rs135503726	23	6500304	0.00032	0.00029	0.00037	3.01892	0.02618	13.3342	0.00127
rs42992336	4	28215295	-0.0003	-0.0003	-0.0003	2.26728	0.02618	12.3233	0.00211
rs43263094	1	122797376	0.0003	0.00021	0.00035	2.10064	0.02182	10.4942	0.00526
rs133363275	8	108416061	-0.0001	-0.0001	-0.0001	0.4351	0.02244	11.994	0.00249
rs43263772	1	122750703	4.1×10^{-5}	2.7×10^{-5}	5.5×10^{-5}	0.04851	0.02618	6.7843	0.03364
rs136325769	23	6379901	2.5×10^{-5}	1.4×10^{-6}	4.8×10^{-5}	0.01797	0.02618	8.29227	0.01583

[1]Only SNPs claimed significant after the refit procedure are reported. k_1, DNA accretion rate. *μ(a)*, *Min(a)*, and *Max(a)* are the mean, minimum and maximum value of the estimated additive gene effect. m^2, proportion of phenotypic variance explained by the SNP. MAF, minor allele frequency. L.Ratio (likelihood ratio) and p-value refer to test statistics obtained for filtering out SNPs (see methods for further explanation). SNPs are ordered by m^2.

Table 2. Information about significant SNPs for α[1]

SNP ID	Chromosome	Position(bp)	*μ(a)*	*Min(a)*	*Max(a)*	m^2	MAF	L.Ratio	p-value
rs41589647	22	33766195	-0.0094	-0.0097	-0.0092	15.8049	0.02182	12.1992	0.00224
rs207906748	7	1983040	0.00791	0.00723	0.00867	12.0914	0.02369	11.515	0.00316
rs109464518	22	34972372	0.0077	0.00732	0.00808	10.2763	0.0212	10.6827	0.00479
rs136499429	21	42710811	-0.0054	-0.0062	-0.0049	5.43515	0.02244	15.3036	0.00048
rs110769696	7	1983458	-0.0038	-0.0046	-0.0032	2.86883	0.02431	11.26	0.00359
rs133789042	21	35367854	-0.0035	-0.0041	-0.003	2.40184	0.02369	14.3295	0.00077
rs43494233	7	1950138	0.00362	0.00317	0.00397	2.21369	0.02057	14.9854	0.00056
rs471481507	7	813448	0.00283	0.00245	0.00315	1.43455	0.02182	11.8727	0.00264
rs41979467	21	42707556	0.00241	0.00192	0.00314	1.18298	0.02494	14.7053	0.00064
rs41981674	21	40396288	-0.0023	-0.0026	-0.0021	0.87778	0.01995	12.2249	0.00222
rs41723543	14	7948988	0.00204	0.00184	0.00221	0.84656	0.02494	14.4295	0.00074
rs41979457	21	42739505	0.0014	0.00088	0.00191	0.45745	0.02868	11.1573	0.00378
rs135317125	21	35389547	0.0011	0.0006	0.00164	0.21589	0.02182	14.606	0.00067
rs43496317	7	2595559	0.001	0.00066	0.00127	0.18907	0.02307	10.7111	0.00472
rs41641673	21	35564875	0.00081	0.00051	0.00107	0.11791	0.02182	6.25241	0.04388
rs133785276	22	33716176	0.00065	0.00031	0.00089	0.06909	0.01995	8.65487	0.0132
rs109487288	21	35255308	0.00059	0.00026	0.0009	0.06026	0.0212	9.63412	0.00809

[1]Only SNPs claimed significant after the refit procedure are reported. α, coefficient of energy for maintenance requirement. *μ(a)*, *Min(a)*, and *Max(a)* are the mean, minimum and maximum value of the estimated additive gene effect. m^2, proportion of phenotypic variance explained by the SNP. MAF, minor allele frequency. L.Ratio (likelihood ratio) and p-value refer to test statistics obtained for filtering out SNPs (see methods for further explanation). SNPs are ordered by m^2.

Table 2. Information about significant SNPs for α[1] (continued)

SNP ID	Chromosome	Position(bp)	*μ(a)*	*Min(a)*	*Max(a)*	m^2	MAF	L.Ratio	p-value
rs109487288	21	35255308	0.00059	0.00026	0.0009	0.06026	0.0212	9.63412	0.00809
rs42576032	21	44054206	0.00045	0.00026	0.00062	0.05114	0.0318	5.27703	0.07147
rs109921315	14	7725751	0.00029	0.00018	0.00039	0.0409	0.06359	27.9864	8.4×10^{-7}
rs208446911	21	36348999	0.00029	0.00017	0.00041	0.03937	0.0586	6.52055	0.03838
rs41658739	7	2565484	0.00028	7.5E-05	0.00049	0.02169	0.03304	7.27203	0.02636
rs109381135	22	34952144	-0.0002	-0.0003	-7×10^{-5}	0.01776	0.08666	11.8635	0.00265
rs133767876	21	37720913	0.00011	4.9E-05	0.00018	0.01741	0.21322	7.56075	0.02281
rs135657217	21	44588851	-0.0001	-0.0002	-4×10^{-5}	0.01589	0.18142	6.04778	0.04861
rs137012373	22	33697308	8.4×10^{-5}	7.7×10^{-7}	0.00017	0.00699	0.1384	4.94925	0.08419
rs134731293	21	40719161	-9×10^{-5}	-0.0002	-1×10^{-7}	0.00627	0.09352	6.60157	0.03685
rs134976779	21	31949409	9×10^{-5}	1.8×10^{-6}	0.00017	0.00572	0.09352	9.14712	0.01032

[1]Only SNPs claimed significant after the refit procedure are reported. α, coefficient of energy for maintenance requirement. *μ(a)*, *Min(a)*, and *Max(a)* are the mean, minimum and maximum value of the estimated additive gene effect. m^2, proportion of phenotypic variance explained by the SNP. MAF, minor allele frequency. L.Ratio (likelihood ratio) and p-value refer to test statistics obtained for filtering out SNPs (see methods for further explanation). SNPs are ordered by m^2.

Table 3. Information about significant SNPs for average daily gain[1]

SNP ID	Chromosome	Position(bp)	*μ(a)*	*Min(a)*	*Max(a)*	m^2	MAF	L.Ratio	p-value
rs110766260	10	93092171	0.10104	0.09296	0.11045	14.4574	0.02941	13.8887	0.00096
rs110668288	11	26603976	0.10524	0.09607	0.1174	14.0623	0.02628	12.6711	0.00177
rs109002831	10	93072865	0.06483	0.05465	0.07257	6.32064	0.03129	14.3284	0.00077
rs109683301	4	23172054	-0.0405	-0.0481	-0.0318	2.08138	0.02628	12.5692	0.00186
rs379649035	4	24044327	-0.037	-0.0436	-0.03	2.01366	0.03066	14.2315	0.00081
rs208819295	4	25721703	0.02563	0.01718	0.03178	0.93058	0.02941	15.0965	0.00053
rs209564777	4	24103163	0.02497	0.01134	0.03493	0.75469	0.02503	10.8498	0.00441
rs109087529	4	24646545	-0.0206	-0.0426	-0.0099	0.55084	0.02691	10.1659	0.0062
rs42724571	10	93151803	-0.0145	-0.0187	-0.0107	0.38802	0.0388	6.89239	0.03187
rs135822287	10	93111107	-0.0155	-0.0224	-0.0087	0.36744	0.03191	14.7964	0.00061
rs41255705	26	21138509	0.01352	0.01066	0.01629	0.26432	0.03004	17.7569	0.00014
rs42604400	4	24020113	0.0125	0.00497	0.02102	0.22124	0.02941	14.464	0.00072
rs42974945	1	59605374	0.01268	0.00256	0.02825	0.17078	0.0219	10.2509	0.00594
rs109580444	23	17725570	0.00852	0.00609	0.01097	0.13	0.03755	6.55036	0.03781

[1]Only SNPs claimed significant after the refit procedure are reported. *μ(a)*, *Min(a)*, and *Max(a)* are the mean, minimum and maximum value of the estimated additive gene effect. m^2, proportion of phenotypic variance explained by the SNP. MAF, minor allele frequency. L.Ratio (likelihood ratio) and p-value refer to test statistics obtained for filtering out SNPs (see methods for further explanation). SNPs are ordered by m^2.

Table 3. Information about significant SNPs for average daily gain[1] (continued)

SNP ID	Chromosome	Position(bp)	*μ(a)*	*Min(a)*	*Max(a)*	m^2	MAF	L.Ratio	p-value
rs43376623	4	24054845	0.00932	0.00039	0.0196	0.12045	0.02879	13.5148	0.00116
rs110261974	4	23127632	-0.0087	-0.0159	-0.001	0.11172	0.03066	11.5318	0.00313
rs135817840	4	23376985	0.00803	0.00339	0.01286	0.10992	0.03567	12.181	0.00226
rs207871286	4	25860516	-0.0036	-0.0062	-0.0011	0.03832	0.06258	7.43393	0.02431
rs210028830	4	23710981	-0.0025	-0.0042	-0.0009	0.03169	0.11076	5.2811	0.07132

[1]Only SNPs claimed significant after the refit procedure are reported. *μ(a)*, *Min(a)*, and *Max(a)* are the mean, minimum and maximum value of the estimated additive gene effect. m^2, proportion of phenotypic variance explained by the SNP. MAF, minor allele frequency. L.Ratio (likelihood ratio) and p-value refer to test statistics obtained for filtering out SNPs (see methods for further explanation). SNPs are ordered by m^2.

Table 4. QTLs and associations with production traits by chromosome for k_1[1]

chr	QTL and association with production traits
4	Average daily feed intake, average daily gain, residual feed intake, feed conversion ratio, body length, chest girth, BW (slaughter), BW (weaning), height (mature), height (yearling), BW (yearling), weaning weight-maternal milk
5	BW (slaughter), BW (mature), height (mature), chest width, rump angle, rump length, rump width, chest depth, hip height
23	Body depth, BW (slaughter), height (mature), height (yearling)

[1]k1, DNA accretion rate. chr, chromosome.

Table 5. QTLs and associations with production traits by chromosome for α[1]

chr	QTLs and association with production traits
7	Dry matter intake, weaning weight-maternal milk
14	Average daily gain, BW (weaning), BW (birth), BW (mean)
21	feed conversion ratio, residual feed intake, BW (birth), BW (yearling), height (mature), height (yearling), weaning weight-maternal milk
22	rump angle, BW (yearling), height (mature), average daily gain

[1]α, coefficient of energy for maintenance requirement. chr, chromosome.

References

ALEXANDRE, P. A. et al. Bovine NR1I3 gene polymorphisms and its association with feed efficiency traits in Nellore cattle. **Meta Gene**, v. 2, n. 1, p. 206–217, 2014.

ARTHUR, J. P. F.; HERD, R. M. Residual feed intake in beef cattle. **Revista Brasileira de Zootecnia**, v. 37, n. spe, p. 269–279, 2008.

AULCHENKO, Y. S. et al. Predicting human height by Victorian and genomic methods. **European journal of human genetics**, v. 17, n. 8, p. 1070–1075, 2009.

BALDWIN, R. L.; BLACK, J. L. Simulation of the effects of nutritional and physiological status on the growth of mammalian tissues: description and evaluation of a computer program. In: **Technical Paper. Animal Research Laboratories**. [s.l.] CSIRO Australia, 1979.

CHIZZOTTI, M.; PAULINO, P.; VALADARES FILHO, S. Possibilidade do uso da ultrassonografia para estimação da composiçao corporal de bovinos. **Anais**...Lavras, MG: 2008

DAS, K. et al. A dynamic model for genome-wide association studies. **Human Genetics**, v. 129, n. 6, p. 629–639, 2011.

FU, G. et al. A dynamic model for functional mapping of biological rhythms. **J Biol Dyn**, v. 29, n. 6, p. 997–1003, 2012.

GJUVSLAND, A. B. et al. Bridging the genotype–phenotype gap: what does it take? **The Journal of Physiology**, v. 591, n. 8, p. 2055–2066, 2013.

GOMES, R. C. et al. Protein synthesis and degradation gene SNPs related to feed intake, feed efficiency, growth, and ultrasound carcass traits in Nellore cattle. **Genetics and Molecular Research**, v. 12, n. 3, p. 2923–2936, 2013.

HAMMER, G. et al. Models for navigating biological complexity in breeding improved crop plants. **Trends in Plant Science**, v. 11, n. 12, p. 587–593, 2006.

HOCH, T.; AGABRIEL, J. A mechanistic dynamic model to estimate beef cattle growth and body composition: 1. Model description. **Agricultural Systems**, v. 81, n. 1, p. 1–15, 2004.

HOTI, F.; SILLANPÄÄ, M. J. Bayesian mapping of genotype x expression interactions in quantitative and qualitative traits. **Heredity**, v. 97, n. 1, p. 4–18, 2006.

HU, Z. L. et al. Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, v. 41, n. D1, p. 871–879, 2013.

KENNEDY, B. W.; WERF, J. H. . VAND DER; MEUWISSEN, T. Genetic and statistical properties of residual feed intake. **Journal of Animal Science**, v. 71, n. 1991, p. 3239–3250, 1993.

LI, J. et al. Bayesian group Lasso for nonparametric varying-coefficient models with application to functional genome-wide association studies. **Annals of Applied Statistics**, v. 9, n. 2, p. 640–664, 2015.

NORDBØ, Ø. et al. Towards causally cohesive genotype–phenotype modelling for characterization of the soft-tissue mechanics of the heart in normal and pathological geometries. **Journal of The Royal Society Interface**, v. 12, n. 106, 1 abr. 2015.

NRC. **Nutrient Requirements of Beef Cattle**. 7. ed. Washington, D.C.: National Academy Press, 2000.

OLTJEN, J. . W. . et al. Development of a Dynamic Model of Beef Cattle Growth and Composition. **Journal of Animal Science**, v. 62, p. 86–97, 1986.

SANTANA, M. H. A et al. Feed efficiency and its correlations with carcass traits measured by ultrasound in Nellore bulls. **Livestock Science**, v. 145, n. 1-3, p. 252–257, 2012.

SANTANA, M. H. A et al. Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. **BMC genetics**, v. 15, p. 21, 2014.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC genomics**, v. 15, n. 1, p. 478, 2014.

SOBRINHO, T. L. et al. Residual feed intake and relationships with performance of Nellore cattle selected for post weaning weight. **Revista Brasileira de Zootecnia**, v. 40, n. 4, p. 929–937, 2011.

SOETAERT, K.; PETZOLDT, T. Inverse Modelling , Sensitivity and Monte Carlo Analysis in R Using Package FME. **Journal of statistical software**, v. 33, n. 3, p. 1–28, 2010.

TEDESCHI, L. O.; FOX, D. G.; GUIROY, P. J. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. **Agricultural Systems**, v. 79, n. 2, p. 171–204, fev. 2004.

VAN DER HEIJDEN, G. et al. SPICY: Towards automated phenotyping of large pepper plants in the greenhouse. **Functional Plant Biology**, v. 39, n. 11, p. 870–877, 2012.

VENTURA, R. V. et al. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. **Journal of Animal Science**, v. 92, n. 4, p. 1433–1444, 2014.

VIK, J. O. et al. Genotype-phenotype map characteristics of an in silico heart cell. **Frontiers in Physiology**, v. 2 DEC, n. December, p. 1–12, 2011.

WANG, Y. et al. Parameters in dynamic models of complex traits are containers of missing heritability. **PLoS Computational Biology**, v. 8, n. 4, 2012.

WILLIAMS, C. B.; JENKINS, T. G. A dynamic model of metabolizable energy utilization in growing and mature cattle . II . Metabolizable energy utilization for gain. **Journal of Animal Science**, p. 1382–1389, 2003.

YIN, X.; STRUIK, P. C.; KROPFF, M. J. Role of crop physiology in predicting gene-to-phenotype relationships. **Trends in Plant Science**, v. 9, n. 9, p. 426–432, 2004.

Appendix– The Davis Growth Model

1. Differential of DNA accretion

$$\frac{dDNA}{dt} = k_1 \times \lambda^{-0.27} \times \left[DNA_{mx} - DNA_{(t)}\right] \times \theta_1 ,$$

$$\lambda = \frac{EBWM}{rEBWM}, \qquad DNA_{mx} = rDNA_{mx} \times \lambda, \qquad \theta_1 = 1.70 \times \delta,$$

$$\delta = \frac{DMI_{(t)} \times ME}{\left(0.438 - 0.2615 \times \frac{EBW}{EBWM}\right) \times EBW^{0.73}}$$

Where $\boldsymbol{k_1}$ is a constant, $\boldsymbol{\lambda}$ is a transformation to account for the effect of frame size, $\boldsymbol{DNA_{mx}}$ is mature DNA, $\boldsymbol{rDNA_{mx}}$ is a reference DNA_{mx}, which was assumed to be 385 g, $\boldsymbol{\theta_1}$ is the adjustment for the effect of energy intake, ***EBW*** is empty body weight, ***EBWM*** is mature EBW, ***rEBWM*** is a reference *EBWM*, which was assumed to be 750 kg, $\boldsymbol{\delta}$ is the proportion of dietary metabolizable energy (***ME***) in dry matter intake (***DMI***) in relation to a reference pattern for ME intake.

2. Differential of protein synthesis and degradation

$$\frac{dProt}{dt} = k_2 \times DNA_{(t)}^{0.73} \times \theta_2 - k_3 \times Prot_{(t)}^{0.73},$$

$$\theta_2 = \frac{0.83 + 0.20 \times \delta}{0.15 + \delta}$$

Where $\boldsymbol{k_2}$ and $\boldsymbol{k_3}$ are constants, and $\boldsymbol{\theta_2}$ is the adjustment for the effect of energy intake.

3. Differential of fat synthesis

$$\frac{dFat}{dt} = \frac{\left[DMI_{(t)} - \frac{(\alpha \times EBW^{0.75})}{NE_m}\right] \times \left[NE_g - Prot_{(t)} \times 5.54\right]}{9.39}$$

Where $\boldsymbol{\alpha}$ is a constant, $\boldsymbol{NE_m}$ and $\boldsymbol{NE_g}$ are net energy concentrations in the feed (Mcal/kg) for maintenance and gain, respectively. Protein and fat gain contains 5.54 and 9.39 Mcal/kg, respectively.

5 A new approach for applied nutritional models: computing parameters of dynamic mechanistic growth models using genome-wide prediction

Abstract

Nutritional models have long been used as decision support tools by the livestock industry. Despite the advance of genomic prediction, these two disciplines have evolved separately. Because model parameters are responsible to describe between-animal variability, we propose an integration of nutritional models with genomics by means of such parameters. Two dynamic mechanistic models of cattle growth were used: Cornell Cattle Value Discovery System (CVDS) and Davis Growth Model (DGM). We estimated SNP marker effects for their parameters and also for observed phenotypes. Then, we compared what would be the best prediction scenario – model simulation with parameters computed from genomic data or genomic prediction directly on higher phenotypes. We found that genomic prediction on dry matter intake (DMI) and average daily gain (ADG) are still a better approach than using CVDS for predictions. Dry matter required (DMR), a CVDS-predicted value for DMI had higher correlation (r = 0.253) with observed DMI than results from genomic prediction (r = 0.07). DGM had better predictive ability (r = 0.38) than genomic prediction on ADG (r = 0.098). This is also the case for whole-body protein (r = 0.496) and fat at slaughter (r = 0.505) whose predictions were better with DGM than genomic prediction performed on the observed traits (r = 0.194 and r = 0.183, respectively). When contrasting simulations with genomically predicted parameters to those with regularly computed ones, CVDS showed moderate correlation and low bias between simulations of DMR (r = 0.966; b = 0.9%) and ADG (r = 0.645; b = 5.5%). Although further model development is necessary, the DGM with subject-specific parameters computed from genotypic data was a better option for predicting phenotypes than genomic prediction alone. In addition, simulations with genomically and regularly computed parameters match at a reasonable extend. This is the main argument to call attention from the research community that our approach may pave the way for the development of a new generation of applied nutritional models, especially towards individual-based simulations with subject-specific parameters computed from genomic information.

5.1 Introduction

Feed efficiency (FE) has been extensively studied by animal scientists due to its potential to enhance profitability in the livestock industry and also because feeding more efficient animals is a sound strategy to mitigate emissions of greenhouse gases (NKRUMAH et al., 2006; WHITE; CAPPER, 2013). For beef cattle, based on the concept of residual feed intake as a proxy for FE, feeding trials to record individual animal intake and growth performance have been conducted (BERRY; CROWLEY, 2013; HERD; ARCHER; ARTHUR, 2006). It was then hypothesized that nutritional mathematical models describing cattle growth could assist efforts to measure FE, appearing as a reasonable and lower cost approach, provided that equations would be able to capture between-animal variability in phenotypes related to animal growth (TEDESCHI et al., 2006; WILLIAMS, 2010).

To date, nutritional systems such as those adopted by the National Research Council (NRC), (2000) still rely on population-based parameters. Between-animal variability is only possible to be captured by models when there is parameter variance. Therefore, computing subject-specific parameters facilitates simulations of between-animal variability.

Previous researches have shown the biological validity of model predictions in terms of heritability and genetic correlations with observed phenotypes (KIRSCHTEN; POLLAK; FOX, 2007). Also, there is the possibility to compute parameters directly from genomic data, a potential that would facilitate the development of a new generation of nutritional models, depicted by the integration between system dynamics and genomics. Strategies towards this integration have been proposed using causally-cohesive phenotype-to-genotype (cGP) modeling as a method in the discipline of computational physiological genomics, which aims to incorporate computational physiology into genetics theory (VIK et al., 2011; WILLIAMS et al., 2006).

Many growth models may serve as a standard for exploring the possibility of obtaining nutritional model predictions with parameters computed from genomic information. Two of those available on the scientific literature offer an interesting contrast when it comes to the aggregation level: the Cornell Cattle Value Discovery System (CVDS) is used in commercial feedlot operations in the United States, and is comprised of sequential equations from the NRC (TEDESCHI; FOX, GUIROY, 2004), and the Davis Growth Model (DGM), which refers to the differential equations developed by Oltjen et al. (1986), describing DNA accretion, protein synthesis and degradation, and fat synthesis.

We combined a phenotypic and a genotypic dataset of Nellore (*Bos indicus*) cattle under feedlot operations to assess accuracy and precision of CVDS and DGM outcomes when simulations are performed with parameters computed from genomic data. We began by regularly deriving parameters from observed data. Then, we conducted a genome-wide prediction (GWP) analysis with a ridge regression procedure on both the parameters and observed phenotypes. We asked what would be the best prediction scenario: i) GWP on observed phenotypes, or ii) model predictions with individual parameters computed from GWP. Model predicted variables were contrasted to both observed values and to those values obtained from simulations conducted with parameters computed from observed data.

Our goal is to discuss the possibilities of using technologies from animal genomics to support the development of a new generation of applied nutritional models. Specifically, in the context of the CVDS model, which was designed to compute individual dry matter intake (DMI) from observed animal performance, this work contributes suggesting directions to novel approaches for research on FE and how it could benefit from its interplay with mathematical modeling.

5.2 Material and methods

5.2.1 Ethics statement

No statement by the local animal ethics committee was necessary for this research. The study described herein used a dataset obtained from experiments published elsewhere (ALEXANDRE et al., 2014; GOMES et al, 2013; SANTANA et al., 2012, 2014), whose authors report phenotype records and DNA samples from tests approved by their respective institutional animal ethics committee and by the cattle owners. All animals' procedures were undertaken within common management practices in beef cattle feedlot operations and in accordance with standard veterinary protocols.

5.2.2 The CVDS model

The CVDS model is comprised of sequential equations based on the NRC (2000) system. Animal characteristics, diet and environmental information are used to simulate daily individual DMI, body weight and carcass composition. The main parameter entering the CVDS model is the adjusted final shrunk body weight (AFSBW), which drives dynamics of retained energy. An exponential decay equation described by Tedeschi, Fox and Guiroy

(2004) is used to adjust the conversion efficiency of dietary metabolizable energy into net energy for gain (k_g) at each time step according to the composition of gain.

The CVDS model computes a reverse simulation. Observed average daily gain (ADG) and diet characteristics may enter the model to predict DMI using a backward calculation technique. In this case, DMI is conveniently named dry matter required (DMR), in a reference to the expected intake that would be necessary for the animal to meet observed performance.

The CVDS model was written in Visual Basic 6 and it is available at http://nutritionmodels.com/cvds.html. All the simulations described herein were performed without the environmental submodel and using the exponential decay adjustment for k_g.

5.2.3 The DGM model

The basic model underlying the DGM is the system of three ordinary differential equations described by Oltjen et al. (1986). Protein, fat and DNA pools are simulated daily. DGM relies on the accretion of whole body protein, and the DNA accretion rate (k_1) and DNA at maturity will shape growth trajectory and animal size. Protein dynamics depends on protein synthesis, which is conditioned to DNA at a given time point. Effects of energy intake on animal growth are incorporated into the model by terms that adjust DNA accretion and protein synthesis, and are obtained empirically. An equation that is also based on the net energy system of NRC (2000) drives residual energy (i.e. not used for maintenance and protein synthesis) towards fat deposition. Maintenance energy requirement is computed with an equation similar to that used by the NRC (2000), whose parameter (α) is important for modeling between-animal variation in protein and fat accretion.

5.2.4 The phenotypic and genotypic dataset

Observations of Nellore (*Bos indicus*) cattle (n = 893) were obtained from fourteen feeding trials conducted in feedlot operations in South (SANTANA et al., 2014), Southeast (SANTANA et al., 2012; GOMES et al., 2013) and Central-West (ALEXANDRE et al., 2014) Brazil. These experiments were carried out in two different facility types (Calan Gates or GrowSafe) to record individual daily feed intake with a length of 73 ± 10.64 days and regular weighing on 21 days. At feedlot entry animals averaged 550 ± 115 days of age and 357 ± 60.34 kg of live weight. ADG was estimated as the slope of a linear regression of weights on testing days. Phenotypic data was tested for normality with the Shapiro-Wilk test, and individuals that were more than twice the interquartile range above the third quartile or below the first quartile were excluded from the dataset.

Genomic DNA of these animals was extracted from either blood samples or hairs pulled from the skin. Samples were obtained from the feeding trials, and DNA was prepared as described by those authors. Imputation from Illumina BovineHD® BeadChip (777,962 SNPs), Affymetrix Axiom® Genome-Wide BOS1 Array (648,874 SNPs), GGP Indicus Neogen HD® (84,379 SNPs) and Illumina BovineSNP50® version 2 BeadChip (54,609 SNPs) to a super-dense panel (1,261,128 SNPs) was performed as described by Santana et al. (2014) and Ventura et al. (2014). As standard practice, genotypic data was submitted to quality control and only autosomal SNPs with minor allele frequency > 2% and deviation from Hardy-Weinberg equilibrium (χ^2-test, 1 *df*, $P > 1 \times 10^{-5}$) were kept for subsequent analysis. In the end, individuals had information on 941,033 SNPs.

5.2.5 Computation of model parameters

The AFSBW parameter of the CVDS model was computed from observed data using equations developed by Guiroy et al. (2001) and Baker et al. (2006), and adapted for *Bos indicus* cattle by Freua et al. (unpublished)[1]. For the DGM, parameters k_1 and α are obtained by fitting the model to the data with a Nelder-Mead method, which is able to optimize nonlinear and multidimensional problems without the need of deriving the objective function to find the solution (PRESS et al., 1990). The method uses an initial simplex, a set of vectors (points) in an *M*-dimensional space, where each vertex represents a possible solution. The number of simplex vertices equals the number of parameters to be fitted, plus one. For the scaling coefficients of the Nelder-Mead method, i.e. the reflection factor (θ), the contraction factor (λ) and expansion factor (γ), we adopted the values suggested by Press et al. (1990), which were 1.0, 0.5 and 2.0, respectively. The tolerance limit was 10^{-8} and the maximum number of iterations was set as 500. The FME package (SOETAERT; PETZOLDT, 2010) developed for R was used for this procedure. To compute the numerical solution of initial value problems for the system of ordinary differential equations in the DGM, the function *ode*(.) with *lsoda* integrator method from the *deSolve* package for R was used (SOETAERT, PETZOLDT, SETZER, 2010). The absolute and relative error tolerance was taken as 10^{-6}.

Parameterization of the DGM requires protein (fProt) and fat (fFat) at slaughter. As observations were not available, body composition was predicted for all animals using ultrasound measurements and equations developed by Chizzotti, Paulino and Valadares Filho et al. (2008). These predictions were then treated as dependent variables.

[1]FREUA, M.C. et al. Parameters of a dynamic mechanistic model of cattle growth retain enough biological interpretation for genotype to phenotype mapping. (unpublished).

5.2.6 Genomic prediction

Several statistical frameworks for genomic prediction have been discussed in the literature. The ridge regression BLUP (rrBLUP) approach was chosen because i) it was straightforward to implement together with CVDS and DGM, and ii) it is equivalent to the basic concept of GBLUP (MEUWISSEN; HAYES; GODDARD, 2001), as this research is a first approach to obtain nutritional models parameters from GWP. Moreover, the ridge regression has the advantage of better handling the very large amount of SNPs markers. The rrBLUP defines genomic prediction in the context of mixed effect models as

$$\begin{aligned} y &= \mu + Zu + \varepsilon \\ u &\sim N(0, K\sigma_u^2) \\ \varepsilon &\sim N(0, I\sigma_\varepsilon^2) \end{aligned} \qquad \text{eq.(1)}$$

Where $\boldsymbol{y}$ is the vector of adjusted phenotypes for sex and contemporaneous group. Depending on the analysis, $\boldsymbol{y}$ assume values of DMI, ADG, fProt, fFat, AFSBW, k_1 or α. $\boldsymbol{\mu}$ is the overall mean, $\boldsymbol{Z}$ is the design matrix for the random effects $\boldsymbol{u}$, and $\boldsymbol{\varepsilon}$ is the random error term. When $\boldsymbol{K}$ is an identity matrix ($\boldsymbol{I}$), the model becomes suitable to estimate marker effects. The solution of eq. (1) assuming an additive model is obtained with a restricted maximum likelihood method as described and implemented by Endelman (2011) in the rrBLUP package for R.

5.2.7 Prediction scenarios for model assessment

Three different scenarios were defined in order to assess the accuracy and precision of model outcomes both when obtained from simulations with regularly computed parameters and when parameters are derived from genomic information. The first scenario (S1) aims to compare GWP predictions of phenotypes with their observed values, and parameters with their regular computation. The second scenario (S2) considers a comparison of model simulations when conducted with parameters obtained from the GWP analysis with observed phenotypes. The third scenario (S3) is intended to contrast model outcomes with parameters computed from GWP regardless the model's ability to simulate animal growth, a matter of the current stage of model development. Thus, S3 compares model outcomes from simulations with genomically computed parameters to those obtained from observed data. In the case of the DGM, when observed performance was used for parameterization, S3 is equivalent to S2.

Scenarios were draw from a Monte Carlo (MC) simulation with 500 iterations. A cross-validation approach was set for each MC iteration in order to split the phenotypic and

genotypic dataset into two different groups (train and testing). For scenarios S2 and S3, simulations from CVDS and DGM were done at each iteration with parameters computed for the testing group using the estimated marker effects obtained from the training subset. The same set of statistics was calculated for model assessment in each scenario, and their mean of 500 MC runs are shown: Pearson correlation coefficient (r), and the regression coefficient and standard error of a linear model comprising observations and predictions through the origin, which serve as a measure of bias (b) and precision, respectively. All computations were done in R v.3.2.0 (R CORE TEAM, 2015).

5.3 Results and discussion

Integration of system dynamics with genomics through model parameters is not only pragmatic and intuitive in terms of modeling between-subject variation, but it also brings a biological explanation. The main assumption of cGP modeling is that causative genetic variation penetrates model parameters more directly than phenotypes, and thus there is a close relation between individual parameter values and the genotype (NORDBØ et al., 2015; VIK et al., 2011). Thus, genetic analysis from such parameters may reveal more heritability than when targeting the complex phenotype (WANG et al., 2012b). In this context, it is possible to differentiate levels of aggregation embedded within selection criteria. Complex traits, which are polygenic and may be understood as outcomes of model behavior, are referred as higher-phenotypes, in a reference to where it stands throughout the biological hierarchy that mechanistically describes a given phenotype. Model core parameters, which are understood as entities that define the system structure, are referred as lower-level phenotypes or component traits of higher-level phenotypes.

At the same time, one of the main components of model development is parameterization: the process of fitting models to the data in an attempt to derive parameter values from measurements (GUNAWARDENA, 2010). It is then very clear to see that models are somewhat dependent on the range of values its variables assume in the dataset used for computing their parameters. This becomes a limitation for modeling. Although the possibility to “foresee” an individual animal characteristic through simulation is what makes modeling attractive for the livestock industry, phenotypic measurements are still needed for parameterization. In addition, between-animal variability is only possible to be captured by models if there is a way to compute individual parameters every time a simulation is

performed. Departing from these concepts, we argued that GWP could be used for inferring individual parameter values in applied nutritional models of cattle growth.

The mean values of statistics obtained from the MC simulation are reported in Tables 1-3. It is difficult to compare the results of GWP obtained here with those reported elsewhere because studies varies largely in the size of their training group and the method used for the analysis. However, our goal was not to explore in detail the potential of GWP and its various methods to predict complex traits such as DMI and ADG, but the focus is on comparing prediction scenarios to contrast GWP to model predictions for the available dataset.

Table 1. Statistics for scenario S1, comparing genomic predictions with observations

Phenotypes[1]	R	SE[2]	Slope[2]
DMI	0.07	0.888	0.998
ADG	0.098	0.255	0.998
fProt	0.194	4.681	1.00
fFat	0.183	11.95	1.00
AFSBW	0.153	16.91	1.00
k_1	0.051	0.002	0.996
A	0.042	0.977	1.00

[1]DMI, dry matter intake; ADG, average daily gain; fProt, final protein content in whole body composition; fFat, final fat content in whole body composition; AFSBW, adjusted final shrunk body weight; k_1, DNA accretion rate; α, constant of energy for maintenance requirement.
[2]SE, standard error, and slope (i.e. bias) are statistics that refer to comparisons between phenotypes predicted by genome-wide selection and observations, or in the case of parameters, with their values computed regularly. Mean values of 500 MC iterations are reported.

Correlation between predicted and observed phenotypes may be used as a performance criterion of GWP analyses, as is the case for scenario S1. Good correlation and accuracy are obtained for fProt (r = 0.194; b = 0%), fFat (r = 0.183; b = 0%), and AFSBW (r = 0.153; b = 0%) when comparing with the other phenotypes. However, this performance is lower than that of GWP for *Bos taurus* reported in the literature for commonly studied traits such as genomic breeding value for DMI and ADG (BOLORMAA et al., 2013; CHEN et al., 2013).

In scenario S2, DMI predicted by CVDS had lower correlation (r = 0.04) and higher bias (b = 6.6%) than scenario S1 (r = 0.07; b = 0.02%). ADG simulated by CVDS in S2 (r = 0.084; b = 31.4%) had also a lower performance than that predicted by GWP in S1 (r = 0.098; b = 0.2%). When individual ADG was known, DMR may be obtained from CVDS runs with

AFSBW predicted using GWP. In scenario S2, DMR showed higher correlation with observed DMI than GWP in S1 to predict DMI. This is the case even when DMR from simulations with regularly computed AFSBW does not perfectly match to that from genomically computed AFSBW, as shown in scenario S3. Thus, under these conditions, CVDS is a better option for predicting DMI than GWP. This is also supported by Kirschten, Pollak and Fox (2007) and Williams et al (2006). Their work computed genetic parameters between model predictions and actual feed intake records, and showed that genetic correlation between them may be as high as 0.85, and heritability of DMR was found to be 0.33 for an Angus cattle dataset. These findings support the idea that selection for model predicted intake values would be effective when considering FE as a breeding goal.

Table 2. Statistics for scenario S2, comparing simulations under genomically computed parameters with observations

Phenotypes[1]	r	SE[2]	Slope[2]
CVDS			
DMI	0.041	1.167	0.934
DMR[3]	0.253	1.727	1.069
ADG	0.084	0.267	0.686
DGM			
ADG	0.38	0.324	0.81
fProt	0.496	6.089	0.984
fFat	0.505	20.19	0.945

[1]DMI, dry matter intake; DMR, dry matter required; ADG, average daily gain; fProt, final protein content in whole body composition; fFat, final fat content in whole body composition.
[2] SE, standard error, and bias are statistics that refer to comparisons between phenotypes predicted by either CVDS or DGM, whose parameters were computed from genomic information, and observations. Mean values of 500 MC iterations are reported.
[3]Statistics for DMR refers to comparisons between observed DMI and DMR.

Although GWP on k_1 and α have not resulted in good accuracy and precision, and despite their influence on protein accretion as described by Oltjen et al. (1986), DGM simulations from scenario S2 were better for predicting ADG, fProt and fFat than S1. These results show that GWP on model parameters may yield better predictability than conducting GWP directly on higher phenotypes. For scenarios S2 and S3 the DGM was superior to CVDS for predicting observed phenotypes, and also for matching simulations with

genomically and regularly computed parameters. This may reveal that parameters of lower level models are more suitable for genomic prediction.

Table 3. Statistics for scenario S3, comparing simulations under genomically and regularly computed AFSBW

Phenotypes[1]	r	SE[2]	Slope[2]
DMR	0.966	0.369	0.991
ADG	0.645	0.172	1.055
k_g	0.875	0.008	0.991

[1]DMR, dry matter required; ADG, average daily gain; kg, conversion efficiency of dietary metabolizable energy to energy for gain.
[2]SE, standard error, and bias are statistics that refer to comparisons between model predictions with genomically and regularly computed parameters. Mean values of 500 MC iterations are reported.

Scenario S3 assess the suitability of model simulations with parameters computed by GWP regardless the model ability to yield accurate predictions. Such analysis indicated that although further model development may be necessary as understood from S2, simulations with genomically and regularly computed parameters match at a reasonable extend. For CVDS, assessment of DMR (r = 0.966; b = 0.9%), ADG (r = 0.645; b = 5.5%), and k_g (r = 0.875; b = 0.9%) showed moderate correlation and low bias between simulations. It is important to emphasize that for the DGM, scenarios S2 and S3 are equivalent once observed animal performance was needed for the optimization procedure used for model fitting. Thus, contrasting simulations with genomically predicted parameters to those under regularly computed ones is the same as designed for S3.

Our study brings generalizations and limitations that may be suppressed in subsequent analysis. While an appropriate framework for conducting GWP with the advantage of its equivalence to GBLUP for this research, the rrBLUP model as defined in eq. (1) assumes equal variance for all markers, and this may not always be the case for a given set of genomic data (WANG et al., 2012a). Methods for GWP from the Bayesian family may yield different outcomes for predicting model parameters from genomic information. Also, it is long recognized that the choice and size of the training group (i.e. reference population) have remarkable influence on the GWP predictability (DAETWYLER; VILLANUEVA; WOOLLIAMS, 2008; VAN GREVENHOF; VAN DER WERF, 2015; SHENGQIANG et al., 2009). Specifically, the dataset for this study was obtained from feeding trials whose animals

had already undergone some kind of selection for entering the efficiency tests. Larger datasets with more heterogeneous individuals may lead to different conclusions from those reported in this study.

The phenotypic dataset was a limitation for computing AFSBW and parameterizing the DGM. Carcass measurements were not available and data from ultrasound had to be used for predicting individual body composition. This scenario certainly contributed to decrease accuracy and precision of models predictions. Therefore, experimental agendas combining genomic and phenotypic observations that allow good model fitting are required to advance towards the interplay between genomic and system dynamics. Also, it seems that the best experimental design would be one very close to measuring the model parameter. This may lead to the development of new phenotyping tools, just as happened in plant physiology (VAN DER HEIJDEN et al., 2012).

In general, model predictions with genomically computed parameters are in accordance with simulations performed with those regularly computed. We expect that as genomic technologies become cheaper in a way to be affordable for commercial operations and with the advance of nutritional models, individual cattle management systems supported by models such as CVDS and DGM would benefit from substituting observations by predictions with GWP to compute parameter values. Animals would be genotyped at the beginning of the feeding period and nutritional model predictions and simulations of between-animal variability would support management strategies that optimize growth, reduce feed costs, and increase profitability in the beef industry.

5.4 Conclusion

Our work has shown that regardless the current state of development of dynamic mechanistic models of cattle growth and their ability to predict phenotypes with good accuracy and precision, simulations conducted with genomically computed parameters are in accordance with those performed with parameters obtained from regular methods. This is the main argument to call attention from the research community that our approach may pave the way for the development of a new generation of applied nutritional models, specifically towards individual-based simulations with subject-specific model parameters computed from genomic information.

References

ALEXANDRE, P. A. et al. Bovine NR1I3 gene polymorphisms and its association with feed efficiency traits in Nellore cattle. **Meta Gene**, v. 2, n. 1, p. 206-217, 2014.

BAKER, M. J. et al. Using ultrasound measurements to predict body composition of yearling bulls. **Journal of Animal Science**, v. 84, n. 10, p. 2666-2672, 2006.

BERRY, D. P.; CROWLEY, J. J. Genetics of feed efficiency in dairy and beef cattle. **Journal of Animal Science**, v. 91, p. 1594-1613, 2013.

BOLORMAA, S. et al. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. **Journal of Animal Science**, v. 91, p. 3088-3104, 2013.

CHEN, L. et al. Accuracy of predicting genomic breeding values for residual feed intake in Angus and Charolais beef cattle. **Journal of Animal Science**, v. 91, p. 4669-4678, 2013.

CHIZZOTTI, M.; PAULINO, P.; VALADARES FILHO, S. Possibilidade do uso da ultrassonografia para estimação da composiçao corporal de bovinos. **Anais**...Lavras, MG: 2008

DAETWYLER, H. D.; VILLANUEVA, B.; WOOLLIAMS, J. A. Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. **PLoS ONE**, v. 3, n. 10, p. e3395, 2008.

ENDELMAN, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. **The Plant Genome Journal**, v. 4, n. 3, p. 250, 2011.

GOMES, R. C. et al. Protein synthesis and degradation gene SNPs related to feed intake, feed efficiency, growth, and ultrasound carcass traits in Nellore cattle. **Genetics and Molecular Research**, v. 12, n. 3, p. 2923--2936, 2013.

GUIROY, P. J. et al. Predicting individual feed requirements of cattle fed in groups. **Journal of Animal Science**, v. 79, p. 1983-1995, 2001.

GUNAWARDENA, J. Models in Systems Biology: The Parameter Problem and the Meanings of Robustness. In: Lodhi, H. M. (Ed.). **Elements of Computational Systems Biology**. John Wiley & Sons, Inc., 2010. p. 19--47.

HERD, R. M.; ARCHER, J. A.; ARTHUR, P. F. The cost of beef production through genetic improvement in residual feed intake: Opportunity and challenges to application. **Journal of Animal Science**, v. 81(E. Suppl. 1), p. E9-E17, 2006.

KIRSCHTEN, D. P.; POLLAK, E. J.; FOX, D. G. Use of a mathematical computer model to predict intake in Angus cattle: Genetic parameters between observed and predicted values, and relationships with other traits. **Anais**...San Antonio, TX: Federation of Animal Science Societies, 2007

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, v. 157, p. 1819-1829, 2001.

NATIONAL RESEARCH COUNCIL. **Nutrient Requirements of Beef Cattle**. 7th ed. Washington, D.C.: National Academy Press, 2000.

NKRUMAH, J. K. et al. Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. **Journal of Animal Science**, v. 84, p. 145-153, 2006.

NORDBØ, Ø. Y. et al. Towards causally cohesive genotype–phenotype modelling for characterization of the soft-tissue mechanics of the heart in normal and pathological geometries. **Journal of The Royal Society Interface**, v. 12, n. 106, abr. 2015.

OLTJEN, J. W. et al. Development of a Dynamic Model of Beef Cattle Growth and Composition. **Journal of Animal Science**, v. 62, p. 86-97, 1986.

PRESS, W. H. et al. **Numerical recipes in C: the art of scientific computing**. Traducao. New York, USA: Cambridge University Press, 1990.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2015.

SANTANA, M. H. A. et al. Feed efficiency and its correlations with carcass traits measured by ultrasound in Nellore bulls. **Livestock Science**, v. 145, n. 1-3, p. 252, 2012.

SANTANA, M. H. et al. Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. **BMC Genetics**, v. 15, n. 1, p. 21, 2014.

SOETAERT, T. P.; PETZOLDT, T. Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package FME. **Journal of Statistical Software**, v. 33, n. 3, fev. 2010.

SOETAERT, K.; PETZOLDT, T.; SETZER, R. W. Package deSolve : Solving Initial Value Differential Equations in R. **Journal of Statistical Software**, v. 33, n. 9, p. 1-25, 2010.

TEDESCHI, L. O.; FOX, D. G.; GUIROY, P. J. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. **Agricultural Systems**, v. 79, n. 2, p. 171, 2004.

TEDESCHI, L. O. et al. Identifying differences in feed efficiency among group-fed cattle. **Journal of Animal Science**, v. 84, p. 767-776, 2006.

VENTURA, R. V. et al. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. **Journal of Animal Science**, v. 92, n. 4, p. 1433-1444, 2014.

VIK, J. O. et al. Genotype–phenotype map characteristics of an in silico heart cell. **Frontiers in Physiology**, v. 2, dez. 2011.

WANG, C. et al. Comparison of five methods for genomic breeding value estimation for the common dataset of the 15th QTL-MAS Workshop. **BMC Proceedings**, v. 6, n. Suppl 2, p. S13, 2012a.

WANG, Y. et al. Parameters in Dynamic Models of Complex Traits are Containers of Missing Heritability. **PLoS Computational Biology**, v. 8, n. 4, p. e1002459, 2012b.

WHITE, R. R.; CAPPER, J. L. An environmental, economic, and social assessment of improving cattle finishing weight or average daily gain within U.S. beef production. **Journal of Animal Science**, v. 91, p. 5801-5812, 2013.

WILLIAMS, C. B. Application of biological simulation models in estimating feed efficiency of finishing steers. **Journal of Animal Science**, v. 88, n. 7, p. 2523, 2010.

WILLIAMS, C. B. et al. Using simulation models to predict feed intake: Phenotypic and genetic relationships between observed and predicted values in cattle. **Journal of Animal Science**, v. 84, p. 1310-1316, 2006.

VAN GREVENHOF, I.; VAN DER WERF, J. Design of reference populations for genomic selection in crossbreeding programs. **Genetics Selection Evolution**, v. 47, n. 1, p. 14, 2015.

VAN DER HEIJDEN, G. et al. SPICY: towards automated phenotyping of large pepper plants in the greenhouse. **Functional Plant Biology**, v. 39, n. 11, p. 870, 2012.

6 General discussion

The research agenda proposed by this dissertation came out not just because the interplay between genomics and nutritional models would contribute to uncover the genetic basis underlying complex traits but mainly because it represents a framework for evolving towards a new generation of applied nutritional models. Ideas for production systems using genomics for cattle management have emerged at the same time genomic science was beginning to be applied to animal breeding (BATES et al., 2005; BUENFIL et al., 2009; DENISE et al., 2004; FLUHARTY; JACKWOOD, 2001; TAYLOR et al., 2002). However, we proposed a different approach from that used in these first attempts, where molecular genetic values are predictors in linear models or animals are simply targeted to specific groups based on genotypic lines.

We adopted an integration of dynamic mechanistic models of cattle growth with genomics by focusing on the model core parameters. Our approach stands out between that of causally-cohesive genotype-to-phenotype (cGP) models (WANG et al., 2012) and functional genome-wide association analysis (*fGWAS*, DAS et al., 2011). We assume that genetic variation of a complex phenotype penetrates model parameters directly, and thus they are important traits of the system. However, considering the current stage of development of the models described in the literature and the phenotyping tools available for research use, we had to estimate these parameters as is the case for *fGWAS* and assume them as observed phenotypes, instead of actually measuring them on the animals, a condition of the cGP approach.

Many growth models could have served as a standard for exploring the possibilities of integrating nutritional models with genomics. However, two of those available on the scientific literature offered an interesting contrast in terms of the aggregation level of biological hierarchies: the Cornell Cattle Value Discovery System (CVDS, TEDESCHI; FOX; GUIROY, 2004) which is widely used in commercial feedlot operations in the United States and is comprised of sequential equations from the National Research Council (NRC), (2000) system; and the Davis Growth Model (DGM), which refers to the differential equations developed by Oltjen et al. (1986) describing DNA accretion, protein synthesis and degradation, and fat synthesis.

First, if we assume that genetic variance penetrates model parameters it is important to check the validity of their biological interpretation – a feature imposed by mathematical formalisms – as seen from the perspective of genomics. A genome-wide association study

(GWAS) was the strategy adopted for such verification. Simply contrasting single nucleotide polymorphisms (SNPs) that are significant for observed phenotypes to those mapped for the parameters was not a consistent method. This is because if we assume that genetic variance penetrates model parameters better than using the higher phenotype as a response in statistical models, it is expected that parameters could yield novel potential genetic markers associated with the trait, and these markers would not necessarily match to those identified by means of the observed higher phenotype. Moreover, SNPs identified for a given production trait and described in the literature are not definitive causative SNPs (BOYLES, 2010; SHIELDS, 2011). Thus, a possible method for verifying the biological interpretation of model parameters was contrasting their mapped genomic regions to previously identified quantitative trait loci (QTLs) surrounding significant SNPs for the parameters. It is expected that regions mapped for the parameters would harbor QTLs associated with the complex trait.

Performing GWAS on two CVDS parameters (AFSBW, adjusted final shrunk body weight, and k_g, the conversion efficiency of dietary metabolizable energy into net energy for gain) and on two parameters of the DGM (k_1, the DNA accretion rate, and α, coefficient of energy for maintenance requirement) revealed that genomic regions associated to them harbor QTLs for correlated complex traits: AFSBW and k_1 with mature body weight, and k_g and α with feed efficiency traits. This provides a proof of concept that CVDS and DGM parameters are indeed phenotypes whose expected interpretations may be stated by means of their genomic regions.

From genotypic probabilities of SNP markers mapped for AFSBW with GWAS we tested our hypothesis that simulations could be done from model parameters computed with genomic information. Using a theoretical model proposed by Aulchenko et al. (2009) we demonstrated that model simulations performed with AFSBW computed from genomic information are quite consistent to those performed with AFSBW computed by means of regular methods. This finding opens up opportunities to explore the use of genomic data to compute subject-specific parameters and to capture between-animal variation in modeling exercises.

Also, motivated by the expectation that component traits (i.e. model parameters) would contribute to uncover the genetic basis of complex traits and that their genetic variation would reveal some of the unexplained phenotypic variance of the complex trait we defined a linear model in an attempt to predict the complex average daily gain (ADG) from estimated genotypic probabilities of SNPs from GWAS performed directly on the ADG or from GWAS performed on parameter k_1. SNPs identified for k_1 were able to additionally explain 2.2% of

the phenotypic variability of ADG, even when SNPs for k_1 did not match to any marker significant for ADG. This suggest that GWAS on the component traits may reveal additional genomic regions underlying the complex phenotype, which otherwise would not be captured by studies conducted on the complex phenotype alone.

Finally, we used a method of genomic prediction (ridge-regression BLUP, ENDELMAN, 2011) to compute parameters for CVDS and DGM. This was different from the method of Aulchenko et al. (2009) because here we definitely estimated marker effects from a reference population for computing AFSBW, k_1 and α for a testing group. We compared different prediction scenarios for desirable phenotypes, which were estimated either by performing genomic prediction directly on them or first predicting model parameters and then using model predictions to infer observations. We found that while genomic prediction on higher phenotypes may still be a better option than predictions from growth models, which are somewhat related to their current stage of development, simulations conducted with genomically computed parameters are in accordance with those performed with parameters obtained from regular methods. This is the main argument to call attention from the research community that this approach may pave the way for the development of a new generation of applied nutritional models, specifically towards individual-based simulations with model parameters computed from genomic information.

References

AULCHENKO, Y. S. et al. Predicting human height by Victorian and genomic methods. **European Journal of Human Genetics**, v. 17, n. 8, p. 1070, 2009.

BATES, S. et al. **Methods and systems for inferring traits to manage non-beef livestock**. [s.l.] Google Patents, 2005.

BOYLES, A. L. The quest for the elusive causative SNP. **Kidney International**, v. 78, n. 1, p. 9, 2010.

BUENFIL, M. A. G. et al. **Animal management**. [s.l.] Google Patents, 2009.

DAS, K. et al. A dynamic model for genome-wide association studies. **Human Genetics**, v. 129, n. 6, p. 629, 2011.

DENISE, S. K. et al. **Compositions, methods, and systems for inferring bovine breed**. [s.l.] Google Patents, 2004.

ENDELMAN, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. **The Plant Genome Journal**, v. 4, n. 3, p. 250, 2011.

FLUHARTY, F.; JACKWOOD, D. J. **Methods for assessing the beef characteristics of live cattle**. [s.l.] Google Patents, 2001.

NATIONAL RESEARCH COUNCIL. **Nutrient Requirements of Beef Cattle**. 7th ed. Washington, D.C.: National Academy Press, 2000.

OLTJEN, J. W. et al. Development of a Dynamic Model of Beef Cattle Growth and Composition. **Journal of Animal Science**, v. 62, p. 86--97, 1986.

SHIELDS, R. Common Disease: Are Causative Alleles Common or Rare? **PLoS Biology**, v. 9, n. 1, p. e1001009, 2011.

TAYLOR, J. F. et al. **Method of managing and marketing livestock based on genetic profiles**. [s.l.] Google Patents, 2002.

TEDESCHI, L. O.; FOX, D. G.; GUIROY, P. J. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. **Agricultural Systems**, v. 79, n. 2, p. 171, 2004.

WANG, Y. et al. Parameters in Dynamic Models of Complex Traits are Containers of Missing Heritability. **PLoS Computational Biology**, v. 8, n. 4, p. e1002459, 2012.