UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

FERRAMENTAS DE SELEÇÃO PARA UNIFORMIDADE DE PRODUÇÃO EM BOVINOS DA RAÇA NELORE

Laiza Helena de Souza Iung
Zootecnista

2018

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

FERRAMENTAS DE SELEÇÃO PARA UNIFORMIDADE DE PRODUÇÃO EM BOVINOS DA RAÇA NELORE

Laiza Helena de Souza Iung

Orientador: Dr. Roberto Carvalheiro

Coorientadores: Dr. Haroldo Henrique de Rezende Neves

Dr. Herman Arend Mulder

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para obtenção do título de Doutora em Genética e Melhoramento Animal

2018

Iung, Laiza Helena de Souza
I92f Ferramentas de seleção para uniformidade de produção em bovinos da raça Nelore / Laiza Helena de Souza Iung. – – Jaboticabal, 2018
xii, 94 p. : il. ; 29 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2018
Orientador: Roberto Carvalheiro
Coorientador: Haroldo Henrique de Resende Neves, Herman Arend Mulder
Banca examinadora: Ricardo Vieira Ventura, Lucia Galvão de Albuquerque, Gerson Antonio de Oliveira Júnior, Diercles Francisco Cardoso
Bibliografia

1. Bovinos de corte. 2. Características de crescimento. 3. Estudo de associação genômica ampla. 4. Método em dois passos. 5. Modelo linear generalizado hierárquico duplo. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.2

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação – Diretoria Técnica de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: FERRAMENTAS DE SELEÇÃO PARA UNIFORMIDADE DE PRODUÇÃO EM BOVINOS DA RAÇA NELORE

AUTORA: LAIZA HELENA DE SOUZA IUNG
ORIENTADOR: ROBERTO CARVALHEIRO
COORIENTADOR: HAROLDO HENRIQUE DE REZENDE NEVES
COORIENTADOR: HERMAN AREND MULDER

Aprovada como parte das exigências para obtenção do Título de Doutora em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:

Pesquisador Dr. ROBERTO CARVALHEIRO (Videoconferência)
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Prof. Dr. RICARDO VIEIRA VENTURA
Departamento de Nutrição e Produção Animal/FMVZ-USP / Pirassununga/SP

Profa. Dra. LUCIA GALVÃO DE ALBUQUERQUE
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Dr. GERSON ANTONIO DE OLIVEIRA JÚNIOR
CRV-Lagoa / Sertãozinho/SP

Pós-doutorando DIERCLES FRANCISCO CARDOSO
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Jaboticabal, 20 de julho de 2018

DADOS CURRICULARES DO AUTOR

Laiza Helena de Souza Iung – nascida em Bom Retiro, Santa Catarina, Brasil, em 23 de Maio de 1989, filha de Simone Aparecida de Souza e Bertoldo Saulo Iung. Iniciou sua graduação em Zootecnia na Universidade Estadual de Santa Catarina (UDESC) em março de 2008. Durante a graduação (2008 – 2012) participou de projetos científicos e foi monitora nas disciplinas de Bioquímica e Melhoramento Animal. Em Agosto de 2012, iniciou seu mestrado no programa de pós-graduação em Ciência Animal e Pastagens da ESALQ/USP, onde foi bolsista CAPES e FAPESP. Durante o mestrado dedicou-se a estudar efeitos simples e de interação associados com produção e composição do leite, incluindo perfil de ácidos graxos, em bovinos da raça Holandesa, sob orientação do prof. Dr. Gerson Barreto Mourão, recebendo o título de mestre em Julho de 2014. No mesmo ano iniciou o doutorado no programa de pós-graduação em Genética e Melhoramento Animal da FCAV/Unesp, sendo bolsista Capes durante todo o período. Durante Abril/2017 à Março/2018 foi bolsista pelo programa de doutorado sanduíche da CAPES (processo número 88881.131673/2016-01), desenvolvendo parte do projeto na Universidade de Wageningen, Holanda, sob orientação do prof. Dr. Herman Arend Mulder.

Brazilient

Dedico

Ao meu eterno amigo, namorado e marido, ***Thiago***, por nunca ter medido esforços em caminhar ao meu lado, onde quer que seja rumo às nossas conquistas.

AGRADECIMENTOS

Agradeço,

Em primeiro lugar a Deus, pela vida e por sempre me guiar e iluminar meus pensamentos e decisões e não me deixar desistir nos momentos de questionamento e indecisão.

Ao meu marido, Thiago, por estar sempre ao meu lado em cada etapa das minhas escolhas profissionais, pelo encorajamento e conversas que me fizeram manter o foco e deixar de pensar no amanhã e viver mais o presente (nem sempre foi fácil), e especialmente pela paciência nos momentos em que ela foi mais necessária, pois além de tudo ele casou com uma geminiana, e só quem convive com uma pessoa desse signo sabe do que estou falando.

Ao meu filho de quatro patas Leleco Junior, pelo amor imensurável, pelos vários lambeijos e brincadeiras que transformaram os meus dias em momentos repletos de risadas e alegrias, por não me deixar ser sedentária também, porque passeios no bairro e na Unesp nos finais de semana sempre fizeram parte da nossa rotina. Se um dia me questionei se a sua vinda à minha vida foi uma boa escolha, hoje digo sem dúvida nenhuma que foi a melhor que eu fiz.

A minha família, em especial aos meus pais, Simone e Alceu, avós e irmãos por me encorajarem na busca dos meus objetivos.

Aos amigos que não deixaram a amizade morrer pela distância física e aos novos amigos conquistados durante estes anos, não só aqui no Brasil mas na Holanda.

Sou imensamente grata aos meus orientadores, prof. Roberto, Haroldo e Han, pela oportunidade, disposição, paciência e atenção em me orientar e por todo o conhecimento que adquiri com vocês ao longo destes anos, que com certeza vai muito além desta tese.

As bancas de qualificação e defesa por todas as considerações e sugestões feitas ao projeto, tornando o mesmo ainda mais completo.

A “Família Carvalheiro” (Grazyella, Giovana, Henrique, Thaise, Daiane, Baltazar, Angel e Helsi) pela companhia, pelas várias celebrações de aniversário,

happy hours, pelas conversas sérias e as não tão sérias assim, enfim, por todos os momentos que passamos juntos.

Aos colegas de departamento, pela companhia quase que diária durante os intervalos, compartilhando café, bolo, paçoquinha e "causos", pelos churrasquinhos e happy hours.

Aos amigos que a Holanda me deu, em especial Rodrigo, Mieke, André e Karol, que proporcionaram a mim e a minha família momentos de alegrias e nos ajudaram a solucionar alguns perrengues também.

A Unesp e Universidade de Wageningen pela oportunidade de desenvolver este estudo e a todos os professores que participaram direta ou indiretamente na construção do mesmo.

A CAPES pelas bolsas concedidas durante todo o período do doutorado.

SUMMARY

FERRAMENTAS DE SELEÇÃO PARA UNIFORMIDADE DE PRODUÇÃO EM BOVINOS DA RAÇA NELORE

Resumo – A existência de um controle genético sobre a variância residual torna possível a seleção para uniformidade de produção. Desta forma, os objetivos do presente estudo foram: i) investigar a presença deste componente genético na variância residual do peso ao sobreano (PS) em bovinos da raça Nelore (N = 194.628, touros = 625), usando duas abordagens: dois passos (*two-step*) e modelo linear generalizado hierárquico duplo (*double hierarchical generalized linear model*, DHGLM); e ii) identificar, através de estudo de associação genômica ampla (*genome-wide association study,* GWAS), regiões genômicas associadas com uniformidade do PS usando differentes variáveis respostas: EBV desregredido ($dEBV_v$) obtido a partir de soluções de um DHGLM assumindo correlação genética não-nula entre média e variância residual ($r_{mv} \neq 0$); e $dEBV_{v\_r0}$ e variância dos resíduos log-transformada ($\ln\_\sigma^2_{\hat{e}}$) obtidos a partir de soluções de um DHGLM assumindo $r_{mv} = 0$. Os resultados confirmam a presença de heterogeneidade genética de variância residual bem como oportunidade de seleção (coeficiente de variação genética da variância residual (GCV_E) variando de 0,10 a 0,17). Porém, a seleção pode ser limitada devido à magnitude da variância genética da variância residual e à forte e positiva correlação genética entre a média e a variância (0,20 e 0,76 para a abordagem em dois passos e DHGLM, respectivamente). Além disso, foi possível observar que um grande número de progênies por touro é necessário para obter EBV com maiores acurácias (herdabilidade da variância residual, $h^2_v < 0,007$). A relação média-variância foi reduzida com a utilização da transformação Box-Cox, o que provocou uma redução nas estimativas dos parâmetros genéticos (h^2_v, GCV_E) e mudou o sinal e a magnitude das correlações genéticas entre média e variância residual. DHGLM foi superior à abordagem em dois passos de acordo com as comparações realizadas, incluindo um delineamento de validação-cruzada. Em relação aos resultados do GWAS, dEBV da média ($dEBV_m$) foi altamente correlacionado com $dEBV_v$ (0,90), resultando em regiões associadas simultaneamente com a média e variância residual do PS, no qual demonstraram estar além dos efeitos de escala. Resultados mais independentes entre média e variância residual foram obtidos quando $r_{mv} = 0$ foi assumida. No total, 13 e 4 polimorfismos de nucleotídeo único (SNPs) mostraram uma forte associação (*Bayes Factor* > 20) com $dEBV_v$ e $\ln\_\sigma^2_{\hat{e}}$, respectivamente, apenas sugestivos sinais (*Bayes Factor* > 3) foram encontrados para $dEBV_{v\_r0}$. Todas as janelas de 1-Mb compartilhadas entre as top 20 entre $dEBV_m$ e $dEBV_v$ foram previamente associadas com características de crescimento. Potenciais genes candidatos do ponto de vista biológico para uniformidade estão envolvidos em metabolismo (principalmente energético e lipídico), resposta ao estresse, resposta inflamatória e imune, mineralização, atividade neuronal e formação óssea. É necessário usar uma estratégia como assumir $r_{mv} = 0$ a fim de obter resultados relacionados à uniformidade e não a média. Nenhuma evidência clara em favor da utilização de uma variável resposta específica foi observada, por isso recomenda-se a utilização de diferentes variáveis para estudar uniformidade a fim de aumentar as evidências

sobre regiões candidatas e os mecanismos biológicos que influenciam este tipo de resposta.

Palavras-chave: bovinos de corte, características de crescimento, estudo de associação genômica ampla, método em dois passos, modelo linear generalizado hierárquico duplo

SELECTION TOOLS FOR UNIFORMITY OF PRODUCTION IN NELLORE CATTLE

Abstract – The existence of genetic control on residual variance makes possible the selection for uniformity. Thus, the objectives of the present study were: i) to investigate the presence of this genetic component on the residual variance of yearling weight (YW) in Nellore cattle (N = 194,628, sires = 625) using two approaches: two step approach and double hierarchical generalized linear model (DHGLM); and ii) to identify, through a genome-wide association study (GWAS), genomic regions associated with uniformity of YW using different response variables: deregressed EBV ($dEBV_v$) obtained from solutions of a DHGLM assuming non-null genetic correlation between mean and residual variance ($r_{mv} \neq 0$); and $dEBV_{v\_r0}$ and log-transformed variance of estimated residuals ($ln\_\sigma^2_{\hat{e}}$) obtained from solutions of a DHGLM assuming $r_{mv} = 0$. The results confirm the presence of genetic heterogeneity of residual variance as well as opportunity of selection (coefficient of genetic variance of residual variance (GCV_E) ranging from 0.10 to 0.17). However, selection may be limited by the magnitude of the genetic variance of the residual variance and the strong and positive genetic correlation between mean and variance (0.20 and 0.76 for two-step approach and DHGLM, respectively). In addition, it was observed that large sire families are required to obtain higher EBV_v accuracies (heritability of residual variance, $h^2_v < 0.007$). The mean-variance relationship was reduced by using Box-Cox transformation, which reduced the genetic parameter estimates (h^2_v, GCV_E) and changed the signal and magnitude of the genetic correlation between mean and residual variance. The DHGLM was superior to the two-step approach according to the comparisons, including a cross-validation design. Regarding the GWAS results, dEBV of the mean ($dEBV_m$) were highly correlated with $dEBV_v$ (0.90), resulting in regions associated simultaneously with mean and residual variance of YW, in which they proved to be beyond scale effects. More independent results between mean and residual variance were obtained when $r_{mv} = 0$ was assumed. In total, 13 and 4 single nucleotide polymorphisms (SNPs) showed a strong association (Bayes Factor > 20) with $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$, respectively, only suggestive signals (Bayes Factor > 3) were found for $dEBV_{v\_r0}$. All 1-Mb windows shared among top 20 between $dEBV_m$ and $dEBV_v$ were previously associated with growth traits. Potential candidate genes from the biological point of view for uniformity are involved in metabolism (mainly energetic and lipid), stress, inflammatory and immune responses, mineralization, neuronal activity, and bone formation. It is necessary to use a strategy like assuming $r_{mv} = 0$ in order to obtain results related only with uniformity. No clear evidence in favor of using a specific response variable was observed, we recommend using different variables to study uniformity to increase evidence on candidate regions and biological mechanisms behind it.

Keywords: beef cattle, double hierarchical generalized linear model, growth traits, genome-wide association study, two-step approach

CHAPTER 1 – General Considerations

1.1 Introduction

Brazil has been consolidated as one of the most important beef producers and exporters in the world (USDA, 2017). In 2017, Brazilian beef production reached about 9.5 million tons, behind only the USA (12.1 million tons), shipping about 20% (1.76 million tons) abroad. Despite the importance of this activity to the country, most production systems still have a low efficiency mainly due to the long cycle of production and consequent commercialization of late animals, characterized by their low carcass quality and lack of carcass standardization.

The ability to adapt to the Brazilian tropical conditions and good growth rate made the Nellore (*Bos taurus indicus*) the most used breed in the country. In such scenario, the use of pastures was favored as the main feed source, which provides a low production cost. Despite being the most used breed for beef production, effective Nellore breeding programs have been established relatively recently (CARVALHEIRO, 2014). Since then, an increasing number of animals have been controlled every year, which allows a faster genetic progress in this breed.

Most of these breeding programs have emphasized growth traits, mainly weaning and yearling weight (CARVALHEIRO, 2014). However, improving only the mean of these traits has not become enough. Many slaughterhouses have adopted special payment programs to encourage farmers to deliver carcasses within specific standards regarding to weight and backfat thickness. They have stimulated the search for more uniform animals, mainly through sire selection, since males have higher selection intensity in cattle breeding, producing more progeny than females.

Reducing phenotypic variability can be achieved through genetic selection, since empirical evidence has shown that genotypes also respond differently to environmental variations leading to genetic differences in residual variance, which is called genetic heterogeneity of residual variance or micro-environmental sensitivity (e.g. MACKAY; LYMAN, 2005; MULDER; BIJMA; HILL, 2007). So far, few genetic studies have been developed to investigate the existence of this genetic component and opportunity to improve uniformity in beef cattle (NEVES; CARVALHEIRO; QUEIROZ, 2011, 2012).

1.2 Literature review

1.2.1 Importance of uniformity and evidence of its genetic control

The search for uniformity of production has become an important breeding goal in several livestock species, such as pigs, chickens and beef cattle. The main and most common reason for it is the economic return that more uniform animal production can bring through differentiated payment according to specific standards by slaughterhouses. Besides that, additional economic losses can be attributed to the lack of homogeneity. In aquaculture, competition and consequently growth rate, mortality, feed efficiency and intake are influenced by a large size variation in a population (GILMOUR; DIBATTISTA; THOMAS, 2005; JANHUNEN; KAUSE; JÄRVISALO, 2012). These effects are more evident in aquaculture species due to a closer interaction in this production environment; however, mortality as a result of competition was already observed in piglets (MILLIGAN; FRASER; KRAMER, 2002). In addition, higher phenotypic variation increases the need for management like size grading, which is harmful to animal welfare.

From a genetic point of view, selection and mating strategies have demonstrated to be efficient ways of achieving uniformity (HOHENBOKEN, 1985; KAUSE et al., 2014; NEVES et al., 2009). The feasibility of selection for uniformity requires the existence of a heritable genetic component related to it. Substantial evidence have supported the hypothesis that residual variance is also under genetic control and it can be used to improve uniformity (reviewed by HILL; MULDER, 2010). Among these there are some studies related to model species such as bristle number in isogenic lines of *Drosophila melanogaster (*MACKAY; LYMAN, 2005) and adult weight in snails (ROS et al., 2004), but also those involving livestock species like litter size in sheep and pigs (SANCRISTOBAL-GAUDY et al., 2001; SORENSEN; WAAGEPETERSEN, 2003) and body weight in chickens (ROWE et al., 2006).

1.2.2 Approaches used to evaluate uniformity

Classical quantitative genetic models normally assume that residual variance is constant, i.e. genotypes differ only on the mean effect (FALCONER; MACKAY, 1996). However, based on the evidence that residual variance can also be under

genetic control, alternative models and approaches accommodating these potential genotypic differences in terms of residual variance were developed.

Following the classical model, P = A + E, basically two extensions were proposed, the first modelled additive genetic effect also on the residual variance: $P = \mu + A_m + \chi\sqrt{\sigma_e^2 + A_v}$, where μ is the mean trait value, σ_e^2 is the mean residual (environmental) variance of the population, A_m and A_v are EBV for the mean and residual variance, respectively (HILL; ZHANG, 2004, MULDER; BIJMA; HILL, 2007), and the second assumed used an exponential model for the residual variance: $P = \mu + A_m + \chi \exp(\frac{1}{2}\ln(\sigma^2_{E,exp}) + A_{v,exp})$, in which $ln(\sigma^2_{E,exp})$ is the natural logarithm of the residual variance in the exponential model and $A_{v,exp}$ is the EBV for residual variance also in the exponential model (SANCRISTOBAL-GAUDY et al., 1998, 2001; SORENSEN & WAAGEPETERSEN, 2003, ROS et al. 2004). Although both can be converted to each other, some disadvantages were previously pointed out by Mulder, Bijma and Hill (2007). According to them, in the first model, $\sigma_e^2 + A_v$ must be higher than 0 and in the latter the residual variance in the population cannot be full splitted into residual variance in the mean part and genetic variance in the residual variance part.

Modelling variance on the log scale as used in the exponential model helps to assure the normal distribution assumption. This idea led to the development of the most used models to evaluate uniformity by using an estimate of residual variance as response variable into two main REML methods: two-step approach and double hierarchical generalized linear model (DHGLM). The first one was proposed initially by Rowe et al. (2006), which analyzed residual variance of body weight in broilers with least squares. In the first step, body weight was fitted in a linear model (mean model), allowing for heterogeneous residual variance among sire families, and in the second step (residual variance model) these estimates of residual variance were log-natural transformed and used as phenotype. In order to account for both genetic and environmental effects on the residual variance at the level of individual records, an extension of this approach was suggested by Mulder et al. (2009). In this new two-step version, natural logarithm of the squared estimated residuals ($\ln(\hat{e}^2)$) were used as response variable in the second step instead of a measure per family, in which the

estimated residuals were obtained from the mean model. An example of the mean and residual variance models used in the two-step are:

$$y = Xb + Za + e \qquad \text{(Step 1 - mean model)}$$

$$\ln\left(\hat{e}^2\right) = Xb_v + Za_v + e_v \qquad \text{(Step 2 – residual variance model)}$$

where y and $\ln\left(\hat{e}^2\right)$ are the response variables for the mean and residual variance models, respectively, Xb (Xb_v), Za (Za_v), e (e_v) are fixed and random effects for the mean ans residual variance models, respectively. In the first step, either homogeneous or heterogeneous residual variance among families can be assumed, allowing for heterogeneity of residual variance. The log transformation of the squared residuals was adopted in order to overcome the asymmetry in their distribution. However, this approach can produce biased estimates of the variance components because it does not account for the uncertainty or imprecision of the estimated residuals (RÖNNEGÅRD et al., 2013).

Recently, a method that can take into account the imprecision of estimated residuals and is based on generalized linear models (LEE et al., 1996) has been widely used in genetic heterogeneity in residual variance studies (e.g. MULDER et al., 2013; RÖNNEGÅRD et al., 2013; VANDENPLAS et al., 2013). Rönnegård et al. (2010) proposed the double hierarchical generalized linear model (DHGLM), an iterative approach that assumes a gamma distribution for the squared residuals, transposing the non-normality issue by using purely squared residuals. Thereafter, Felleki et al. (2012) extended this approach by fitting the trait by itself (also termed as mean) and residual variance at the same time through a bivariate model, as follows:

$$\begin{bmatrix} y \\ \psi \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & X_v \end{bmatrix} \begin{bmatrix} b \\ b_v \end{bmatrix} + \begin{bmatrix} Z & 0 \\ 0 & Z_v \end{bmatrix} \begin{bmatrix} a \\ a_v \end{bmatrix} + \begin{bmatrix} e \\ e_v \end{bmatrix}$$

Where *y* is the trait and *ψ* is the linearized response variable for the residual variance, defined as: $\psi_i = \log(\hat{\sigma}^2_{e_i}) + \left(\left\{\left[\hat{e}_i^2/(1 - h_i)\right] - \hat{\sigma}^2_{e_i}\right\} / \hat{\sigma}^2_{e_i}\right)$, which $\hat{\sigma}^2_{e_i}$ is the predicted residual variance for observation *i* obtained in the previous iteration, $\hat{e}_i^2$ is the squared residual from the mean model estimated for observation *i* and h_i is the leverage for observation *i*, i.e. the i^{th} diagonal element of the hat matrix (HOAGLIN & WELSCH, 1978). Fixed and random effects are represented by Xb (X_vb_v), Za (Z_va_v) and e (e_v), for the mean and residual variance (subscript v), respectively.

An alternative approach, rarely used, is the hierarchical Bayesian models, applying Markov chains (MCMC, e.g. IBÁÑEZ-ESCRICHE et al., 2008; SORENSEN; WAAGEPETERSEN, 2003). This method allows estimating genetic effects on the mean and the residual variance simultaneously; however, it has a large computational demand, mainly for a huge amount of data.

The partial genetic control of residual variance found through these approaches can be used to improve uniformity of production. However, this is not the only way to achieve genetic progress on phenotypic variability since uniformity can also be measured in different ways. Basically, three main factors limit the choice of the response variable for uniformity: trait, data structure and availability of genotypes, the latter being required mainly for genomic studies. These limitations happen because statistical analysis for uniformity requires within-individual variance based on repeated observations per animal or within-family variance based on large offspring groups per family. When individuals have repeated observations of a trait, such as harvest weight within family and birth weight within litter, standard deviations and coefficient of variation can be taken from these groups and used to quantify uniformity, as done by Khaw et al. (2016), Zhang et al. (2016) and Wang et al. (2017). The advantage of using these measures is the easier interpretation and because they are on the same scale as the trait, compared to the response variables used in the two-step and DHGLM approaches.

1.2.3 Genetic parameters

The interpretation and comparison of results on genetic heterogeneity of residual variance have been made through two genetic parameters: heritability of residual variance (h^2_v) and genetic coefficient of variation (GCV_E), also known as evolvability (MULDER et al., 2007). According to these authors, h^2_v is a measure to determine the accuracy of EBV for residual variance, and it is calculated as: $h^2_v = \sigma^2_{a_{v,add}} / (2\sigma^4_P + 3\sigma^2_{a_{v,add}})$, where $\sigma^2_{a_{v,add}}$ is the additive genetic variance estimated for the residual variance on the additive scale and σ^4_P is the squared phenotypic variance. The $\sigma^2_{a_{v,add}}$ was obtained as: $\sigma^2_{a_{v,add}} = h^2_{res} 2(\bar{\sigma}^2_{\hat{e}})^2$, where h^2_{res} is the heritability of the response variable used in the residual variance model, and $\bar{\sigma}^2_{\hat{e}}$ is the average

residual variance in the mean model (MULDER et al., 2009). The GCV_E allows evaluating the potential response to selection for residual variance, and it is calculated as: $GCV_E = \sigma_{a_{v,add}} / \bar{\sigma}^2_{\hat{e}}$, which $\sigma_{a_{v,add}}$ is the standard deviation of the additive genetic variance for the residual variance on the additive scale. In a previous review, Hill and Mulder (2010) found values ranging from 0.0 to 0.05 for h^2_v and 0.0 to 0.60 for GCV_E. Thereafter, the majority of the estimates were reported within these ranges, and in some cases these values were close to 0.10 for h^2_v and 0.79 for GCV_E (Table 1a to Table1e). These results indicate the need for large family data to obtain accurate EBV for residual variance and highlight the opportunity to change residual variance through selection, through h^2_v and GCV_E, respectively.

An additional genetic parameter to be considered, especially to guide selection decisions when uniformity is part of breeding goal, is the genetic correlation between mean and residual variance (r_{mv}). Estimates of r_{mv} covering the whole parametric space (-1.0 to 1.0) were reported for different livestock species (Table 1a to Table 1e). The most negative values are the most favorable if the breeding goals are to increase the mean and reduce the variance; otherwise, mean and residual variance need to be included in a selection index to obtain the desired response.

Partially, the wide range of r_{mv} estimates in terms of sign and magnitude can be attributed to the use of transformation, such as Box-Cox and log, as observed in Felleki and Lundeheim (2015), Sae-Lim et al. (2015), Sonesson et al. (2013) and Yang et al. (2011). Data transformation is used to account for scale effects, i.e. higher means associated with higher variances, and then reduce the mean-variance relationship (BOX; COX, 1964).

Results from divergent selection experiments in mice and rabbits have shown the feasibility of improve residual variance to achieve uniformity. In mice, response to selection for birth weight variability was verified from the second generation as well as evidence of correlated responses for production and welfare traits (FORMOSO-RAFFERTY et al., 2016a, 2016b). Direct and correlated responses were also observed for litter size in rabbits (GARREAU et al., 2008; BLASCO et al., 2017; ARGENTE et al., 2017).

Table 1a. Summary of the studies on uniformity in sheep

Traits	Method	h^2_v	GCV_E	r_{mv}	Reference
Litter size	REML	0.048	0.51	0.19	SanCristobal-Gaudy et al. (2001)
Litter size	REML	0.10	0.50	0.88 (0.07)	Cottle et al. (2016)
Birth weight	REML	0.08	0.58	0.39	Sae-Lim et al. (2018)

h^2_v: heritability of residual variance or uniformity; GCV_E: genetic coefficient of variation of residual variance or uniformity; r_{mv}: genetic correlation between mean and residual variance; REML: restricted estimate maximum likelihood; when availabele, standard errors are presented between brackets.

Table 1b. Summary of the studies on uniformity in pigs

Traits	Method	h^2_v	GCV_E	r_{mv}	Reference
pH muscle	REML	0.039	0.40	0.79	SanCristobal-Gaudy et al. (1998)
Litter size	MCMC	0.026	0.31	-0.62	Sorensen and Waagepetersen (2003)
Slaughter weight	MCMC	0.011	0.34	-0.07 (-0.26;0.12)	Ibáñez-Escriche et al. (2008)
Litter size	MCMC	-	-	-0.64 (-0.82;-0.45)	Yang, Christensen and Sorensen (2011)
Litter size[1]	MCMC	-	-	0.70 (0.44;0.98)	Yang, Christensen and Sorensen (2011)
Birth weight	REML	0.008 to 0.011	0.087 to 0.105	0.23 (0.41) to 0.62 (0.12)	Sell-Kubiak et al. (2015)
Birth weight	REML	0.011 to 0.015	0.10 to 0.12	0.53 (0.08) to 0.59 (0.06)	Sell-Kubiak et al. (2015b)
Litter size	REML	0.006 (0.001) to 0.01 (0.006)	0.087 (0.004)	0.49 (0.04) to 0.68 (0.04)	Sell-Kubiak et al. (2015c)
Teat count	REML	0.06 to 0.07	0.48 to 0.51	0.76 (0.033) to 0.79 (0.025)	Felleki and Lundeheim (2015)
Teat count[1]	REML	0.04	0.40 to 0.41	0.50 (0.056) to 0.54 (0.044)	Felleki and Lundeheim (2015)
Growth traits	REML	0.006 (0.001) to 0.008 (0.001)	0.24 (0.017) to 0.27 (0.018)	0.12 to 0.21	Gorssen, Buys and Janssens (2018)

h^2_v: heritability of residual variance or uniformity; GCV_E: genetic coefficient of variation of residual variance or uniformity; r_{mv}: genetic correlation between mean and residual variance; REML: restricted estimate maximum likelihood; MCMC: Markov chain Monte Carlo; when availabele, standard errors or posterior intervals (in case of MCMC method) are presented between brackets.
[1]: using a transformed dataset

Table 1c. Summary of the studies on uniformity in chicken

Traits	Method	h^2_v	GCV_E	r_{mv}	Reference
Body weight[1]	REML	0.029 to 0.030	0.30 to 0.32	-0.17 to -0.11	Rowe et al. (2006)
Body weight[1]	REML	0.030 (0.05) to 0.038 (0.006)	0.32 (0.030) to 0.37 (0.032)	-0.27 to -0.23	Wolc et al. (2009)
Conformation score[1]	REML	0.023 (0.004) to 0.032 (0.005)	0.25 (0.023) to 0.31 (0.025)	0.33 to 0.40	Wolc et al. (2009)
Body weight[1]	REML	0.046 (0.005) to 0.047 (0.004)	0.49 to 0.57	-0.45 (0.040) to -0.41 (0.032)	Mulder et al. (2009)
Eggshell color	REML	0.010 (0.001) to 0.011 (0.003)	0.26 (0.039) to 0.28 (0.019)	-0.057 (0.084) to 0.43 (0.10)	Mulder, Visscher and Fablet (2016)

h^2_v: heritability of residual variance or uniformity; GCV_E: genetic coefficient of variation of residual variance or uniformity; r_{mv}: genetic correlation between mean and residual variance; REML: restricted estimate maximum likelihood; when availabele, standard errors are presented between brackets.
[1]: estimates separated by sex

Table 1d. Summary of the studies on uniformity in fish

Traits	Method	h_v^2	GCV_E	r_{mv}	Reference
Body weight	REML	0.024 (0.006)	0.37	-0.16 (0.039)	Janhunen et al. (2012b)
Body weight	REML	-	-	0.42	Sonesson, Ødegård and Rönnegård (2013)
Body weight[1]	REML	-	-	-0.17 to 0.14	Sonesson, Ødegård and Rönnegård (2013)
Body weight	REML	0.010 to 0.011	0.190 to 0.211	0.30 (0.15) to 0.79 (0.13)	Sae-Lim et al. (2015)
Body weight[1]	REML	0.010 to 0.024	0.174 to 0.296	-0.62 (0.21) to -0.83 (0.10)	Sae-Lim et al. (2015)
Body size and weight	REML	-	0.39 to 0.58	0.11 (0.16) to 0.60 (0.09)	Marjanovic et al. (2016)
Body weight	REML	0.036 (0.019) to 0.038 (0.020)	0.48 (0.114) to 0.523 (0.116)	0.95	Sae-Lim et al. (2017)
Body weight[1]	REML	0.014 (0.013) to 0.015 (0.013)	0.298 (0.100) to 0.299 (0.095)	-0.093 to 0.024	Sae-Lim et al. (2017)

h_v^2: heritability of residual variance or uniformity; GCV_E: genetic coefficient of variation of residual variance or uniformity; r_{mv}: genetic correlation between mean and residual variance; REML: restricted estimate maximum likelihood; when availabele, standard errors are presented between brackets.
[1]: using a transformed dataset

Table 1e. Summary of the studies on uniformity in cattle

Traits	Method	h_v^2	GCV_E	r_{mv}	Reference
Beef production traits	REML	0.004 (0.002) to 0.099 (0.006)	0.12 to 0.70	-0.21 to 0.42	Neves, Carvalheiro and Queiroz (2011)
Beef growth traits[1]	REML	0 to 0.058	0 to 0.787	-0.01 to 0.36	Neves, Carvalheiro and Queiroz (2012)
Milk yield and somatic cell score	REML	0.005	0.20	0.38 to 0.60	Rönnegård et al. (2013)
Milk yield, somatic cell score and fatty acids	REML	0.001 to 0.004	0.12 to 0.17	0.22 (0.13;0.30) to 0.47 (0.40;0.54)	Vandenplas et al. (2013)
Calving to first service[2]	REML	0.005	0.12	0.69 (0.09)	Ghiasi and Felleki (2016)

h_v^2: heritability of residual variance or uniformity; GCV_E: genetic coefficient of variation of residual variance or uniformity; r_{mv}: genetic correlation between mean and residual variance; REML: restricted estimate maximum likelihood; when availabele, standard errors are presented between brackets.
[1]: estimates separated by sex
[2]: using a transformed dataset

1.2.4 Genomic studies on uniformity

Currently, the wide availability of genomic information has promoted substantial gain in knowledge about the genetic mechanisms underlying important traits and accuracy of EBV, which increases the response to selection (MEUWISSEN; HAYES; GODDARD, 2001). Adding genomic data to investigate uniformity may lead to the discovery of variance-controlling genes through genome-wide association study (GWAS). In livestock, few GWAS have been developed so far

for uniformity, and a lack of consensus about the appropriate response variable used in these studies is an additional challenge.

Residual variance per individual from DHGLM were used as response variable by Mulder et al. (2013) to identify genomic regions associated with residual variance of somatic cell score in dairy cattle. Deregressed EBV (dEBV), using solutions from a DHGLM, were used by Sell-Kubiak et al. (2015c) to identify genomic regions associated with variability of litter size in pigs (SELL-KUBIAK et al., 2015a). Using a simpler way to measure uniformity, Wolc et al. (2012) and Wang et al. (2017) used phenotypic standard deviation variation of egg weight in chickens and coefficient of birth weight in pigs, respectively, as response variables in traditional linear models. Such variety of response variables can be explained because the trait, data structure and availability of genotypes determine the way to quantity uniformity, respecting the principle of having within-individual variance based on repeated observations per animal or within-family variance based on large offspring groups per family.

Similar genetic parameters for mean and residual variance were estimated when using either pedigree-based or genomic relationship matrices (GRM; MULDER et al., 2013; SELL-KUBIAK et al., 2015b). However, benefits of the GRM compared to the pedigree-based relationship were greater in terms of accuracy of EBV for uniformity of harvest weight in salmon, as found Sae-Lim et al. (2017). A previous study showed an increase in accuracy of EBV for mean and residual variance of body weight in pigs when considering genomic information, although the difference was not significant, probably due to the small size of the reference population used for genomic prediction (SELL-KUBIAK et al., 2015b).

Thus, unraveling the genetic and genomic basis of heterogeneity of residual through quantitative and genomic analysis will help to: find a better way to improve uniformity, understand the biology behind it and increase selection response by increasing accuracy of EBV for residual variance.

1.3 Objectives

- Investigate the genetic heterogeneity of residual variance on yearling weight (YW) in a Nellore beef cattle population, comparing the results from two different approaches (two-step and double hierarchical generalized linear model –

DHGLM) and evaluating the effectiveness of phenotypic transformation to accommodate scale effects.

- Identify genomic regions associated with within-family residual variance of YW in Nellore cattle, using different response variables formed by residuals and solutions of DHGLM assuming null and non-null genetic correlation between mean and residual variance.

1.4 References

ARGENTE, M. J.; CALLE, E. W.; GARCÍA, M. L.; BLASCO, A. Correlated response in litter size components in rabbits selected for litter size variability. **Journal of Animal Breeding and Genetics**, Berlin, v. 134, p. 505-511, 2017.

BLASCO, A.; MARTÍNEZ-ÁLVARO, M.; GARCÍA, M. L.; IBÁÑEZ-ESCRICHE, N.; ARGENTE, M. J. Selection for environmental variance of litter size in rabbits. **Genetics Selection Evolution**, Jouy-en-Josas, v. 49, p. 48, 2017.

BOX, G. E. P; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society. Series B (Methodological),** London, p. 211, 252, 1964.

CARVALHEIRO, R. Genomic selection in Nelore cattle in Brazil. **In Proceedings of the 10th World Congress on Genetic Applied to Livestock Production**: 17-22 August 2014; Vancouver.

COTTLE, D. J.; GILMOUR, A. R.; PABIOU, T.; AMER, P. R.; FAHEY, A. G. Genetic selection for increased mean and reduced variance of twinning rate in Belclare ewes. **Journal of Animal Breeding and Genetics**, Berlin, v. 133, p. 126–37, 2016.

FALCONER, D.S.; MACKAY, T.F.C. **Quantitative Genetics**, 1996.

FELLEKI, M.; LEE, D.; LEE, Y.; GILMOUR, A. R.; RÖNNEGÅRD, L. Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. **Genetics Research**, Cambridge, v. 94, p. 307–317, 2012.

FELLEKI, M.; LUNDEHEIM, N. Genetic heteroscedasticity of teat count in pigs. **Journal of Animal Breeding and Genetics**, Berlin, v. 132, p. 392-398. 2015.

FORMOSO-RAFFERTY, N.; CERVANTES, I.; IBÁÑEZ-ESCRICHE, N.; GUTIÉRREZ, J. P. Correlated genetic trends for production and welfare traits in a mouse population divergently selected for birth weight environmental variability. **Animal,** Cambridge, v. 10, p. 1770–1777, 2016.

FORMOSO-RAFFERTY, N.; CERVANTES, I.; IBÁÑEZ-ESCRICHE, N.; GUTIÉRREZ, J. P. Genetic control of the environmental variance for birth weight in seven generations of a divergent selection experiment in mice. **Journal of Animal Breeding and Genetics**, Berlin, v. 133, p. 227–237, 2016b.

GARREAU, H.; BOLET, G.; LARZUL, C.; ROBERT-GRANIÉ, C.; SALEIL, G.; SANCRISTOBAL, M.; BODIN, L. Results of four generations of a canalising selection for rabbit birth weight. **Livestock Science**, Amsterdam, v. 119, p. 55–62, 2008.

GHIASI, H.; FELLEKI, M. Joint estimation of (co) variance components and breeding values for mean and dispersion of days from calving to first service in Holstein cow. **Animal Production Science**, Melbourne, v. 57, p. 760-766, 2016.

GILMOUR, K. M.; DIBATTISTA, J. D.; THOMAS, J. B. Physiological Causes and Consequences of Social Status in Salmonid Fish. **Integrative and Comparative Biology,** McLean, v. 45, p. 263–273, 2005.

GORSSEN, W.; BUYS, N.; JANSSENS, S. Genetic parameters for uniformity of fattening pigs' growth traits. **In Proceedings of the 11th World Congress on Genetic Applied to Livestock Production**: 11-16 February 2018; Auckland.

HILL, W. G.; ZHANG, X. S. Effects of phenotypic variability of directional selection arising through genetic differences in residual variability. **Genetical Research**, London, v. 83, p. 121–132, 2004.

HILL, W. G.; MULDER, H. A. Genetic analysis of environmental variation. **Genetics Research**, Cambridge, v. 92, p. 381–395, 2010.

HOAGLIN, D. C.; WELSCH, R. E. The hat matrix in regression and ANOVA. **The American Statistician**, Washington, v.. 32, p. 17–22, 1978.

HOHENBOKEN, W. D. The Manipulation of Variation in Quantitative Traits : A Review of Possible Genetic Strategies. **Journal of Animal Science**, Champaign, v. 60, p. 101–110, 1985.

IBÁÑEZ-ESCRICHE, N.; VARONA, L.; SORENSEN, D.; NOGUERA, J. L. A study of heterogeneity of environmental variance for slaughter weight in pigs. **Animal**, Cambridge, v. 2, p. 19–26, 2008.

JANHUNEN, M.; KAUSE, A.; JÄRVISALO, O. Costs of being extreme - Do body size deviations from population or sire means decrease vitality in rainbow trout? **Aquaculture,** Amsterdam**,** v. 370, p. 123–129, 2012.

JANHUNEN, M.; KAUSE, A.; VEHVILÄINEN, H.; JÄRVISALO, O. Genetics of microenvironmental sensitivity of body weight in rainbow trout (*Oncorhynchus mykiss*) selected for improved growth. **PLoS One**, San Francisco, v. 7, p. e38766, 2012b.

KAUSE, A.; JANHUNEN, M.; VEHVILÄINEN, H.; KOSKINEN, H.; NOUSIAINEN, A. Improving uniformity of growth by mating and selection strategies in rainbow trout. **In Proceedings of the 10th World Congress on Genetic Applied to Livestock Production**: 17-22 August 2014; Vancouver.

KHAW, H. L.; PONZONI, R. W.; YEE, H. Y.; BIN AZIZ; M. A., MULDER, H. A., MARJANOVIC, J.; Bijma, P. Genetic variance for uniformity of harvest weight in Nile tilapia (Oreochromis niloticus). **Aquaculture**, Amsterdam, v. 451, p. 113-120, 2016.

LEE, Y.; NELDER, J. A.; ENGEL, B.; KEEN, A.; CLAYTON, D.; LONDFORD, N. T.; et al. Hierarchical generalized linear models. **Journal of the Royal Statistical Society. Series B (Methodological),** London, p. 619-679, 1996.

MACKAY, T. F. C.; LYMAN, R. F. Drosophila bristles and the nature of quantitative genetic variation. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, London, v. 360, p. 1513–1527, 2005.

MARJANOVIC, J.; MULDER, H. A.; KHAW, H. L.; BIJMA, P. Genetic parameters for uniformity of harvest weight in the GIFT strain of Nile tilapia estimated using double hierarchical generalized linear models. **Genetics Selection Evolution,** Jouy-en-Josas, v. 48, p. 41, 2016.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics,** Baltimore**,** v. 157, p. 1819–1829, 2001.

MILLIGAN, B. N.; FRASER, D.; KRAMER, D. L. Within-litter birth weight variation in the domestic pig and its relation to pre-weaning survival, weight gain, and variation in weaning weights. **Livestock Production Science,** Amsterdam, v. 76, p. 181–191, 2002.

MULDER, H. A.; BIJMA, P.; HILL, W. G. Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. **Genetics**, Baltimore, v. 175, p. 1895-1910, 2007.

MULDER, H. A.; HILL, W. G.; VEREIJKEN, A.; VEERKAMP, R. F. Estimation of genetic variation in residual variance in female and male broilers. **Animal,** Cambridge, v. 3, p. 1673–1680, 2009.

MULDER, H. A.; CRUMP, R.; CALUS, M. P. L.; VEERKAMP, R. Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms. **Journal of Dairy Science**, Champaign, v. 96, p. 306–17, 2013.

MULDER, H. A.; VISSCHER, J.; FABLET, J. Estimating the purebred–crossbred genetic correlation for uniformity of eggshell color in laying hens. **Genetics Selection Evolution**, Jouy-en-Josas, v. 48, p. 39, 2016.

NEVES, H. H. R.; CARVALHEIRO, R.; CARDOSO, V.; FRIES, L. A.; QUEIROZ, S. A. Acasalamento dirigido para aumentar a produção de animais geneticamente superiores e reduzir a variabilidade da progênie em bovinos. **Revista Brasileira de Zootecnia**, Viçosa, v. 38, p. 1201–1204, 2009.

NEVES, H. H. R.; CARVALHEIRO, R.; QUEIROZ, S. A. Genetic variability of residual variance of production traits in Nellore beef cattle. **Livestock Science,** Amsterdam**,** v. 142, p. 164–169, 2011.

NEVES, H. H. R.; CARVALHEIRO, R.; QUEIROZ, S. A. Genetic and environmental heterogeneity of residual variance of weight traits in Nellore beef cattle. **Genetics Selection Evolution,** Jouy-en-Josas, v. 44, n. 1, p.19, 2012.

RÖNNEGÅRD, L.; FELLEKI, M.; FIKSE, F.; MULDER, H. A.; STRANDBERG, E. Genetic heterogeneity of residual variance—Estimation of variance components using double hierarchical generalized linear models. **Genetics Selection Evolution**, Jouy-en-Josas, v. 42, p. 1–10, 2010.

RÖNNEGÅRD, L.; FELLEKI, M.; FIKSE, W. F.; MULDER, H. A.; STRANDBERG, E. Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. **Journal of Dairy Science**, Champaign, v. 96, p. 2627-2636, 2013.

ROS, M.; SORENSEN, D.; WAAGEPETERSEN, R.; DUPONT-NIVET, M.; SANCRISTOBAL, M.; BONNET, J. C.; et al. Evidence for genetic control of adult weight plasticity in the snail Helix aspersa. **Genetics**, Baltimore, v. 168, p. 2089–2097, 2004.

ROWE, S. J.; WHITE, I. M. S.; AVENDANO, S.; HILL, W. G. Genetic heterogeneity of residual variance in broiler chickens. **Genetics Selection Evolution,** Jouy-en-Josas, v. 38, p. 617–635, 2006.

SAE-LIM, P., KAUSE, A.; JANHUNEN, M.; VEHVILÄINEN, H.; KOSKINEN, H.; GJERDE, B.; LILLEHAMMER, M.; MULDER, H. A. Genetic (co)variance of rainbow trout (*Oncorhynchus mykiss*) body weight and its uniformity across production environments. **Genetics Selection Evolution**, Jouy-en-Josas, v. 47, p. 46, 2015.

SAE-LIM, P.; KAUSE, A.; LILLEHAMMER, M.; MULDER, H. A. Estimation of breeding values for uniformity of growth in Atlantic salmon (*Salmo salar*) using pedigree relationships or single-step genomic evaluation. **Genetics Selection Evolution**, Jouy-en-Josas, v. 49, p. 33, 2017.

SAE-LIM, P.; JAKOBSEN, J. H.; MULDER, H. A. Uniformity in birth weight is heritable in Norwegian White Sheep. **In Proceedings of the 11th World Congress on Genetic Applied to Livestock Production**: 11-16 February 2018; Auckland.

SANCRISTOBAL-GAUDY, M.; ELSEN, J. M.; BODIN, L.; CHEVALET, C. Prediction of the response to selection for canalization of a continuous trait in animal breeding. **Genetics Selection Evolution**, Jouy-en-Josas, v. 39, p. 423–451, 1998.

SANCRISTOBAL-GAUDY, M.; BODIN, L.; ELSEN, J. M.; CHEVALET, C. Genetic components of litter size variability in sheep. **Genetics Selection Evolution**, Jouy-en-Josas, v. 33, p. 249–271, 2001.

SELL-KUBIAK, E.; BIJMA, P. KNOL, E. F.; MULDER, H. A. Comparison of methods to study uniformity of traits: Application to birth weight in pigs. **Journal of Animal Science**, Champaign**,** v. 93, p. 900–911, 2015.

SELL-KUBIAK, E.; WANG, S.; KNOL, E. F.; MULDER, H. A. Genetic analysis of within-litter variation in piglets' birth weight using genomic or pedigree relationship matrices. **Journal of Animal Science**, Champaign**,** v. 93, p. 1471-1480, 2015b.

SELL-KUBIAK, E.; DUIJVESTEIJN, N.; LOPES, M. S.; JANSS, L. L. G.; KNOL, E. F.; BIJMA, P.; MULDER, H. A. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. **BMC Genomics**, London, v. 16, p. 1049, 2015c.

SONESSON, A. K.; ØDEGÅRD, J.; RÖNNEGÅRD, L. Genetic heterogeneity of within family variance of body weight in Atlantic salmon (*Salmo salar*). **Genetics Selection Evolution**, Jouy-en-Josas, v. 45, p. 41, 2013.

SORENSEN, D.; WAAGEPETERSEN, R. Normal linear models with genetically structured residual variance heterogeneity: a case study. **Genetics Research**, Cambridge, v. 82, p. 207–222, 2003.

UNITED STATES DEPARTMENT OF AGRICULTURE - USDA. Livestock and Poultry: World Markets and Trade. Available in: <http://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf>. Acessed on: 31 Aug. 2017.

VANDENPLAS, J.; BASTIN, C.; GENGLER, N.; MULDER, H. A. Genetic variance in micro-environmental sensitivity for milk and milk quality in Walloon Holstein cattle. **Journal of Dairy Science**, Champaign, v. 96, p. 5977–5990, 2013.

WANG, Y.; DING, X.; TAN, Z.; NING, C.; XING, K.; YANG, T.; PAN, Y.; SUN, D.; WANG, C. Genome-Wide Association Study of Piglet Uniformity and Farrowing Interval. **Frontiers in genetics**, Lausanne, 8, 194, 2017.

WOLC, A.; WHITE, I. M. S.; AVENDANO, S.; HILL, W. G. Genetic variability in residual variation of body weight and conformation scores in broiler chickens. **Poultry Science**, Danville, v. 88, p. 1156-1161, 2009.

WOLC, A.; ARANGO, J.; SETTAR, P.; FULTON, J. E.; O'SULLIVAN, N. P.; PREISINGER, R.; HABIER, D.; FERNANDO, R.; GARRICK, D. J.; HILL, W. G.; DEKKERS, J. C. M. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. **Animal Genetics**, New York, v. 43, p. 87-96, 2012.

YANG, Y.; CHRISTENSEN, O.; SORENSEN, D. A. Analysis of a genetically structured variance heterogeneity model using the Box–Cox transformation. **Genetics Research**, Cambridge, v. 93, p. 33-46, 2011.

ZHANG, T.; WANG, L. G.; SHI, H. B.; YAN, H.; ZHANG, L. C.; LIU, X.; PU, L. et al. Heritabilities and genetic and phenotypic correlations of litter uniformity and litter size in Large White sows. **Journal of Integrative Agriculture**, Beijing, v. 15, n. 4, p. 848-854, 2016.

CHAPTER 2 - Genetic control of residual variance of yearling weight in Nellore beef cattle[a]

Abstract – There is evidence for genetic variability in residual variance of livestock traits, which offers the potential for selection for increased uniformity of production. Different statistical approaches have been employed to study this topic, however, little is known about the concordance between them. The aim of our study was to investigate the genetic heterogeneity of residual variance on yearling weight (YW; 291.15 ± 46.67) in a Nellore beef cattle population, compare the results of the statistical approaches two-step and double hierarchical generalized linear model (DHGLM), and evaluate the effectiveness of power transformation to accommodate scale differences. The comparison was based on genetic parameters, accuracy of estimated breeding values (EBV) for residual variance and cross-validation to assess predictive performance of both approaches. A total of 194,628 yearling weight records from 625 sires were used in the analysis. The results supported the hypothesis of genetic heterogeneity of residual variance on YW in Nellore beef cattle and the opportunity of selection, measured through genetic coefficient of variation of residual variance (0.10-0.12 for two-step and 0.17 for DHGLM, using untransformed dataset). However, low estimates of genetic variance associated with positive genetic correlations between mean and residual variance (about 0.20 for two-step and 0.76 for DHGLM for untransformed dataset) limit the genetic response to selection for uniformity of production when simultaneously increasing YW itself. Moreover, large sire families are needed to obtain accurate estimates of genetic merit for residual variance as indicated by the low heritability estimates (< 0.007). Box-Cox transformation was able to decrease the dependence of the variance on the mean and decreased the estimates of genetic parameters for residual variance. The transformation reduced but did not eliminate all the genetic heterogeneity of residual variance, highlighting its presence beyond the scale effect. DHGLM showed higher predictive ability of EBV for residual variance and, therefore, should be preferred over the two-step approach.

[a]Article published in the “Journal of Animal Science” doi: 10.2527/jas.2016.1326

Keywords: beef cattle, cross-validation, double hierarchical generalized linear model, genetic heterogeneity of residual variance, uniformity of production, yearling weight

2.1 Introduction

The demand for more uniform animal production, stimulated mainly by economic incentives, has increased the interest in reducing variability of traits through selection. Evidences of genetic control on the residual variance of some traits or so-called genetic heterogeneity of residual variance have made this a potential response variable in genetic studies on uniformity (e.g. Hill and Mulder, 2010). In beef cattle, more uniformity has been sought in weight and carcass related traits, such as yearling weight (YW). Neves et al. (2012), using a two-step approach, observed the existence of a genetic component on residual variance of YW.

Genetic heterogeneity of residual variance has been studied two ways. Mulder et al. (2009), present a two-step approach in which the variance components were first estimated for the mean of the trait and afterwards the log-transformed squared estimated residuals were used as response variable in the residual variance model. Recently, Rönnegård et al. (2010) and Felleki et al. (2012) proposed the double hierarchical generalized linear model (DHGLM), simultaneously adjusting two sets of mixed model equations, one on the level of the mean and the other on the level of the residual variance. In this approach, a correction is used to adjust for the imprecision of using estimated residuals as a proxy for the residual variance at the individual level.

However, to our knowledge, no study has directly compared the two-step approach with the DHGLM. The aim of our study was to investigate the genetic heterogeneity of residual variance on YW in a Nellore beef cattle population, compare the results of the two-step approach and the DHGLM, and evaluate the effectiveness of phenotypic transformation to accommodate scale effects.

2.2 Materials and Methods

Animal Care and Use Committee approval was not required for this study because the data were obtained from an existing database.

2.2.1 Data

Data of YW, in kg, available in the Aliança Nelore database (www.gensys.com.br) were used in this study. This trait was measured around 550

days of age, with a minimum of 331 days and a maximum of 629 days, in animals born between 1986 and 2013, which were raised on pasture in herds from Brazil and Paraguay.

The contemporary groups (CG) were formed by concatenating sex, herd, year and season of birth and management group identification (considering weaning and yearling groups). Two birth seasons were defined: September to March and April to August. Genetic connectedness among CG was evaluated using the AMC software (Roso and Schenkel, 2006), based on direct genetic links between CG due to common ancestors (Fries, 1998). Only data from CG with at least ten direct genetic links to the main set (largest number of direct genetic links between groups) of connected data were maintained. During the data editing, only records within 3.5 standard deviations from the mean, CG with at least 25 animals and sires with at least 50 progeny were kept in subsequent analyses (described later).

Phenotypic variance of YW was 69% higher in males than females. Thus, additional analyses for each sex were performed, applying the same data editing criteria as applied to the full database. A summarized description of each data set is shown in Table 1.

To deal with the possible existence of scaling effects, Box-Cox transformation was used (Box and Cox, 1964). The transformed data were obtained as: $y^{(\lambda)} = \left[(y^{\lambda}-1)/\lambda\right]$, where *y* corresponds to the phenotypic records of YW and λ is the transformation parameter that allows obtaining an approximately normal distribution of the transformed phenotypes ($y^{(\lambda)}$). The optimal λ (-0.33, 0.05 and -0.386 for full, male and female datasets, respectively) was obtained under a model including fixed effect of CG and age (linear and quadratic, nested within sex) as covariate, using the MASS package (Venables and Ripley, 2002) in R (R Development Core Team, 2016).

2.2.2 General description of the models

Heterogeneity of residual variance of untransformed and Box–Cox-transformed YW data was studied under two-step and DHGLM approaches. For all analyses, single-trait sire models were adopted. A sire model was chosen because a single observation per animal was available and, therefore, an animal model would

provide biased variance component estimates under this condition (Mulder et al., 2013; Sonesson et al., 2013). All analyses were carried out using the DMU package (Madsen and Jensen, 2013).

Table 1. Data structure and descriptive statistics of yearling weight according to the dataset

Dataset	Full	Male	Female
Data structure			
Number of records	194,628	82,994	78,271
Number of contemporary groups (CG)	3,905	1,708	1,615
Number of sires	625	303	288
Number of dams	143,918	71,600	67,908
Number of herds	276	218	220
Average number of progeny per sire	311	274	272
Descriptive statistics			
Mean, kg	291.15	309.67	273.43
Standard deviation, kg	46.67	48.29	37.17
Skewness	0.62	0.44	0.52
Kurtosis	0.39	0.15	0.24

2.2.2.1 Two-step approach

The following sire model was fitted in the first step:

$$\mathbf{y = Xb + Zs + e,} \quad (1)$$

where **y** is the vector of observations, **b** is the vector of fixed effects (CG) and covariates (linear and quadratic effects of animal age, nested within sex); **s** is the vector of random sire genetic effects, with $\mathbf{s} \sim N(0, \mathbf{A}\sigma_s^2)$ and **e** is the vector of random residual effects, assuming that $\mathbf{e} \sim N(0, \mathbf{R}\sigma_e^2)$. **X** and **Z** are corresponding incidence matrices, **A** is the numerator relationship matrix, **R** is the residual variance matrix, σ_s^2 and σ_e^2 are sire genetic variance and residual variance, respectively. Either homogeneous (HOM) or heterogeneous residual variance among sire families (HET) was assumed in this step. Thus, **R** is a diagonal matrix, whose diagonal elements were either all equal to σ_e^2 (HOM model) or equal to $\sigma_{e_i}^2$ for sire family *i* (HET model).

The residuals associated with each observation, estimated from model (1), were then used to obtain the log-transformed squared estimated residuals ($\ln(\hat{e}^2)$), which were analyzed as a measure of the residual variance according to the following model (second step):

$$\mathbf{\ln\left(\hat{e}^2\right) = Xb_v + Zs_v + e_v,} \quad (2)$$

where $\mathbf{ln(\hat{e}^2)}$ is the vector of log transformed squared estimated residuals for YW; $\mathbf{b_v}$ is the vector of fixed effects (CG and linear and quadratic effects of animal age, nested within sex); $\mathbf{s_v}$ is the vector of random sire genetic effects, with $\mathbf{s_v} \sim N(0, \mathbf{A}\sigma^2_{s_v})$, $\mathbf{e_v}$ is the vector of random residual effects, with $\mathbf{e_v} \sim N(0, \mathbf{I}\sigma^2_{e_v})$, $\mathbf{I}$ is an identity matrix, $\sigma^2_{s_v}$ and $\sigma^2_{e_v}$ are sire genetic variance and residual variance at the level of the log-transformed residuals, respectively.

HOM and HET models (first step) were fitted using the Average Information Restricted Maximum Likelihood (AI-REML) and Reversible Jump Monte Carlo sampling (RJMC) modules of the DMU package, respectively. RJMC module was used because it allows adjusting heterogeneous residual variance per sire. For HET model (first step), chains of 250,000 (untransformed data) and 500,000 (transformed data) cycles were required for convergence. No burn-in period was used and samples were saved every 50 cycles. The convergence was evaluated through visual inspection and Geweke test (Geweke, 1992), as implemented in the CODA R package (Plummer et al., 2006). In the second step, AI-REML module was used to fit the model described in equation (2).

2.2.2.2 DHGLM

Rönnegård et al. (2010) introduced a DHGLM algorithm to study genetic heterogeneity of residual variance. They proposed fitting simultaneously a linear model for the mean and a gamma generalized linear mixed model (GLMM) for the residual variance, by using an iterative procedure with a log link function to account for the non-normality of the response variable at the level of residual variance.

Later, Felleki et al. (2012) extended this approach by using a linearized response variable instead of using a gamma GLMM to model the residual variance, allowing to fit a bivariate linear mixed model and to estimate the correlation between genetic effects on mean and residual variance. Such linearized response variable is defined as: $\psi_i = \log(\hat{\sigma}^2_{e_i}) + \left(\left\{\left[\hat{e}^2_i/(1 - h_i)\right] - \hat{\sigma}^2_{e_i}\right\} / \hat{\sigma}^2_{e_i}\right)$, where $\hat{\sigma}^2_{e_i}$ is the predicted residual variance for observation *i* obtained in the previous iteration, $\hat{e}^2_i$ is the squared residual from the mean model estimated for observation *i* and h_i is the leverage for

observation *i*, i.e. the i^{th} diagonal element of the hat matrix (Hoaglin and Welsch, 1978). Thus, the DHGLM fitted in the present study can be represented as follows:

$$\begin{bmatrix}\mathbf{y}\\ \boldsymbol{\psi}\end{bmatrix} = \begin{bmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{0} & \mathbf{X_v}\end{bmatrix}\begin{bmatrix}\mathbf{b}\\ \mathbf{b_v}\end{bmatrix} + \begin{bmatrix}\mathbf{Z} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z_v}\end{bmatrix}\begin{bmatrix}\mathbf{s}\\ \mathbf{s_v}\end{bmatrix} + \begin{bmatrix}\mathbf{e}\\ \mathbf{e_v}\end{bmatrix}, \quad (3)$$

where **y** and **ψ** are vectors of response variables for the mean and residual variance models, respectively, and other effects are the same as described previously for the two-step approach. It was assumed that: $\begin{bmatrix}\mathbf{s}\\ \mathbf{s_v}\end{bmatrix} \sim N\left(0, \begin{bmatrix}\sigma_s^2 & \sigma_{s,s_v}\\ \sigma_{s,s_v} & \sigma_{s_v}^2\end{bmatrix} \otimes \mathbf{A}\right)$, and $\begin{bmatrix}\mathbf{e}\\ \mathbf{e_v}\end{bmatrix} \sim N\left(0, \begin{bmatrix}\mathbf{W^{-1}}\sigma_e^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{W_v^{-1}}\sigma_{e_v}^2\end{bmatrix} \otimes \mathbf{I}\right)$, where $\mathbf{W}=\text{diag}(\hat{\psi}^{-1})$ and $\mathbf{W_v}=\text{diag}\left(\frac{1-h}{2}\right)$ contain the residual variance per observation, since σ_e^2 and $\sigma_{e_v}^2$ are scaling variance, which are expected to be equal to 1. The model described in equation (3) was fitted iteratively, updating **ψ** , **W** and $\mathbf{W_v}$ in each iteration until convergence (Felleki et al., 2012). Convergence was assumed to be achieved when the relative difference in the estimates of variance components obtained between successive interactions was smaller than 10^{-6}. The AI-REML module of the DMU package was used to perform the DHGLM analyses.

2.2.3 Genetic Parameters

Variance components obtained under two-step approach and DHGLM were used to compute the heritability of residual variance (h_v^2) and the genetic coefficient of variation of residual variance (GCV_E). In general, h_v^2 is required to assess the accuracy of EBV for residual variance (EBV_v) and GCV_E can be used to assess the opportunity to change the variability through selection.

In the two-step approach, h_v^2 was calculated as $h_v^2 = \sigma_{a_{v,add}}^2 / (2\sigma_P^4 + 3\sigma_{a_{v,add}}^2)$, where $\sigma_{a_{v,add}}^2$ is the additive genetic variance estimated for the residual variance on the additive scale and σ_P^4 is the squared phenotypic variance (Mulder et al. 2007). The $\sigma_{a_{v,add}}^2$ was obtained as: $\sigma_{a_{v,add}}^2 = h_{res}^2 2(\bar{\sigma}_{\hat{e}}^2)^2$, in which h_{res}^2 is the heritability of the response variable used in the residual variance model, and $\bar{\sigma}_{\hat{e}}^2$ is the average residual variance in the mean model (Mulder et al. 2009). The standard error of h_v^2 was

calculated according to Mulder et al. (2009). The GCV_E was calculated as: $GCV_E = \sigma_{a_{v,add}} / \overline{\sigma}^2_{\hat{e}}$, where $\sigma_{a_{v,add}}$ is the standard deviation of the additive genetic variance for the residual variance on the additive scale. Note that the residual variance in these calculations is on the animal model scale rather than sire model scale, so it is calculated as $\overline{\sigma}^2_{\hat{e}} - 3\sigma^2_s$.

For DHGLM, h^2_v and GCV_E were obtained as in the two-step approach, however, the $\sigma^2_{a_{v,add}}$ was calculated as: $\sigma^2_{a_{v,add}} = \sigma^4_{e,exp} \exp\left(2\sigma^2_{a_{v,exp}}\right) - \sigma^4_{e,add}$, where $\sigma^2_{e,exp} = \left[\overline{(1/\mathbf{W})}\,\sigma^2_{\hat{e}}\right] / \left(\exp\left(0.5\sigma^2_{a_{v,exp}}\right)\right)$, $\sigma^2_{a_{v,exp}}$ is the additive genetic variance estimated for the residual variance of YW on the exponential scale and $\sigma^2_{e,add} = \left[\overline{(1/\mathbf{W})}\,\sigma^2_{\hat{e}}\right]$, where $\sigma^2_{\hat{e}}$ and $\overline{(1/\mathbf{W})}$ are residual variance and the average of the reciprocal weights obtained in the mean model, respectively. Here, the residual variance also was considered on the animal model scale rather than sire model scale, i.e. $\overline{\sigma}^2_{\hat{e}} - 3\sigma^2_s$.

As we fitted sire models, it was assumed that the estimates of sire genetic variance ($\sigma^2_{\hat{s}}$) represented one quarter of the additive genetic variance (σ^2_a), for both mean and residual variance models. Therefore, $\sigma^2_{a_{v,exp}}$ was obtained after multiplying the estimated sire genetic variance by 4. The standard errors for h^2_v and GCV_E were obtained according to Mulder et al. (2016).

Moreover, in the two-step approach, we calculated the Pearson's correlation coefficient between sire EBV on mean and residual variance (EBV_m and EBV_v, respectively)**.** Such procedure was adopted because the direct estimation of the r_{mv} is not possible under this approach.

2.2.4 Comparison between two-step approach and DHGLM

The comparison between two-step approach and DHGLM was realized similarly as in Sell-Kubiak et al. (2015), i.e. in terms of estimates of variance components and genetic parameters, EBV_v and the accuracies obtained under each approach. The accuracies of EBV_v for each sire were calculated using estimates of prediction error variance (PEV) and the sire genetic variance for the residual variance.

Moreover, a cross-validation procedure was carried out to compare the predictive ability of these approaches. The entire dataset was split into two subsets to form the training and validation sets, so that the training set was composed of phenotypic information of about 50% of the progeny of each sire, and the remaining progeny used as the validation set. Data from CG with at least 25 animals (for the training set) were kept for cross-validation. The predictive ability of each model was assessed by the Pearson's correlation between sire EBV obtained from training and validation sets. Ten replicates of the aforementioned data splitting strategy were carried out. Predictive ability was evaluated for both mean and the residual variance. Correlation estimates obtained across all replicates were summarized using mean and standard error.

Additional analyses for the DHGLM approach were performed assuming null correlation between EBV_m and EBV_v, to give a fairer comparison between the two-step and DHGLM approaches, as this information is not used by the two-step approach.

2.3 Results

The linear dependence of variance on mean of YW (scale effect) was reduced after Box-Cox transformation (Fig. 1). The coefficient of determination of the linear regression (R^2) of the within sire family phenotypic variance on sire family mean decreased from 0.29 to 0.04. In addition, the slight skewness of the original data was absent after the transformation, which resulted in a distribution closer to the normal (Fig. 2).

Estimates of variance components and genetic parameters are presented in Table 2. In general, similar estimates of heritabilities were obtained for untransformed and transformed data, especially for the mean. Moderate estimates of h^2 were obtained for YW (ranging from 0.31 to 0.37), being the lowest values observed with DHGLM and two-step using the HET model. When each sex was analyzed separately, higher estimates of h^2 were obtained for females than for males in the two-step approach, and similar values were observed for DHGLM.

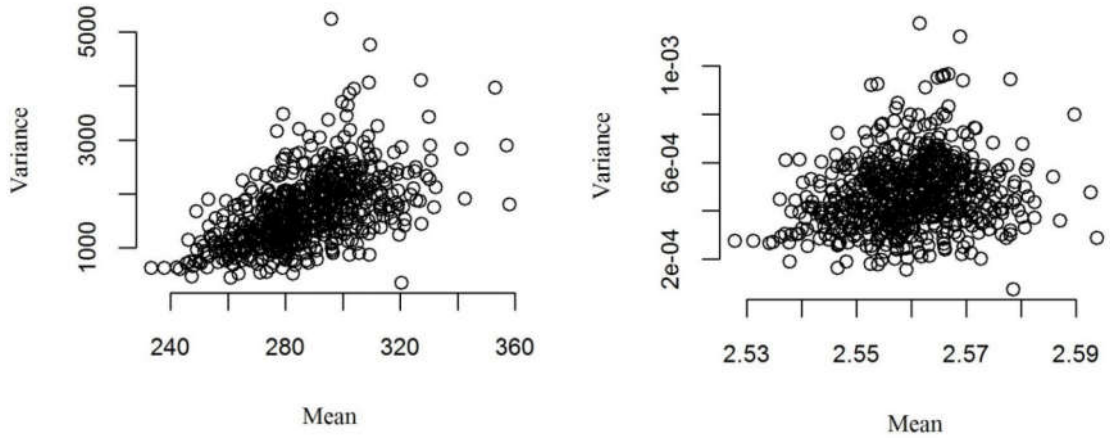

Figure 1. Relationship between mean and variance of yearling weight (YW) within sire families. Left: untransformed data. Right: Box-Cox transformed data.

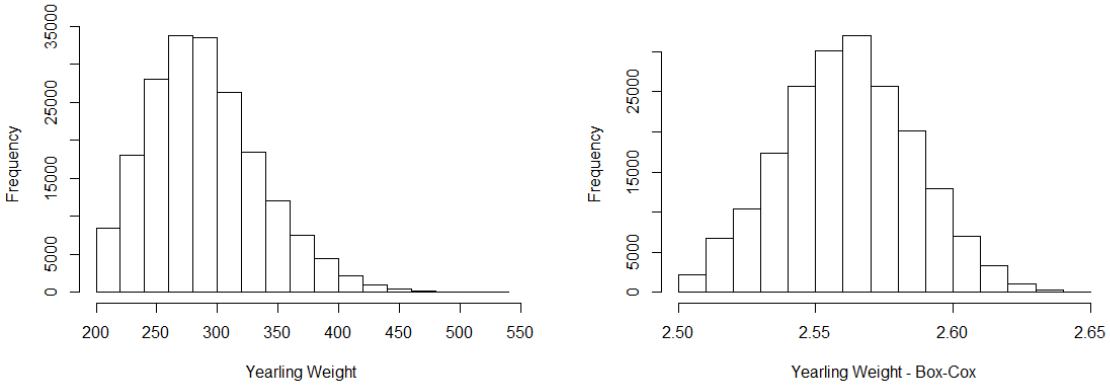

Figure 2. Distribution of yearling weight (YW) before (left) and after Box-Cox transformation (right).

For the residual variance part, very low h_v^2 (< 0.01) were estimated for YW in all models, being even lower after transformation. Compared to the two-step approach, the DHGLM resulted in only slightly higher estimates of h_v^2 (0.007 and 0.003 for untransformed and transformed data, respectively). All estimates of GCV_E were lower than 0.20; the highest values were estimated using DHGLM (0.17 and 0.12 for untransformed and transformed data, respectively).

Estimates of genetic parameters for residual variance obtained using the full dataset were closer to the results obtained with the female dataset. Comparing sex-specific analyses, higher estimates of h_v^2 and GCV_E were observed for females than

for males. Estimates of h_v^2 and GCV_E for males were always lower than 0.0029 and 0.10, respectively, whereas for females the corresponding estimates achieved values were equal to 0.011 and 0.20, under the DHGLM approach.

The relationship between EBV of the mean and residual variance (r_{mv}) is presented in Table 2. Positive estimates were observed before transformation, being of moderate to high magnitude for DHGLM (0.76) and smaller for the two-step approach (about 0.20). Conversely, negative estimates of r_{mv} were obtained after Box-Cox transformation, under both approaches (varying from -0.04 to -0.15).

Higher estimates of r_{mv} were also observed for females than for males in all approaches (Table 2). DHGLM yielded higher estimates of r_{mv} compared to the two-step approach, for the full dataset and also for the sex-specific analyses.

The comparison between approaches through the correlation between EBV_v is presented in Fig. 3. Moderate correlation estimates were obtained between HOM and DHGLM (0.45) and between HET and DHGLM (0.46). In addition, a very high correlation estimate between EBV_v from HOM and HET was found using the two-step approach (0.97, not shown).

The accuracies of the sire's EBV_v obtained under two-step approach and DHGLM are presented in Fig. 4. At least 100 progeny per sire were required to obtain moderate accuracies. Using sire families with about 100 progeny, DHGLM resulted in higher EBV_v accuracy (0.81) than the two-step approach (about 0.50).

With respect to predictive ability for the mean, an average correlation coefficient of about 0.79 was estimated for both two-step and DHGLM approaches (Table 3). For the residual variance part, HET model yielded better result than the HOM model (0.22 versus 0.19 and 0.11 versus 0.07 for untransformed and transformed data, respectively). However, higher average correlations were observed for DHGLM, 0.72 for untransformed data and 0.22 after transformation. When we fitted a model with null correlation between mean and variance in DHGLM, the average correlations between EBV_v from training and validation sets for untransformed and transformed data were similar (0.32 and 0.26, respectively) and higher than those obtained for the two-step approach (Table 3). These results show that DHGLM yields higher predictive ability for EBV_v than the two-step approach.

Table 2. Estimates of variance components and genetic parameters for untransformed and Box-Cox transformed (BC) yearling weight (YW) under two-step and DHGLM approaches according to the dataset

Dataset	Model	σ^2_S	σ^{2*}_e	$\sigma^2_{S_{V,exp}}$	h^2	h^2_v	GCV_E	r_{mv}
Full	HOM	59.34 (4.06)	589.46 (1.91)	0.65E-2 (0.18E-2)	0.37	0.002 (0.56E-3)	0.10	0.20 (0.04)
	HET	56.05 (4.21)	621 (46.86)[1]	0.83E-2 (0.20E-2)	0.33	0.003 (0.71E-3)	0.12	0.19 (0.04)
	DHGLM	49.79 (3.40)	589.47	0.73E-2 (0.11E-2)	0.31	0.007 (0.11E-2)	0.17 (0.013)	0.76 (0.05)
BC-Full	HOM	0.17E-4 (0.12E-5)	0.17E-3 (0.54E-6)	0.22E-2 (0.11E-2)	0.37	0.0007 (0.37E-3)	0.06	-0.04 (0.04)
	HET	0.16E-4 (0.12E-5)	0.18E-3 (0.13E-4)[1]	0.33E-2 (0.13E-2)	0.32	0.0012 (0.49E-3)	0.07	-0.06 (0.04)
	DHGLM	0.15E-4 (0.10E-5)	0.17E-3	0.34E-2 (0.84E-3)	0.33	0.0030 (0.76E-3)	0.12 (0.01)	-0.15 (0.11)
Male	HOM	50.29 (5.14)	687.36 (3.42)	0.71E-3 (0.15E-2)	0.27	0.0003 (0.14E-2)	0.03	0.14 (0.06)
	HET	48.07 (5.36)	724.4 (55.61)[1]	0.50E-2 (0.26E-2)	0.25	0.0022 (0.11E-2)	0.09	0.14 (0.06)
	DHGLM	45.85 (4.61)	687.2617	0.26E-2 (0.10E-2)	0.25	0.0029 (0.11E-2)	0.10 (0.02)	0.80 (0.13)
Female	HOM	41.44 (4.22)	481.16 (2.46)	0.68E-2 (0.28E-2)	0.32	0.0025 (0.10E-1)	0.10	0.33 (0.06)
	HET	38.22 (4.28)	512.4 (51.37)[1]	0.95E-2 (0.33E-2)	0.28	0.0039 (0.14E-2)	0.12	0.34 (0.06)
	DHGLM	33.08 (3.40)	481.73	0.01 (0.19E-2)	0.26	0.011 (0.22E-2)	0.20 (0.02)	0.89 (0.05)

σ^2_S = sire genetic variance for the mean ($\sigma^2_A = 4\sigma^2_S$); σ^2_e = residual variance for the mean; $\sigma^2_{S_{V,exp}}$ = sire genetic variance for the residual variance on the exponential scale; h^2 and h^2_v = heritability for the mean and residual variance, respectively; GCV_E = genetic coefficient of variation for the residual variance; r_{mv} = genetic correlation between the mean and the residual variance; standard errors are in parenthesis. HOM and HET = two-step approach assuming either homogeneous or heterogeneous residual variance in the first step, respectively. DHGLM = double hierarchical generalized linear model.

*Residual variance from sire model

[1]Average of residual variance among sire families and standard deviation in parenthesis.

Table 3. Average Pearson's correlations between EBV for mean and residual variance obtained from training set and EBV obtained from validation set of yearling weight (YW)

	Mean model			Variance model		
	Two-step approach			Two-step approach		
	HOM	HET	DHGLM	HOM	HET	DHGLM
Untransformed	0.80 (0.38E-2)	0.80 (0.33E-2)	0.78 (0.42E-2)	0.19 (0.03)	0.22 (0.02)	0.72 (0.01)
Transformed	0.79 (0.42E-2)	0.79 (0.34E-2)	0.78 (0.42E-2)	0.07 (0.04)	0.11 (0.03)	0.22 (0.04)
Untransformed/Null correlation*			0.78 (0.38E-2)			0.32 (0.02)
Transformed/Null correlation			0.78 (0.41E-2)			0.26 (0.02)

HOM and HET = homoscedastic and heteroscedastic models, respectively. Standard errors are presented between brackets.

*DHGLM assuming null correlation between mean and residual variance.

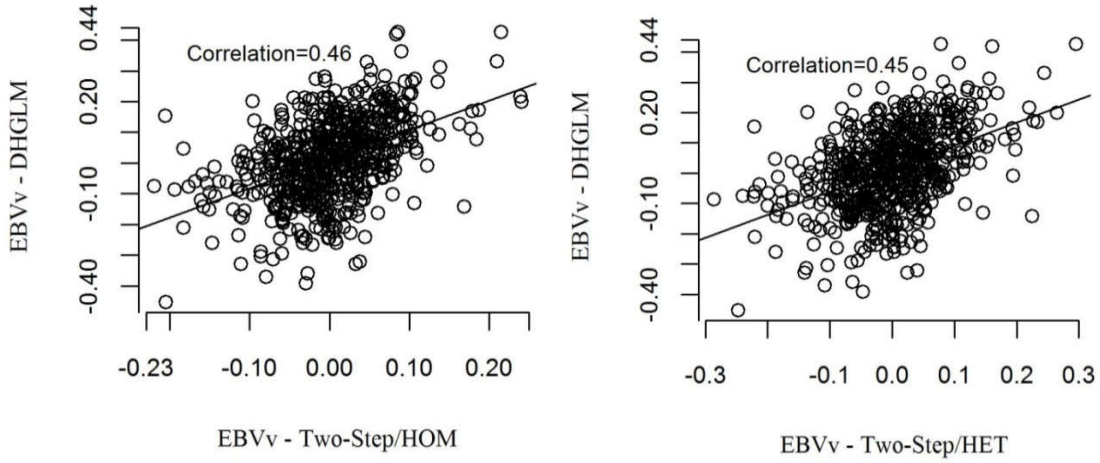

Figure 3. Scatter plot and correlation between sire's EBV for residual variance (EBV_v) estimated by two-step and DHGLM approaches for yearling weight (YW) using untransformed data. Left: two-step homoscedastic model (HOM). Right: two-step heteroscedastic model (HET).

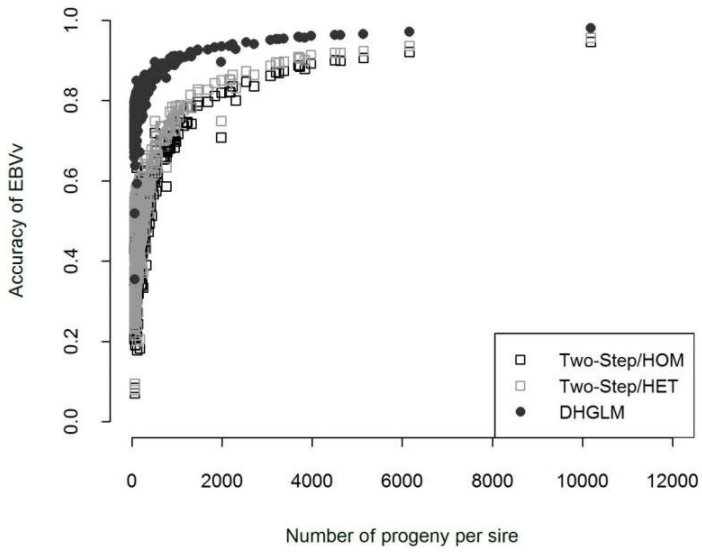

Figure 4. Comparison between accuracies of the sire's EBV for residual variance (EBV_v) of untransformed yearling weight (YW) estimated by two-step, using homoscedastic (Two-Step/HOM) and heteroscedastic (Two-Step/HET) models, and by DHGLM, according to the progeny size per sire.

2.4 Discussion

2.4.1 Genetic parameters

Estimates of h^2 for the mean obtained in this study are slightly lower than those observed in previous studies for YW in Nellore cattle (varying from 0.36 to 0.42; Boligon et al., 2011; Neves et al., 2012; Santana et al., 2013). For residual variance, the low estimates of h_v^2 observed are within the range reported in the literature, especially for growth traits (< 0.05; reviewed by Hill and Mulder 2010; Neves et al., 2012; Sell-Kubiak et al., 2015; Sae-Lim et al., 2015). Such result highlights the need of a large amount of information to predict EBV_v accurately, as previously pointed out by Mulder et al. (2007).

There are several potential reasons for the low estimates of h_v^2 found in this study. The first reason is the definition of the h_v^2 on the individual record level. Estimating a variance based on a single phenotype must be by definition very inaccurate, as the sampling variance of a variance estimate is typically larger than the sampling variance of a mean. These sampling effects appear as 'environmental effects' on the variability. Moreover, the squared estimated residuals considered as response variable in the residual variance model can provide limited information about the true environmental variation, possibly resulting in even smaller h_v^2. Despite this, different studies reinforce the hypothesis that uniformity can be improved through selection. For instance, substantial decrease in residual variance of body weight and litter size in rabbits was achieved in the first generations of selection experiments (Garreau et al., 2008; Argente et al., 2010).

Despite the low h_v^2, the GCV_E values showed the feasibility to improve the uniformity through selection. Based on the results of our study, it was expected that the change of 1 genetic standard deviation will change the residual variance of YW by 10 to 17% for untransformed data, or 6 to 12% for transformed data. The GCV_E values obtained here are lower than the range observed for growth traits in other studies (0.30 to 0.60; Hill and Mulder, 2010; Janhunen et al., 2012; Sonesson et al., 2013; Sae-Lim et al., 2015; Khaw et al., 2016). However, similar values were reported by Neves et al. (2012) for YW using a two-step approach in a subset of the present data.

The h^2 estimates obtained for each sex are in agreement with studies that compared body weight traits between sex, where higher estimates were found in females (Ibáñez-Escriche et al., 2008; Mulder et al., 2009; Wolc et al., 2009). The same behavior was observed in relation to the genetic parameters for the residual variance part (h_v^2 and GCV_E). This result can be explained because the difference in sire genetic variance between males and females is not proportional to the difference in residual variance, resulting in higher heritability for females.

2.4.2 Scale effect

Whenever increased uniformity is part of a breeding goal, knowledge about the r_{mv} is essential to guide selection decisions. However, the r_{mv} may reflect a scale effect, i.e. higher means are associated with higher variances. Data transformation is an alternative to reduce the impact of such scale effect on genetic heterogeneity of residual variance (Yang et al. 2011). In our study, Box-Cox transformation was able to decrease the dependence of the variance on the mean, reducing the estimates of h_v^2 and GCV_E, but it did not eliminate the existence of genetic heterogeneity of residual variance.

Changes in sign and/or magnitude of estimated r_{mv} arising from transformation, as observed here, is not unusual in studies of genetic heterogeneity of residual variance (Yang et al., 2011; Felleki and Lundeheim, 2015). Yang et al. (2011), using a Bayesian hierarchical model to analyze litter size, reported changes in r_{mv} from -0.73 to 0.28 in rabbits and -0.64 to 0.70 in pigs, after transformation. Felleki and Lundeheim (2015) analyzed litter size in pigs using DHGLM and verified a decrease on the estimates of r_{mv} after transformation, from 0.79 to 0.54 and 0.76 to 0.50 on full-sib and half-sib models, respectively. For body weight in fish, Sonesson et al. (2013) and Sae-Lim et al. (2015), using DHGLM approach, found r_{mv} changing from moderate and positive to negative after applying log-transformation. Such findings reinforce the importance of considering possible scale effects when studying genetic heterogeneity of residual variance.

2.4.3 Two-step approach versus DHGLM

Few studies compared different statistical approaches to model genetic heterogeneity at the level of residual variance, most of which restricted the comparison to variance component estimates (Wolc et al., 2009; Rönnegård et al., 2010; Felleki et al., 2012). Recently, Sell-Kubiak et al. (2015) compared estimates obtained from DHGLM and an analysis using within-litter variance of birth weight as a response variable, in pigs, and observed a small advantage of DHGLM to predict EBV_v, using cross-validation.

The correlation between predicted EBV_v from two-step and DHGLM indicated substantial re-ranking of sires between approaches, reinforcing the importance of defining a proper method to predict EBV_v. DHGLM outperformed the two-step approach for the different comparison criteria used in the present study. Regarding the accuracy of EBV_v, DHGLM provided higher accuracies than the two-step approach, especially for sires with progeny of moderate size. As observed by Mulder et al. (2007), moderate accuracies of EBV_v were obtained just when the sires had at least 100 progeny, irrespective of the method in our case. The advantage of DHGLM over two-step, with respect to accuracies of EBV_v, were less pronounced when transformed data were used or when a null r_{mv} was assumed (not shown), indicating that part of the differences on the accuracies of EBV_v were due to indirect information coming from the mean . DHGLM was more advantageous in predicting EBV_v, as also observed by Sell-Kubiak et al. (2015). The high predictive ability of DHGLM with untransformed data can be partially attributed to the high r_{mv}. Nevertheless, even after adjusting a model with null correlation between mean and residual variance (for untransformed and transformed data), the DHGLM outperformed the two-step approach in terms of predictive ability for residual variance.

2.5 Conclusions

We found a genetic component on residual variance of yearling weight and opportunity to change it through selection, assessed through genetic coefficient of variation of residual variance. However, the magnitude of genetic variance and heritability of residual variance, combined with positive genetic correlation between mean and residual variance, may restrict the genetic response to selection for

uniformity of production, especially when simultaneously aiming to increase yearling weight itself. Box-Cox transformation reduced the mean-variance relationship, nevertheless, genetic heterogeneity of residual variance remained beyond the scale effect. DHGLM proved to be more advantageous in predicting EBV_v compared to the two-step approach.

2.6 Literature Cited

Argente, M. J., M. L. Garcia, R. Muelas, M. A. Santacreu, and A. Blasco. 2010. Preliminary results in a divergent selection experiment on variance for litter size in rabbits. In: Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Communication 0546.

Boligon, A. A. and L. G. Albuquerque. 2011. Genetic parameters and relationships of heifer pregnancy and age at first calving with weight gain, yearling and mature weight in Nelore cattle. Livest. Sci. 141:12-16. doi:10.1016/j.livsci.2011.04.009

Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations. J. R. Stat. Soc. Series B Stat. Methodol. 26:211–252.

Felleki, M., D. Lee, Y. Lee, A. R. Gilmour, and L. Rönnegård. 2012. Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. Genet. Res. 94(6):307–317. doi:10.1017/S0016672312000766

Felleki, M., and N. Lundeheim. 2015. Genetic heteroscedasticity of teat count in pigs. J. Anim. Breed. Genet.132(5):392-398. doi:10.1111/jbg.12134

Fries. L. A. 1998. Connectability in beef cattle genetic evaluation: the heuristic approach used in MILC.FOR. In: Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia. p. 449.

Garreau, H., G. Bolet, C. Larzul, C. Robert-Granie, G. Saleil, M. SanCristobal M, and L. Bodin. 2008. Results of four generations of a canalising selection for rabbit birth weight. Livest. Sci. 119:55–62. doi:10.1016/j.livsci.2008.02.009

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, Bayesian Statistics 4. Oxford Univ. Press, UK, p. 164–193.

Hill, W. G., and H. A. Mulder. 2010. Genetic analysis of environmental variation. Genet. Res. 92:381-395. doi:10.1017/S0016672310000546

Hoaglin, D. C., and R. E. Welsch. 1978. The hat matrix in regression and ANOVA. Am. Stat. 32:17–22.

Ibáñez-Escriche, N., L. Varona, D. Sorensen, and J. L. Noguera. 2008. A study of heterogeneity of environmental variance for slaughter weight in pigs. Animal 2:19–26. doi: 10.1017/S1751731107001000

Janhunen, M., A. Kause, H. Vehviläinen, and O. Järvisalo. 2012. Genetics of microenvironmental sensitivity of body weight in rainbow trout (*Oncorhynchus mykiss*) selected for improved growth. PLoS ONE 7(6):e38766. doi:10.1371/journal.pone.0038766

Khaw, H. L., R. W. Ponzoni, H. Y. Yee, M. A. Aziz, H. A. Mulder, J. Marjanovic, and P. Bijma. 2016. Genetic variance for uniformity of harvest weight in Nile tilapia (*Oreochromis niloticus*). Aquaculture 451:113-120. doi: 10.1016/j.aquaculture.2015.09.003

Madsen, P., and J. Jensen. 2013. A user's guide to DMU. Version 6, release 5.2. Danish Institute of Agricultural Sciences, Research Centre Foulum, Denmark.

Mulder, H. A., P. Bijma, and W. G. Hill. 2007. Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. Genetics 175(4):1895–1910. doi:10.1534/genetics.106.063743.

Mulder, H. A., W. G. Hill, A. Vereijken, and R. F. Veerkamp. 2009. Estimation of genetic variation in residual variance in female and male broiler chickens. Animal 3(12):1673–1680. doi:10.1017/S1751731109990668

Mulder, H. A., L. Rönnegård, W. F. Fikse, R. F. Veerkamp, and E. Strandberg. 2013. Estimation of genetic variance for macro- and micro-environmental sensitivity using double hierarchical generalized linear models. Genet. Sel. Evol. 45:23. doi:10.1186/1297-9686-45-23.

Mulder, H. A., J. Visscher, and J. Fablet. 2016. Estimating the purebred–crossbred genetic correlation for uniformity of eggshell color in laying hens. Genet. Sel. Evol. 48:39. doi:10.1186/s12711-016-0212-2

Neves, H. H. R., R. Carvalheiro, and S. A. Queiroz. 2012. Genetic and environmental heterogeneity of residual variance of weight traits in Nellore beef cattle. Genet. Sel. Evol. 44:19. doi:10.1186/1297-9686-44-19

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6:7-11.

R Development Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rönnegård, L., M. Felleki, F. Fikse, H. A. Mulder, and E. Strandberg. 2010. Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. Genet. Sel. Evol. 42(8):1-10. doi:10.1186/1297-9686-42-8

Roso, V. M., and F. S. Schenkel. 2006. Proc. AMC – a computer program to assess the degree of connectedness among contemporary groups. In: Proc. 8th World Congr. Genet. Appl. Livest. Prod. Belo Horizonte, Brazil. p. 26.

Sae-Lim, P., A. Kause, M. Janhunen, H. Vehviläinen, H. Koskinen, B. Gjerde, M. Lillehammer, and H. A. Mulder. 2015. Genetic (co)variance of rainbow trout (*Oncorhynchus mykiss*) body weight and its uniformity across production environments. Genet. Sel. Evol. 47:46. doi:10.1186/s12711-015-0122-8

Santana, M. L., J. P. Eler, A. B. Bignardi, and J. B. S. Ferraz. 2013. Genetic associations among average annual productivity, growth traits, and stayability: A parallel between Nelore and composite beef cattle. J. Anim. Sci. 91(6):2566-2574. doi:10.2527/jas.2012-5856

Sell-Kubiak, E., P. Bijma, E. F. Knol, and H. A. Mulder. 2015. Comparison of methods to study uniformity of traits: Application to birth weight in pigs. J. Anim. Sci. 93(3):900-911. doi:10.2527/jas2014-8313

Sonesson, A. K., J. Ødegård, and L. Rönnegård. 2013. Genetic heterogeneity of within family variance of body weight in Atlantic salmon (*Salmo salar*). Genet. Sel. Evol. 45(1):41. doi:10.1186/1297-9686-45-41

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S, 4th ed. Springer, New York.

Wolc, A., I. M. S. White, S. Avendano, and W. G. Hill. 2009. Genetic variability in residual variation of body weight and conformation scores in broiler chickens. Poultry Sci. 88:1156– 1161. doi: 10.3382/ps.2008-00547

Yang, Y., O. F. Christensen, and D. Sorensen. 2011. Analysis of a genetically structure variance heterogeneity model using the Box-Cox transformation. Genet. Res. 93:33-46. doi:10.1017/S0016672310000418

CHAPTER 3 – Genomic regions underlying uniformity of yearling weight in Nellore cattle evaluated under different response variables[a]

Abstract

Background: In livestock, residual variance has been studied because of the interest to improve uniformity of production. Several studies have provided evidence that residual variance is partially under genetic control; however, few investigations have elucidated genes that control it. The aim of this study was to identify genomic regions associated with within-family residual variance of yearling weight (YW; N = 423) in Nellore bulls with high density SNP data, using different response variables. For this, solutions from double hierarchical generalized linear models (DHGLM) were used to obtain different response variables, as follows: a DGHLM assuming non-null genetic correlation between mean and residual variance ($r_{mv} \neq 0$) to obtain deregressed EBV for mean ($dEBV_m$) and residual variance ($dEBV_v$); and a DHGLM assuming $r_{mv} = 0$ to obtain two alternative response variables for residual variance, $dEBV_{v\_r0}$ and log-transformed variance of estimated residuals ($ln\_\sigma^2_{\hat{e}}$).

Results: The $dEBV_m$ and $dEBV_v$ were highly correlated (0.90), resulting in common regions associated with mean and residual variance of YW. However, higher standardized effects on variance than the mean showed that these regions had effects on the variance beyond scale effects. More independent association results between mean and residual variance were obtained when null r_{mv} was assumed. While 13 and 4 single nucleotide polymorphisms (SNPs) with a strong association (Bayes Factor > 20) with $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$, respectively, were identified, only suggestive signals were found for $dEBV_{v\_r0}$. All shared 1-Mb windows among top 20 between $dEBV_m$ and $dEBV_v$ were previously associated with growth traits. The potential candidate genes for uniformity of YW are involved in metabolism, stress, inflammatory and immune responses, mineralization, neuronal activity and bone formation.

Conclusions: It is necessary to use a strategy like assuming null r_{mv} to obtain genomic regions associated with uniformity that are not associated with the mean. Genes involved not only in metabolism, but also stress, inflammatory and immune responses, mineralization, neuronal activity and bone formation were the most

[a]Article resubmitted for publication in the “BMC Genomics” on 23/05/2018.

potential biological candidates for uniformity of YW. Since no clear evidence in favor of using a specific response variable was found, we recommend considering multiple response variables to study uniformity to increase evidence on candidate regions and biological mechanisms behind it.

Keywords: beef cattle, DHGLM, genetic heterogeneity of residual variance, growth traits, GWAS, micro-environmental sensitivity

3.1 Introduction

Uniformity is becoming increasingly important among livestock species. In meat production systems, this is more evident by the increasing adoption of economic incentives by slaughterhouses to stimulate farmers to deliver animals that meet specific carcass standards. If uniformity is partly under genetic control, genetic selection could be used to improve uniformity of animals. Genetic control of uniformity can be the result of how genotypes respond differently towards unknown micro-environmental factors [1,2]. Such a phenomenon is called genetic heterogeneity of residual variance or genetic variance in micro-environmental sensitivity. Up to now, several studies support the existence of a genetic component on residual variance and draw attention for its potential to improve uniformity through selection (e.g. [3-6]). Unraveling the genetic basis of heterogeneity of residual variance through genome-wide association studies (GWAS) will help to understand the biology behind it and increase selection response by identifying candidate genes affecting uniformity and considering this information in genomic prediction.

Few GWAS for uniformity traits have been performed in livestock and there is no consensus about the most appropriate response variable to be used in such studies. Phenotypic standard deviation and coefficient of variation were used to address uniformity of egg weight in chickens by Wolc et al. [7] and birth weight in pigs by Wang et al. [8,9]. Residual variance per individual from double hierarchical generalized linear model (DHGLM; obtained according to Rönnegård et al. [10]) were used as response variable by Mulder et al. [11] to identify genomic regions related to residual variance of somatic cell score in dairy cattle. Estimated breeding values (EBV) from a DHGLM, using an extension developed by Felleki et al. [12], were deregressed (dEBV) and used to identify genomic regions associated with variability of litter size in pigs by Sell-Kubiak et al. [13]. Such variety of response variables can be explained because uniformity can be measured in different ways depending on trait and data structure. Statistical analysis requires either the within-individual variance based on repeated observations per animal or the within-family variance based on large offspring groups per family. Furthermore, the availability of genotypes influences the choice for the response variable as well. The choice for dEBV in the study by Sell-Kubiak et al. [13] was made because of availability of genotypes on

boars with many phenotyped offspring and sows with genotypes and phenotypes for litter size. In the case of Mulder et al. [11], genotypes were only available on cows with many phenotypic observations on experimental farms.

In beef cattle, growth traits are often used as selection criteria in breeding programs due to its economic impact in meat production. Several genes and genomic regions influencing the mean of these traits have been identified (e.g. [14-16]). In addition, the existence of a genetic component on residual variance of growth traits was previously reported [17-19]. However, no further investigations have been carried out to study the genetic mechanisms underlying uniformity of growth traits. Thus, the aim of this study was to identify genomic regions and genes potentially associated with within-family residual variance of yearling weight (YW) in Nellore cattle through GWAS, using different response variables.

3.2 Results

Our aim was to identify genomic regions associated with within-family residual variance of YW in Nellore cattle, using different response variables in GWAS. For this, we used solutions from DHGLM assuming: i) non-null genetic correlation between mean and residual variance of YW ($r_{mv} \neq 0$) to obtain deregressed EBV for mean ($dEBV_m$) and residual variance ($dEBV_v$); and ii) $r_{mv} = 0$ to obtain $dEBV_{v\_r0}$ and log-transformed variance of estimated residuals ($ln\_\sigma^2_{\hat{e}}$).

The Pearson's correlations between the different response variables and the number of common 1-Mb windows (among the top 20 that explained the largest proportion of genetic variance in the association studies) among them are shown in Table 1. As expected, $dEBV_m$ was highly correlated with $dEBV_v$, 0.90, given the high and positive r_{mv} (0.76; [19]), resulting in eight out of the top 20 windows shared between both $dEBV_m$ and $dEBV_v$ (Table 2). The $dEBV_m$ was also positively correlated with $dEBV_{v\_r0}$, 0.19, but no overlap was found among the top 20 windows.

The response variables for residual variance of YW had low to moderate correlations with each other (Table 1). The largest correlation was observed between $dEBV_v$ and $dEBV_{v\_r0}$ (0.53), which shared one window among the top 20, Chr12_5. On the other hand, the smallest correlation was between $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$ (0.08), which also shared only one window among the top 20, Chr22_52. The $dEBV_{v\_r0}$ and

$ln\_\sigma^2_{\hat{e}}$ showed a correlation of 0.24 and had two common regions among the top 20: Chr6_105 and Chr9_8.

Table 1. Number of common 1-Mb windows[1] (above diagonal) and Pearson's correlation (below diagonal) between response variables

	$dEBV_m$	$dEBV_v$	$dEBV_{v\_r0}$	$ln\_\sigma^2_{\hat{e}}$
$dEBV_m$		8	0	0
$dEBV_v$	0.90 (0.02)		1	1
$dEBV_{v\_r0}$	0.19 (0.05)	0.53 (0.04)		2
$ln\_\sigma^2_{\hat{e}}$	-0.01 (0.05)	0.08 (0.05)	0.24 (0.05)	

[1]Considering only the top 20 windows that explained the largest proportion of genetic variance for each response variable; $dEBV_m$ and $dEBV_v$: deregressed EBV for mean and residual variance of yearling weight, respectively; $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$: deregressed EBV for residual variance and log-transformed variance of estimated residuals, respectively, both assuming null genetic correlation between mean and residual variance. Standard errors are presented between brackets.

Table 2. Common 1-Mb windows shared by deregressed EBV for mean ($dEBV_m$) and residual variance ($dEBV_v$)

	% Var[2]			
Windows[1]	$dEBV_m$	$dEBV_v$	Candidate gene and/or QTL region[3]	Reference[4]
Chr1_92	2.57	0.51	Chr1_70.2:94.6, Chr1_87.2:98.9	[20,21]
Chr3_42	2.89	1.54	Chr3_41	[15]
Chr3_45	1.00	1.65	Chr3_41	[15]
Chr13_59	1.07	0.60	Chr13_58, *GNAS*	[15,22]
Chr14_24:26	3.76	3.07	*RB1CC1, NPBWR1, PLAG1, PENK*	[14-16,23-27]
Chr16_21	1.26	5.13	Chr16_8.4:34, Chr16_10.6:22.5, Chr16_19.4:40	[21,28]

In bold are the windows harboring strong associations (Bayes Factor > 20) for both response variables.
[1] Represented by chromosome (Chr) and physical position (in Mb) (Chr_Mb and Chr_Mb:Mb).
[2] Proportion of genetic variance explained by each 1-Mb window.
[3] Ensembl Database UMD3.1 and QTLdb database (represented by Chr_Mb and Chr_Mb:Mb).
[4] Previous studies reporting QTL regions and/or candidate genes related with growth traits that are next or within the windows found in this study.

Based on the GWAS results, 40, 13 and 4 single nucleotide polymorphisms (SNPs) showed a strong association (BF > 20) with $dEBV_m$, $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$, respectively (Figure 1 and Additional file 1). Box plots of $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$ by genotype of the markers with higher BF were presented to give an overview about the signals that these SNPs are capturing (Figure 2). In both cases, the first homozygous genotype (0 or AA) was more uniform, with lower mean and lower dispersion of the corresponding response variable, compared to the other genotypes, although AA was less frequent in relation to AB and BB. Only a few suggestive association signals (2; BF > 3) were found for $dEBV_{v\_r0}$. All these SNPs are located

within the top 20 1-Mb windows that explained the largest proportion of genetic variance in each response variable. The top 20 windows explained together 34.6%, 32.6%, 4.4% and 20.0% of the genetic variance of $dEBV_m$, $dEBV_v$, $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$ (data not shown). The range of the proportion of variance explained by individual windows suggests that uniformity of YW behaves as a polygenic trait determined by several genes, as well as its mean.

Within the common windows shared by $dEBV_m$ and $dEBV_v$, genes and QTL regions were mapped (Table 2). Several candidate genes for growth traits in beef cattle were identified on chromosome 14 (*RB1CC1, NPBWR1, PLAG1* and *PENK*). Such findings highlight that most of the effects captured by $dEBV_v$ can be due to the strong r_{mv} and could be potentially scale effects [19].

In order to investigate whether these common regions can really affect the variance beyond a simple scale effect, we applied a similar procedure as described in Wolc et al. [7]. For this, only the common SNP with the highest BF in each window was considered. The absolute effect of these SNPs on $dEBV_m$ and $dEBV_v$ were divided by the mean of YW and mean of σ^2_e, respectively. If SNP would have an effect on mean and variance due to scaling, then the standardized effects should be the same or higher on mean. However, all SNPs showed a greater standardized effect on $dEBV_v$ than $dEBV_m$ indicating that these regions affect the residual variance beyond a simple scale effect (data not shown). These results suggest that even with a high and strong r_{mv} it is possible to find regions determining mean growth and its variability simultaneously, while the effects on the variance are larger than can be explained by a simple scale effect.

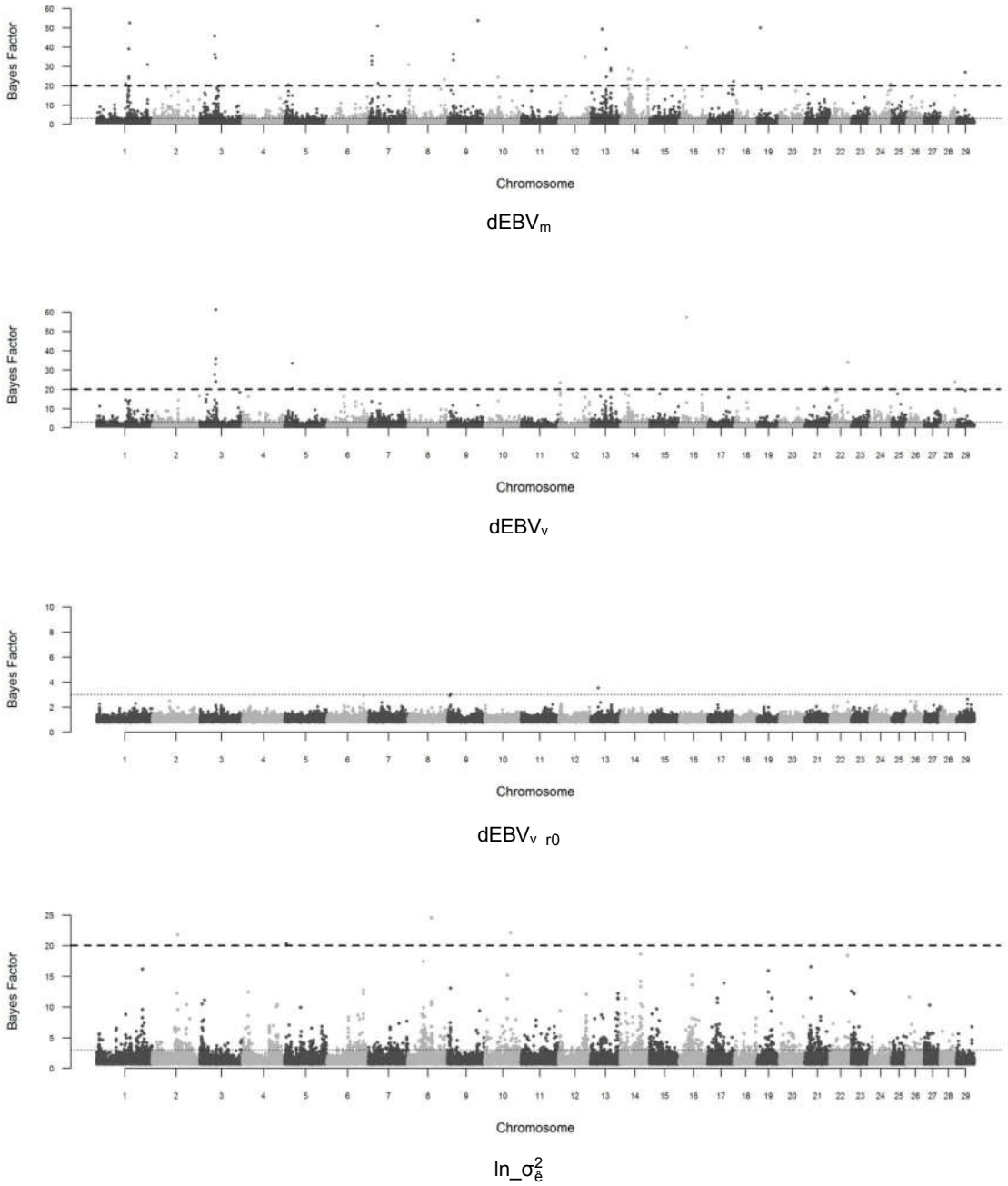

Figure 1. Manhattan plot for the mean (dEBV$_m$) and residual variance (dEBV$_v$, dEBV$_{v_r0}$ and ln_$\sigma^2_{\hat{e}}$) of yearling weight. dEBV$_m$ and dEBV$_v$: deregressed EBV for mean and residual variance, respectively; dEBV$_{v_r0}$ and ln_$\sigma^2_{\hat{e}}$: deregressed EBV for residual variance and log-transformed variance of estimated residuals, respectively, both obtained from solutions of a DHGLM assuming null genetic correlation between mean and residual variance. The horizontal dotted and dashed lines represent Bayes Factor of 3 (suggestive) and 20 (strong), respectively.

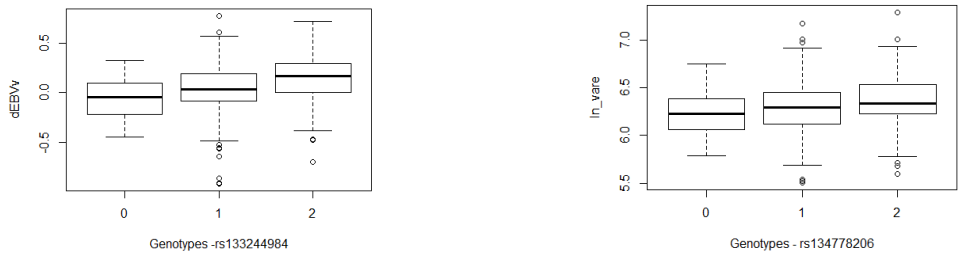

Figure 2. Box plots for the $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$ according to rs133244984 and rs134778206 genotypes, respectively.

In the shared windows by $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$, Chr6_105 and Chr9_8, genes involved in metabolism, mineralization and neuronal activity were considered potential candidates for uniformity (Table 3). No gene was identified within Chr12_5, shared by $dEBV_v$ and $dEBV_{v\_r0}$. However, several potential genes were found on Chr22_52 shared by $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$; among them, genes related to metabolism, stress and inflammatory and immune responses (Table 3).

For uniformity, potential genes related to energy and protein metabolism, stress, inflammatory and immune responses, neuronal plasticity, bone formation and were observed in the windows of the SNPs that showed strong association with $ln\_\sigma^2_{\hat{e}}$ and $dEBV_v$ (Table 4 and 5).

Table 3. Overlapped 1-Mb windows between response variables for uniformity ($dEBV_v$, $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$)

Windows[1]	Shared by			% Variance Explained[2] (BF[3])			Candidate Genes[4]
	$dEBV_v$	$dEBV_{v\_r0}$	$ln\_\sigma^2_{\hat{e}}$	$dEBV_v$	$dEBV_{v\_r0}$	$ln\_\sigma^2_{\hat{e}}$	
Chr12_5	X	X		1.69 (23.48)	0.19 (1.99)		-
Chr22_52	X		X	1.73 (33.97)		0.72 (18.37)	*UBA7, IP6K1, GMPPB, RNF123, MST1, APEH, AMT, RHOA, USP4, USP19, IMPDH2, NDUFAF3, SLC25A20, PRKAR2A, IP6K2, UQCRC1, UCN2, PFKFB4, SHISA5*
Chr6_105		X	X		0.29 (2.94)	0.91 (12.74)	*NUDT9, SPARCL1, DSPP, DMP1, MAN2B2, PPP2R2C, JAKMIP1*
Chr9_8		X	X		0.28 (3.05)	0.67 (13.08)	*ADGRB3*

$dEBV_v$: deregressed EBV for residual variance of yearling weight; $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$: deregressed EBV for residual variance and log-transformed variance of estimated residuals, respectively, both assuming null genetic correlation between mean and residual variance.

[1] Represented by chromosome (Chr) and physical position (in Mb) (i.e. Chr_Mb).
[2] Proportion of genetic variance in each 1-Mb window.
[3] Single nucleotide polymorphism (SNP) with the highest Bayes Factor (BF) within each 1-Mb window.
[4] Ensembl Database UMD3.1. See the complete gene list in each window in Additional file 1.

Table 4. Significant SNPs[1] associated with log-variance of estimated residuals ($\ln\_\sigma^2_{\hat{e}}$)

SNP	Chr	Position (Mb)	BF	MAF	SNP effect	Windows[2]	% Var[3]	Candidate Genes[4]
rs134778206	8	67.46	24.51	0.40	0.00075	Chr8_68	2.34	*LPL, SLC18A1, ATP6V1B2, LZTS1*
rs133984762	10	74.34	22.11	0.48	0.00068	Chr10_75	1.27	*HIF1A*
rs132792897	2	73.42	21.75	0.50	-0.00066	Chr2_74	1.97	*GLI2, CLASP1*
rs135481650	5	5.38	20.32	0.49	-0.00061	Chr5_6	1.04	*BBS10, OSBPL8*

[1] Single nucleotide polymorphisms (SNPs) that showed a strong association with $\ln\_\sigma^2_{\hat{e}}$ according Bayes Factor (BF > 20).
[2] Represented by chromosome (Chr) and physical position (in Mb) (i.e. Chr_Mb).
[3] Proportion of genetic variance in each 1-Mb window.
[4] Ensembl Database UMD3.1. See the complete gene list in each window in Additional file 1.

Table 5. Significant SNPs[1] associated with deregressed EBV for residual variance ($dEBV_v$)

SNP	Chr	Position (Mb)	BF	MAF	SNP effect	Windows[2]	% Var[3]	Candidate Genes[4]
rs133244984	3	45.30	61.24	0.32	0.00221	Chr3_46	8.67	*DPYD*
rs137051934	16	20.84	57.29	0.40	0.00190	Chr16_21	5.13	-
rs135408640	3	45.29	35.74	0.41	0.00117	Chr3_46	8.67	*DPYD*
rs42014753	22	51.60	33.97	0.37	0.00112	Chr22_52	1.73	*UBA7, IP6K1, GMPPB, RNF123, MST1, APEH, AMT, RHOA, USP4, USP19, IMPDH2, NDUFAF3, SLC25A20, PRKAR2A, IP6K2, UQCRC1, UCN2, PFKFB4, SHISA5*
rs110480161	5	22.01	33.44	0.45	0.00110	Chr5_23	2.38	*BTG1*
rs136349671	3	44.37	32.98	0.39	0.00106	Chr3_45	1.65	*PLPPR4, PLPPR5*
rs137143404	3	41.54	27.60	0.50	-0.00087	Chr3_42	1.54	-
rs135488380	3	45.34	23.95	0.34	0.00079	Chr3_46	8.67	*DPYD*
rs137135953	28	41.34	23.88	0.37	0.00074	Chr28_42	0.84	*BMPR1A, ADIRF, GLUD1*
rs133451489	12	4.46	23.48	0.39	-0.00072	Chr12_5	1.69	-
rs109757968	21	62.41	20.46	0.33	0.00065	Chr21_63	0.67	*TCL1B, TCL1A, BDKRB2, BDKRB1, AK7*
rs136908466	5	22.06	20.28	0.43	-0.00063	Chr5_23	2.38	*BTG1*
rs137014141	14	14.47	20.16	0.35	0.00062	Chr14_15	0.85	-

[1] Single nucleotide polymorphisms (SNPs) that showed a strong association with $dEBV_v$ according Bayes Factor (BF > 20).
[2] Represented by chromosome (Chr) and physical position (in Mb) (i.e. Chr_Mb).
[3] Proportion of genetic variance in each 1-Mb window.
[4] Ensembl Database UMD3.1. See the complete gene list in each window in Additional file 1.

3.3 Discussion

In this study, we identified genomic regions associated with within-family residual variance of YW in Nellore cattle by using different response variables ($dEBV_v$, $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$). Thirteen and four strong associations (BF > 20) were observed using $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$, respectively. Common windows among the top 20 that explained most of the genetic variance were observed between $dEBV_v$ and the other response variables including $dEBV_m$. The $dEBV_v$ shared 8 out of the top 20 windows with $dEBV_m$, one with $dEBV_{v\_r0}$ and another one with $ln\_\sigma^2_{\hat{e}}$. Potential candidate genes related to metabolism (carbohydrate, energy, lipid, nucleotide and amino acid), stress, inflammatory and immune responses, mineralization, neuronal activity and bone formation were found for $ln\_\sigma^2_{\hat{e}}$ and $dEBV_v$.

3.3.1 Common regions between $dEBV_m$ and $dEBV_v$

All common regions shared by $dEBV_m$ and $dEBV_v$ contained QTL previously associated with birth, yearling and carcass weight in other breeds [15,20,28,21]. The region found on chromosome 14 (24 to 26 Mb) was associated with growth, feed efficiency and carcass quality traits in beef cattle and several candidate genes were reported [14-16,23-27]. According to these studies, *pleiomorphic adenoma gene 1* (*PLAG1*) seems to be the major gene due to its role in regulating insulin-like growth factors (IGF; [29]). However, some of these genes, *like RB1-inducible coiled-coil 1* (*RB1CC1*), *neuropeptides B/W receptor 1 (NPBWR1)* and *proenkephalin* (*PENK*), are also involved in processes that can contribute to determine uniformity of growth traits, like YW. *RB1CC1* is involved in cell growth and differentiation, senescence, apoptosis and autophagy, and was one of the genes differentially expressed in heat stressed chickens [30,31]. *NPBWR1* modulates feeding behavior and energy homeostasis and *PENK* is responsible for producing enkephalins in response to stress [32,33]. Based on these findings and the fact that these shared regions between $dEBV_m$ and $dEBV_v$ showed effects beyond scale effects, we can conclude that the same gene can affect both, mean and its variability.

3.3.2 Potential genes associated with $ln\_\sigma^2_{\hat{e}}$, $dEBV_v$ and $dEBV_{v\_r0}$

Genes involved in metabolism (*ATP6V1B2*, *LPL*, *OSBPL8* and *GLI2*), immune response (*GLIPR1*), neuronal plasticity (*LZTS1*) and stress response (*SLC18A1*, *HIF1A* and *BBS10* and *CLASP1*) were considered the most potential biological candidates for uniformity using $ln\_\sigma^2_{\hat{e}}$. For $dEBV_v$, genes related to metabolism (*DPYD, BTG1, AK7, ADIRF, GLUD1, IP6K1, GMPPB, APEH, AMT, USP4, USP19, IMPDH2, NDUFAF3, SLC25A20, PRKAR2A, UQCRC1, PFKFB4* and *SHISA5*), stress response (*BDKRB2*, *BDKRB1*, *IP6K2* and *UCN2*), inflammatory and immune responses (*TCL1A, TCL1B, UBA7, RNF123, MST1* and *RHOA*), neuronal plasticity (*PLPPR4* and *PLPPR5*) and bone formation (*BMPR1A*) were identified as potential candidates. Such findings support partially the hypothesis that most likely the mean of a trait and its uniformity share processes, like those related to metabolism. However, uniformity may be a trait which is even more complex and may be controlled by several processes and mechanisms beyond metabolism, in order to reach homeostasis and maintain the body in balance.

Among the metabolism-related genes for $ln\_\sigma^2_{\hat{e}}$, *ATPase, H+ transporting, V1 subunit B2 (ATP6V1B2)* encodes a component of the vacuolar ATPase (V-ATPase), which is activated according to the energy level, like during glucose starvation [34]. Previously, the *ATP6V1B2* gene was found to be associated with feed efficiency traits in beef cattle [35]. *Lipoprotein lipase (LPL*) and *oxysterol binding protein-like 8 (OSBPL8)* genes are involved in lipid metabolism. *LPL* is responsible for hydrolysis of circulating triglycerides and very low-density lipoprotein [36]. Associations between *LPL* and growth and carcass quality traits were also reported previously [37-40]. *OSBPL8* modulates lipid homeostasis through sterol regulatory element binding proteins (SREBPs; [41]). In addition, *glioma-associated oncogene family zinc finger 2 (GLI2)* gene is a transcription factor of the Hedgehog (Hh) signaling pathway, which is involved in muscle development mainly during embryonic development, but also in adult tissue homeostasis and repair [42,43].

For $dEBV_v$, most of the genes are involved in metabolic pathways or related processes to it, such as metabolism of carbohydrate, energy and lipid, nucleotide and amino acid (*GMPPB, UQCRC1, PFKFB4, AMT, GLUD1, AK7, IMPDH2, NDUFAF3, DPYD, SLC25A20, PRKAR2A* and *IP6K1*). Genes regulating muscle cell

proliferation, differentiation and myogenesis (*BTG1* [41], *USP4* and *USP19* [45,46], and *SHISA5*), adipogenic differentiation (*ADIRF* [47]) and proteolysis (*APEH*) also showed metabolic functions that potentially can contribute to uniformity of YW.

Genes involved in the energy and protein metabolism may play an important role to keep body homeostasis against environmental perturbations. Usually, metabolic responses to these stressors involve catabolic processes such as energy mobilization and protein degradation [48,49]. In such situations, genes related to metabolism are required to recover homeostasis and consequently part of the energy is diverted from growth which can potentially affect performance and therefore explain the potential trade-off between energy for growth and energy for homeostasis.

Genes related to stress response were also found using $\ln\_\sigma^2_{\hat{e}}$ (*SLC18A1*, *HIF1A*, *BBS10* and *CLASP1*) and $dEBV_v$ (*BDKRB2*, *BDKRB1*, *IP6K2* and *UCN2*). Stress is a response to new environmental conditions, like nutrition, housing or any stimuli, that threatens homeostasis. Therefore, stress genes may have a direct effect on uniformity, such that animals that are worse in dealing with environmental perturbations tend to be less uniform (e.g. [50]).

Vesicular monoamine transporter 1 (SLC18A1) and *urocortin 2 (UCN2*) genes are involved in the hypothalamic-pituitary-adrenal (HPA) axis. *SLC18A1* acts in the final stage of the HPA-axis, transporting catecholamines like dopamine, noradrenaline and adrenaline. Catecholamines are released as a physiological stress response [51]. *UCN2* is a member of the corticotropin-releasing hormone (CRH) family, which is a key mediator of the stress response by activating the HPA axis [52,53].

The *hypoxia-inducible factor 1-alpha (HIF1A*) gene encodes a transcription factor that is induced by hypoxia, promoting angiogenesis, cell proliferation/survival and glucose/iron metabolism [54]. Angiogenesis is the process of new vessel formation, which is related to heat stress. Recently, *HIF1A* was highly expressed in the ruminal epithelium of highly-efficient animals [55], which can be explained by its functions, mainly angiogenesis and glucose metabolism. It indicates that animals can differ in relation to absorption and energy generation, changing individual growth performance.

The *Bardet-Biedl Syndrome 10 (BBS10*) gene, a member of heat shock protein (HSP) family, encodes a chaperonin-like protein [56]. The HSP family plays a crucial role in response to several stressful conditions in addition to heat stress [57], and it is not the first time that a member of this family is related to uniformity. Previously, *Hsp90* was considered a candidate gene for developmental stability of morphological traits in *Drosophila melanogaster* and *Arabidopsis thaliana* (e.g. [58-61]) and variability of litter size in pigs [13]. It supports the evidence that the function of chaperones is extended to several organisms, making these HSP protein genes potential candidates for uniformity.

Another gene related to HSP family is the *inositol hexakisphosphate kinase 2 (IP6K2)*, a metabolism related-gene that also plays a role in apoptosis [62]. However, its catalytic activity is regulated by *HSP90* gene [63], which suggests that cellular response can also be expected due to environmental stimuli.

Cytoplasmic linker associated protein 1 (CLASP1) and *bradykinin receptor B1 and B2 (BDKRB1* and *BDKRB2)* genes are involved in regulation of wound healing and blood pressure regulation, mechanisms that can be used to deal with environmental stressors. Previously, Coble et al. [64] found *BDKRB1* among up-regulated genes during heat stress in chickens.

Some genes involved in inflammatory and immune responses, *GLIPR1, EEA1, TCL1A, TCL1B, UBA7, RNF123, MST1* and *RHOA*, were also found. *GLIPR1, UBA7, RNF123* and *EEA1* participate in several immune system-related pathways. *T-cell leukemia/lymphoma 1A and 1B (TCL1A and TCL1B*) genes are involved in T- and B-cell development [65]. The *macrophage stimulating 1 (MST1*) gene regulates macrophage activity during inflammation [66,67]. Ras homolog family member A (RHOA) is a small GTPases of the Rho family, which regulates several cellular processes including actin cytoskeleton reorganization and inflammation [68,69].

Four other potential candidate genes for uniformity were identified. *Leucine zipper putative tumor suppressor 1 (LZTS1*), *phospholipid phosphatase related 4 and 5 (PLPPR4* and *PLPPR5*) are involved in neuronal plasticity [70,71], and *bone morphogenetic protein receptor type 1A* (*BMPR1A*) playing an important role in bone formation, which was recently found associated with body size in sheep [72].

Neuronal plasticity is the ability of the nervous system to adapt to environmental changes, which denotes another potential class of genes determining uniformity.

In the common regions between $\ln\_\sigma^2_{\hat{e}}$ and $dEBV_{v\_r0}$, genes related to metabolism (*NUDT9*, *SPARCL1* and *MAN2B2*) and mineralization of tissues like bones and teeth (*DMP1* and *DSPP*; [73]), are likely to be linked to uniformity on Chr6_105. It is not unusual to find common genes related to stature, height and body weight in beef cattle (e.g. [14,74-76]). Such findings highlight the strong relationship between height and weight, partially explained by common regulatory mechanisms.

Additionally, both common regions (Chr6_105 and 9_8) contain genes with neuronal functions*: PPP2R2C, JAKMIP1* and *ADGRB3*. *Protein phosphatase 2, regulatory subunit B, gamma (PPP2R2C)* and *adhesion G protein-coupled receptor B3 (ADGRB3)* were previously related to learning and memory (synaptic plasticity) and synaptogenesis in mice [77,78]. *Janus kinase and microtubule-interacting protein 1 (JAKMIP1)* was associated with social behavior in mice through its function on neuronal translation and was related to T cell-mediated cytotoxicity [79,80]. These results together with $dEBV_v$ and $\ln\_\sigma^2_{\hat{e}}$ provide evidence that the nervous system is also modulated according to environmental perturbations. Previously, studies reporting such neuronal plasticity in response to changes in nutrition and temperature were observed in *Drosophila* [81-83]. Therefore, together with other tissues, the nervous system seems to be a potential modulator of the mechanisms underlying uniformity.

3.3.3 Response variables

The high correlation between $dEBV_m$ and $dEBV_v$ resulted in 8 common regions affecting simultaneously mean and residual variance. Such findings are in line with Wolc et al. [7] who found correlations ranging from 0.54 to 0.74 and a region explaining a large proportion of the genetic variance of the mean and standard deviation (SD) of egg weight at different ages. The authors noted that it was not a simple scale effect, because the effect of the QTL on mean egg weight (measured between 26 and 28 weeks of age) was about 4% of the mean egg weight and on SD egg weight at the same age was about 5% of the mean SD. More recently, Sell-

Kubiak et al. [13] also reported a high genetic correlation, 0.49, between litter size and its variation, but no overlapped genomic regions were found.

The $dEBV_v$ not only captured common effects with the mean of YW but also its residual variance by sharing genomic regions (among top 20 by variance explained) with $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$, Chr12_5 and Chr22_52, respectively. These results suggest that the mean and uniformity of YW are partially under different genetic control. Though some of the genes found on chromosome 22 and most of those in common between $dEBV_m$ and $dEBV_v$ are involved in metabolism (e.g. *PLAG1*, *DPYD*, *AK7*, *GMPPB*, *AMT* and *GLUD1*), a process most probably related to both, mean and uniformity.

The weak correlations among $dEBV_m$ and the response variables that assumed null r_{mv} ($dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$) emphasize the role of the latter to find genomic regions that are specifically affecting residual variance and not the mean of YW. Generally, data transformation is used as the main tool to reduce mean-variance relationship in studies on uniformity (e.g. [84-86]). This is one of the first studies on uniformity reporting null r_{mv} as a strategy to deal with scale effects and showing the differences in terms of associated regions when comparing to GWAS using $dEBV_v$ when non-null r_{mv} was assumed.

Knowing the importance of considering potential scale effects and confounding between mean and residual variance, the next step is to discuss the suitability of the different response variables to be used in a GWAS for uniformity. Based on our results, we recommend considering the following aspects regarding the response variable and data structure. Firstly, we need to take into account the nature of the trait because small genetic variance and low heritabilities are often found for uniformity. Previously we estimated a heritability for residual variance of YW even smaller (0.007; [19]) than has been reported for similar traits in other livestock populations [87]. Probably this also may have contributed to the small (or none) number of SNPs that reached the significance threshold (BF > 20) when we analyzed the response variables for residual variance. In this context, irrespective of the response variable, a larger sample size is required to increase the power of GWAS [88].

Secondly, we used two different procedures to obtain response variables to perform GWAS for uniformity. The first one was the deregressed EBV; deregression, which removes both the contribution of parents and the shrinkage present in the EBV. On the other hand, the second one resulted in a measure, $\ln\_\sigma^2_{\hat{e}}$, with less adjustments compared to deregression. However, some non-genetic effects, such as the contemporary group (CG) effect on the variance may have affected the within-family variance estimate $\ln\_\sigma^2_{\hat{e}}$, while they are accounted for when we used dEBV since non-genetic effects, were fitted in the DHGLM in the residual variance part of the model. In addition, null and non-null r_{mv} were assumed to measure the influence of the mean and/or potential scale effects on the residual variance. Therefore, analyzing each response variable, we can conclude that: i) $dEBV_v$ was greatly affected by the mean; ii) $dEBV_{v\_r0}$ had low accuracy by assuming null r_{mv}, which requires a large amount of phenotypic information and/or genotyped sires; and iii) $\ln\_\sigma^2_{\hat{e}}$ was less related to the mean and less regressed compared to dEBV, but it may include some non-genetic effects, like CG. It is clear that all three approaches have advantages and disadvantages. Given the fact that $\ln\_\sigma^2_{\hat{e}}$ is likely more independent to the mean than $dEBV_v$, there may be in this case a small preference for this response variable as it behaved also better in the Markov chain Monte Carlo (MCMC). Furthermore, considering multiple response variables can be beneficial, windows that appear in multiple approaches may have a higher credibility and may help in understanding the role of genes affecting the mean, the residual variance or both and whether genes controlling the variance beyond scale effects.

3.4 Conclusions

Using solutions from a DHGLM that assumed null genetic correlation between mean and residual variance was a suitable strategy to identify genomic regions affecting uniformity less dependent on the mean. Thirteen and four SNPs were associated with $dEBV_{v\_r0}$ and $\ln\_\sigma^2_{\hat{e}}$, respectively, and additional common regions were observed among the responses variables. In all these regions, potential biological candidate genes for uniformity of YW were related not only with metabolic genes, but also stress, inflammatory and immune responses, mineralization, neuronal activity and bone formation. No strong evidence was found in favor of using

a specific response variable for GWAS, while using multiple response variables was beneficial from the biological point of view, since more evidence on candidate regions for being shared can be found.

3.5 Materials and Methods

The phenotypic data used in this study are described in Iung et al. [19]. Here, genotypic information from 423 influential Nellore sires with a large number of evaluated progeny was available. The number of progeny per sire considered for the study ranged from 50 to 10,180, with an average of 411 (see distribution of progeny per sire in Additional file 2). In total, 194,628 yearling weight measurements from the progeny were used to estimate the response variables used in the GWAS.

3.5.1 Response variables in the association studies

All responses variables used solutions or residuals from DHGLM fitted in our previous study [19]. The DHGLM is an iterative approach that estimates simultaneously genetic parameters for the mean and residual variance. In our previous study, sire, contemporary group and age (linear and quadratic, nested within sex) were included as random and fixed effects, and covariates, respectively, in both mean and residual variance models. This approach was chosen because of its better predictability compared to a two-step approach, as previously stated by Iung et al. [19]. Mean of YW was evaluated through $dEBV_m$ obtained following the deregression procedure proposed by Garrick et al. [89] and using EBV of the mean from DHGLM [12]. For the residual variance of YW, different response variables were analyzed: i) $dEBV_v$: deregressed EBV of the residual variance from a DHGLM assuming non-null genetic correlation between mean and residual variance ($r_{mv} \neq 0$); ii) $dEBV_{v\_r0}$: deregressed EBV of the residual variance from a DHGLM assuming $r_{mv} = 0$; and iii) $ln\_\sigma^2_{\hat{e}}$, log-transformed variance of estimated residuals for each sire family, using residuals from the mean model of a DHGLM also assuming $r_{mv} = 0$. The log-transformation was used to reduce mean-variance relationship and correct for non-normality of the residual variances. Fitting the DHGLM assuming null r_{mv} was intended to reduce the impact of the mean of YW may have on the EBV in the residual variance part of the model, given the strong estimate of r_{mv} on YW previously

reported for this population [19]. For instance, the EBV_v and EBV_m would be even higher correlated than the r_{mv} because of using the information on mean YW for both sets of EBV. When using null r_{mv}, it is expected to have a higher possibility to find regions only affecting the residual variance. Descriptive statistics of each response variable are presented in Table 6 and the distribution of them in Additional file 2.

3.5.2 Genotypes

Genotypic information from 423 Nellore bulls was used, being 415 genotyped with the Illumina® BovineHD chip (HD) and 8 with Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) and then imputed to HD (777k) using FImpute software [90]. In the quality control of genotypes, SNPs located on the sex chromosomes, with minor allele frequency (MAF) lower than 0.02, p-value for Hardy-Weinberg equilibrium test (HWE) less than 10^{-5} and highly correlated SNPs ($r^2 > 0.99$) within a window of 100 consecutive SNPs were removed, so that 333,877 SNPs remained for analysis. All genotyped bulls had a call rate higher than 0.90, passing the quality control.

Table 6. Descriptive statistics for mean and residual variance of yearling weight

Phenotypes	N	Mean (SD)	Minimum	Maximum
Mean				
$dEBV_m$	423	6.084 (15.109)	-67.120	45.790
Uniformity				
$dEBV_v$	423	0.073 (0.248)	-0.920	0.775
$dEBV_{v\_r0}$	423	2.55E-05 (0.442)	-1.403	1.494
$ln\_\sigma^2_{\hat{e}}$	423	6.313 (0.273)	5.505	7.295

$dEBV_m$ and $dEBV_v$: deregressed EBV for mean and residual variance of yearling weight, respectively; $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$: deregressed EBV for residual variance and log-transformed variance of estimated residuals, respectively, both assuming null genetic correlation between mean and residual variance; N: number of observations; SD: standard deviation.

3.5.3 GWAS

The SNP effects were estimated using Bayes C method [91], a mixture model which assumes that there is a smaller fraction of SNPs (1-π) with large effects and a large proportion (π) with zero or near zero effects on the trait, as follows:

$$y = 1\mu + \sum_{i=1}^{N} z_i a_i \delta_i + e$$

where **y** is the vector of response variables ($dEBV_m$, $dEBV_v$, $dEBV_{v\_r0}$ or $ln\_\sigma^2_{\hat{e}}$), **μ** is the overall mean, **1** is a vector of ones, $\mathbf{z_i}$ is the vector of genotypes of the animals for the ith SNP, $\mathbf{a_i}$ is the allele substitution effect of the ith SNP, $\boldsymbol{\delta}_i$ is an indicator variable set to 1 if the ith SNP has an non-zero effect on the trait and to 0 otherwise, **e** is the vector of random residual effects and N is the number of SNPs. It was assumed $\mathbf{a}_i \sim N(0, \sigma^2_a)$ and $\mathbf{e} \sim N(0, \mathbf{R}\sigma^2_e)$, where σ^2_a is the variance of SNP effects, σ^2_e is the residual variance and **R** is a diagonal matrix whose elements account for heterogeneous residual variance across observations. When dEBV was the response variable, the diagonal elements of **R** were derived following Garrick et al. [89], whereas when $ln\_\sigma^2_{\hat{e}}$ was the response, such elements were equal to the reciprocal of the number of progeny of each sire. The **R** matrix accounted for differences in sampling variance due to different numbers of progeny per sire. In addition, scaled inverse chi-squared prior distributions with v degrees of freedom and scale parameter S were assumed for σ^2_a and σ^2_e. In our study, π was fixed at 0.999, which means that 0.1% of the SNPs fitted in the model, i.e. about 334, have an effect on the trait. Such a criterion was assumed aiming to obtain stronger signals of candidate QTL and also due to the limiting number of genotypes.

The analyses were performed using the GS3 software [92]. Chains of 550,000 and 250,000 iterations, after discarding the first 150,000 and 50,000 as burn-in, were generated for analyses pertaining to the three response variables for residual variance and mean of YW, respectively, saving samples every 50 iterations in both cases. Convergence was assessed through Geweke test [93] using CODA R package [94].

The significance of each association was evaluated through the Bayes Factor (BF):

$$BF = \frac{\left(\frac{p}{1-p}\right)}{\left(\frac{\pi}{1-\pi}\right)}$$

where **p** and **π** are the posterior and prior probability of a SNP being included in the model, i.e. having a non-zero effect, respectively. Suggestive and strong evidences were assigned to SNPs with BF greater than 3 and 20, respectively [95,96]. In addition, the proportion of genetic variance explained by SNPs within non-

overlapping 1-Mb windows was calculated. A window size of 1-Mb was chosen since SNPs within this distance presented moderate to high (> 0.2) pairwise linkage disequilibrium (LD; measured by r^2) (see Additional file 3). In total, SNPs were allocated to 2,522 windows, with an average (SD) of 132 (42) SNPs. The top 20 windows that explained the largest proportion of genetic variance was compared between response variables.

3.5.4 Functional annotation

The shared 1-Mb windows (among the top 20 that explained the largest proportion of genetic variance) between response variables and windows with SNPs that showed a strong association (BF > 20) were screened to identify potential candidate genes using Ensembl Genome Browser (http://www.ensembl.org), based on the UMD3.1 bovine genome assembly [97]. Previously reported QTL regions overlapping such windows were identified from Cattle QTLdb [98]. Functional enrichment based on Gene Ontology (GO) terms and pathway analysis using Blast2GO PRO and Reactome [99,100], respectively, was performed on these regions in order to help us to determine potential candidate genes for uniformity. Only the potential candidate genes, according to their function, pathway or related biological process, were discussed here (the complete gene list and annotations are shown in Additional file 1). The LD maps within each 1-Mb window discussed here were obtained using the R package LDheatmap [101] (Additional file 3).

3.6 References

1. Falconer DS, Mackay TFC. Introduction to quantitative genetics. Essex: Pearson; 1996.

2. Mulder HA, Rönnegård L, Fikse WF, Veerkamp RF, Strandberg E. Estimation of genetic variance for macro- and micro-environmental sensitivity using double hierarchical generalized linear models. Genet Sel Evol. 2013;45:23.

3. Vandenplas J, Bastin C, Gengler N, Mulder HA. Genetic variance in micro-environmental sensitivity for milk and milk quality in Walloon Holstein cattle. J Dairy Sci. 2013;96:5977–90.

4. Sell-Kubiak E, Bijma P, Knol EF, Mulder HA. Comparison of methods to study uniformity of traits: Application to birth weight in pigs. J Anim Sci. 2015;93:900–11.

5. Marjanovic J, Mulder HA, Khaw HL, Bijma P. Genetic parameters for uniformity of harvest weight and body size traits in the GIFT strain of Nile tilapia. Genet Sel Evol. 2016;48:41.

6. Mulder HA, Visscher J, Fablet J. Estimating the purebred-crossbred genetic correlation for uniformity of eggshell color in laying hens. Genet Sel Evol. 2016;48:39.

7. Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, et al. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. Anim Genet. 2012;43:87–96.

8. Wang X, Liu X, Deng D, Yu M, Li X. Genetic determinants of pig birth weight variability. BMC Genet. 2016;17:41.

9. Wang Y, Ding X, Tan Z, Ning C, Xing K, Yang T, et al. Genome-Wide Association Study of Piglet Uniformity and Farrowing Interval. Front Genet. 2017;8:1–9.

10. Rönnegård L, Felleki M, Fikse F, Mulder HA, Strandberg E. Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. Genet Sel Evol. 2010;42:8.

11. Mulder HA, Crump RE, Calus MPL, Veerkamp RF. Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms. J Dairy Sci. 2013;96:7306–17.

12. Felleki M, Lee D, Lee Y, Gilmour AR, Rönnegård L. Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. Genet Res. 2012;94:307–17.

13. Sell-Kubiak E, Duijvesteijn N, Lopes MS, Janss LLG, Knol EF, Bijma P, et al. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. BMC Genomics. 2015;16:1049.

14. Utsunomiya YT, do Carmo AS, Carvalheiro R, Neves HHR, Matos MC, Zavarez LB, et al. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. BMC Genet. 2013;14:52.

15. Saatchi M, Schnabel RD, Taylor JF, Garrick DJ. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics. 2014;15:442.

16. Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, et al. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. BMC Genomics. 2017;18:386.

17. Neves HHR, Carvalheiro R, Roso VM, Queiroz SA. Genetic variability of residual variance of production traits in Nellore beef cattle. Livest Sci. 2011;142:164–9.

18. Neves HHR, Carvalheiro R, Queiroz SA. Genetic and environmental heterogeneity of residual variance of weight traits in Nellore beef cattle. Genet Sel Evol. 2012;44:19.

19. Iung LHS, Neves HHR, Mulder HA, Carvalheiro R. Genetic control of residual variance of yearling weight in Nellore beef cattle. J Anim Sci. 2017;95:1425–33.

20. Casas E, Shackelford SD, Keele JW, Koohmaraie M, Smith TPL, Stone RT. Detection of quantitative trait loci for growth and carcass composition in cattle. J Anim Sci. 2003;81:2976–83.

21. McClure MC, Morsci NS, Schnabel RD, Kim JW, Yao P, Rolf MM, et al. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. Anim Genet. 2010;41:597–607.

22. Sikora KM, Magee DA, Berkowicz EW, Berry DP, Howard DJ, Mullen MP, et al. DNA sequence polymorphisms within the bovine guanine nucleotide-binding protein Gs subunit alpha (Gsα)-encoding (GNAS) genomic imprinting domain are associated with performance traits. BMC Genet. 2011;12:4.

23. Lindholm-Perry AK, Kuehn LA, Smith TPL, Ferrell CL, Jenkins TG, Freetly HC, et al. A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle. Anim Genet. 2012;43:216–9.

24. Júnior GAF, Costa RB, de Camargo GMF, Carvalheiro R, Rosa GJM, Baldi F, et al. Genome scan for postmortem carcass traits in Nellore cattle. J Anim Sci. 2016;94:4087–95.

25. Magalhães AFB, de Camargo GMF, Fernandes GA, Gordo DGM, Tonussi RL, Costa RB, et al. Genome-Wide Association Study of Meat Quality Traits in Nellore Cattle. PLoS One. 2016;11:e0157845.

26. Silva RMO, Stafuzza NB, Fragomeni BO, de Camargo GMF, Ceacero TM, Cyrillo JNSG, et al. Genome-wide association study for carcass traits in an experimental Nelore cattle population. PLoS One. 2017;12:e0169860.

27. Ramayo-Caldas Y, Fortes MRS, Hudson NJ, Porto-Neto LR, Bolormaa S, Barendse W, et al. A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G, and FOXP3 in intramuscular fat deposition of beef cattle. J Anim Sci. 2014;92:2832–45.

28. Casas E, Keele JW, Shackelford SD, Koohmaraie M, Stone RT. Identification of quantitative trait loci for growth and carcass composition in cattle. Anim Genet. 2004;35:2–6.

29. Voz ML, Agten NS, Van De Ven WJM, Kas K. PLAG1, the main translocation target in pleomorphic adenoma of the salivary glands, is a positive regulator of IGF-II. Cancer Res. 2000;60:106–13.

30. Nishimura I, Chano T, Kita H, Matsusue Y, Okabe H. RB1CC1 protein suppresses type II collagen synthesis in chondrocytes and causes dwarfism. J Biol Chem. 2011;286:43925–32.

31. Luo QB, Song XY, Ji CL, Zhang XQ, Zhang DX. Exploring the molecular mechanism of acute heat stress exposure in broiler chickens using gene expression profiling. Gene. 2014;546:200–5.

32. Li H, Kentish SJ, Wittert GA, Page AJ. The role of neuropeptide W in energy homeostasis. Acta Physiol. 2018;222:e12884.

33. Bali A, Randhawa PK, Jaggi AS. Stress and opioids: role of opioids in modulating stress-related behavior and effect of stress on morphine conditioned place preference. Neurosci Biobehav Rev. 2015;51:138–50.

34. Zhang CS, Jiang B, Li M, Zhu M, Peng Y, Zhang YL, et al. The lysosomal v-ATPase-ragulator complex is a common activator for AMPK and mTORC1, acting as a switch between catabolism and anabolism. Cell Metab. 2014;20:526–40.

35. Abo-Ismail MK, Kelly MJ, Squires EJ, Swanson KC, Bauck S, Miller SP. Identification of single nucleotide polymorphisms in genes involved in digestive and metabolic processes associated with feed efficiency and performance traits in beef cattle. J Anim Sci. 2013;91:2512–29.

36. Mead JR, Irvine SA, Ramji DP. Lipoprotein lipase: structure, function, regulation, and role in disease. J Mol Med. 2002;80:753–69.

37. Wang XP, Luoreng ZM, Li F, Wang JR, Li N, Li SH. Genetic polymorphisms of lipoprotein lipase gene and their associations with growth traits in Xiangxi cattle. Mol Biol Rep. 2012;39:10331–8.

38. Oh D, La B, Lee Y, Byun Y, Lee J, Yeo G, et al. Identification of novel single nucleotide polymorphisms (SNPs) of the lipoprotein lipase (LPL) gene associated with fatty acid composition in Korean cattle. Mol Biol Rep. 2013;40:3155–63.

39. Ding XZ, Liang CN, Guo X, Xing CF, Bao PJ, Chu M, et al. A novel single nucleotide polymorphism in exon 7 of LPL gene and its association with carcass traits and visceral fat deposition in yak (Bos grunniens) steers. Mol Biol Rep. 2012;39:669–73.

40. Gui L, Jia C, Zhang Y, Zhao C, Zan L. Association studies on the bovine lipoprotein lipase gene polymorphism with growth and carcass quality traits in Qinchuan cattle. Mol Cell Probes. 2016;30:61–5.

41. Zhou T, Li S, Zhong W, Vihervaara T, Béaslas O, Perttilä J, et al. Osbp-related protein 8 (ORP8) regulates plasma and liver tissue lipid levels and interacts with the nucleoporin Nup62. PLoS One. 2011;6:e21078.

42. Straface G, Aprahamian T, Flex A, Gaetani E, Biscetti F, Smith RC, et al. Sonic hedgehog regulates angiogenesis and myogenesis during post-natal skeletal muscle regeneration. J Cell Mol Med. 2009;13:2424–35.

43. Petrova R, Joyner AL. Roles for Hedgehog signaling in adult organ homeostasis and repair. Development. 2014;141:3445–57.

44. Rodier A, Rochard P, Berthet C, Rouault JP, Casas F, Daury L, et al. Identification of functional domains involved in BTG1 cell localization. Oncogene. 2001;20:2691–703.

45. Yun S Il, Kim KK. Ubiquitin-specific protease 4 (USP4) suppresses myoblast differentiation by down regulating MyoD activity in a catalytic-independent manner. Cell Signal. 2017;35:48–60.

46. Wing SS. Deubiquitinating enzymes in skeletal muscle atrophy - An essential role for USP19. Int J Biochem Cell Biol. 2016;79:462–8.

47. Ni Y, Ji C, Wang B, Qiu J, Wang J, Guo X. A Novel pro-adipogenesis factor abundant in adipose tissues and over-expressed in obesity acts upstream of PPARγ and C/EBPα. J Bioenerg Biomembr. 2013;45:219–28.

48. Brockman RP, Laarveld B. Hormonal regulation of metabolism in ruminants; a review. Livest Prod Sci. 1986;14:313–34.

49. Belhadj Slimen I, Najar T, Ghram A, Abdrrabba M. Heat stress effects on livestock: Molecular, cellular and metabolic aspects, a review. J Anim Physiol Anim Nutr. 2016. 401–12.

50. Blackshaw JK, Blackshaw AW. Heat stress in cattle and the effect of shade on production and behaviour: A review. Aust J Exp Agric. 1994;34:285–95.

51. Möstl E, Palme R. Hormones as indicators of stress. Domest Anim Endocrinol. 2002. 67–74.

52. Vale W, Spiess J, Rivier C, Rivier J. Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and beta-endorphin. Science. 1981;213:1394–7.

53. Bale TL, Vale WW. CRF AND CRF R ECEPTORS : Role in Stress Responsivity and Other Behaviors. Annu Rev Pharmacol Toxicol. 2004;44:525–57.

54. Lee J-W, Bae S-H, Jeong J-W, Kim S-H, Kim K-W. Hypoxia-inducible factor (HIF-1)alpha: its protein stability and biological functions. Exp Mol Med. 2004;36:1–12.

55. Elolimy AA, McCann JC, Shike DW, Loor JJ. 443 Residual feed intake in beef cattle and its association with ruminal epithelium gene expression. J Anim Sci. 2017;95:217–8.

56. Kampinga HH, Hageman J, Vos MJ, Kubota H, Tanguay RM, Bruford EA, et al. Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones. 2009;14:105–11.

57. Sørensen JG, Kristensen TN, Loeschcke V. The evolutionary and ecological role of heat shock proteins. Ecol Lett. 2003;6:1025–37.

58. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. Nature. 2002;417:618–24.

59. Rutherford S, Hirate Y, Swalla BJ. The Hsp90 Capacitor, Developmental Remodeling, and Evolution: The Robustness of Gene Networks and the Curious Evolvability of Metamorphosis. Crit Rev Biochem Mol Biol. 2007;42:355–72.

60. Sangster TA, Bahrami A, Wilczek A, Watanabe E, Schellenberg K, McLellan C, et al. Phenotypic diversity and altered environmental plasticity in Arabidopsis thaliana with reduced Hsp90 levels. PLoS One. 2007;2:e648.

61. Sangster TA, Salathia N, Undurraga S, Milo R, Schellenberg K, Lindquist S, et al. HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. Proc Natl Acad Sci U S A. 2008;105:2963–8.

62. Morrison BH, Bauer JA, Kalvakolanu D V., Lindner DJ. Inositol Hexakisphosphate Kinase 2 Mediates Growth Suppressive and Apoptotic Effects of Interferon-β in Ovarian Carcinoma Cells. J Biol Chem. 2001;276:24965–70.

63. Chakraborty A, Koldobskiy MA, Sixt KM, Juluri KR, Mustafa AK, Snowman AM, et al. HSP90 regulates cell survival via inositol hexakisphosphate kinase-2. Proc Natl Acad Sci U S A. 2008;105:1134–9.

64. Coble DJ, Fleming D, Persia ME, Ashwell CM, Rothschild MF, Schmidt CJ, et al. RNA-seq analysis of broiler liver transcriptome reveals novel responses to high ambient temperature. BMC Genomics. 2014;15:1–12.

65. Kang SM, Narducci MG, Lazzeri C, Mongiovì AM, Caprini E, Bresin A, et al. Impaired T- and B-cell development in Tcl1-deficient mice. Blood. 2005;105:1288–94.

66. Leonard JE, Johnson DE, Felsen RB, Tanney LE, Royston I, Dillman RO. Establishment of a human B-cell tumor in athymic mice. Cancer Res. 1987;47:2899–902.

67. Skeel A, Leonard EJ. Action and target cell specificity of human macrophage-stimulating protein (MSP). J Immunol. 1994;152:4618–23.

68. Hall A. Rho GTPases and the actin cytoskeleton. Science. 1998;279:509–14.

69. Singh R, Wang B, Shirvaikar A, Khan S, Kamat S, Schelling JR, et al. The IL-1 receptor and rho directly associate to drive cell activation in inflammation. J Clin Invest. 1999;103:1561–70.

70. Konno D. The postsynaptic density and dendritic raft localization of PSD-Zip70, which contains an N-myristoylation sequence and leucine-zipper motifs. J Cell Sci. 2002;115:4695–706.

71. Savaskan NE, Bräuer AU, Nitsch R. Molecular cloning and expression regulation of PRG-3, a new member of the plasticity-related gene family. Eur J Neurosci. 2004;19:212–20.

72. Kominakis A, Hager-Theodorides AL, Zoidis E, Saridaki A, Antonakos G, Tsiamis G. Combined GWAS and “guilt by association”-based prioritization analysis identifies functional candidate genes for body size in sheep. Genet Sel Evol. 2017;49:41.

73. Fisher LW, Fedarko NS. Six genes expressed in bones and teeth encode the current members of the SIBLING family of proteins. Connect Tissue Res. 2003;44(Suppl1):33–40.

74. Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, et al. Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. BMC Genet. 2012;13:40.

75. Cao XK, Zhan ZY, Huang YZ, Lan XY, Lei CZ, Qi XL, et al. Variants and haplotypes within MEF2C gene influence stature of chinese native cattle including body dimensions and weight. Livest Sci. 2016;185:106–9.

76. Han YJ, Chen Y, Liu Y, Liu XL. Sequence variants of the LCORL gene and its association with growth and carcass traits in Qinchuan cattle in China. J Genet. 2017;96:9–17.

77. Backx L, Vermeesch J, Pijkels E, de Ravel T, Seuntjens E, Van Esch H. PPP2R2C, a gene disrupted in autosomal dominant intellectual disability. Eur J Med Genet. 2010;53:239–43.

78. Sigoillot SM, Iyer K, Binda F, González-Calvo I, Talleur M, Vodjdani G, et al. The secreted protein C1QL1 and its receptor BAI3 control the synaptic connectivity of excitatory inputs converging on cerebellar purkinje cells. Cell Rep. 2015;10:820–32.

79. Penney J, Tsai LH. JAKMIP1: Translating the Message for Social Behavior. Neuron. 2015;88:1070–2.

80. Libri V, Schulte D, van Stijn A, Ragimbeau J, Rogge L, Pellegrini S. Jakmip1 is expressed upon T cell differentiation and has an inhibitory function in cytotoxic T lymphocytes. J Immunol. 2008;181:5847–56.

81. Lin S, Marin EC, Yang CP, Kao CF, Apenteng BA, Huang Y, et al. Extremes of lineage plasticity in the drosophila brain. Curr Biol. 2013;23:1908–13.

82. Lanet E, Gould AP, Maurange C. Protection of Neuronal Diversity at the Expense of Neuronal Numbers during Nutrient Restriction in the Drosophila Visual System. Cell Rep. 2013;3:587–94.

83. Mellert DJ, Williamson WR, Shirangi TR, Card GM, Truman JW. Genetic and environmental control of neurodevelopmental robustness in Drosophila. PLoS One. 2016;11:e0155957.

84. Yang Y, Christensen OF, Sorensen D. Analysis of a genetically structured variance heterogeneity model using the Box–Cox transformation. Genet Res. 2011;93:33–46.

85. Sonesson AK, Ødegård J, Rönnegård L. Genetic heterogeneity of within-family variance of body weight in Atlantic salmon (Salmo salar). Genet Sel Evol. 2013;45:41.

86. Sae-Lim P, Kause A, Janhunen M, Vehviläinen H, Koskinen H, Gjerde B, et al. Genetic (co)variance of rainbow trout (Oncorhynchus mykiss) body weight and its uniformity across production environments. Genet Sel Evol. 2015;47:46.

87. Hill WG, Mulder HA. Genetic analysis of environmental variation. Genet Res. 2010;92:381–95.

88. Spencer CCA, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 2009;5:e1000477.

89. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet Sel Evol. 2009;41:55.

90. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014;15:478.

91. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12:186.

92. Legarra A, Ricard A, Filangi O. GS3: Genomic Selection, Gibbs Sampling, Gauss Seidel (and BayesCπ). 2014. http://snp.toulouse.inra.fr/~alegarra/manualgs3_last.pdf.

93. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo JM, Berger J, Dawid AP, Smith AFM, editors. Bayesian Statistics 4. Oxford: Oxford University Press; 1992. p. 169–93.

94. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence diagnostics and output analysis for MCMC. R News. 2006;6:7–11.

95. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc. 1995;90:773.

96. Vidal O, Noguera JL, Amills M, Varona L, Gil M, Jiménez N, et al. Identification of carcass and meat quality quantitative trait loci in a Landrace pig population selected for growth and leanness. J Anim Sci. 2005;83:293–300.

97. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 2009;10:R42.

98. Hu ZL, Park CA, Wu XL, Reecy JM. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. 2013;41:871–9.

99. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.

100. Croft D, Mundo A, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014;42:D472–7.

101. Shin J-H, Blay S, McNeney B, Graham J. LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. J Stat Softw. 2006;16.

CHAPTER 4 – Final Considerations

In general, our results confirm the existence of genetic control on residual variance of yearling weight in Nellore cattle. Estimates of genetic parameters for mean and residual variance are in line with the literature. The existence of a genetic control on residual variance remained even after applying data transformation to reduce the mean-variance relationship. As a consequence of this transformation, lower genetic parameter estimates and change in sign and magnitude of the genetic correlation between mean and residual variance (r_{mv}) were verified. The better predictive performance of EBV for residual variance obtained with the double hierarquichal generalized linear model (DHGLM; even when a null r_{mv} was assumed to make it more similar to the two-step approach), highlights the advantage of the DHGLM by accounting for the accuracy of the squared residuals, represented by the diagonal elements of the hat matrix.

In a breeding program, selection to reduce residual variance is attractive since it can lead to more uniform (homogeneous) performance without depleting both genetic variance and genetic progress at the mean phenotypic level of the trait. Based on our findings, some limitation in relation to selection for uniformity of YW is expected due to the magnitude of the genetic variance of residual variance, low heritability of residual variance and positive and high r_{mv}. Some strategies may be further investigated to overcome these challenges. The first one is using genomic information to increase prediction accuracy of EBV for residual variance. Few studies on uniformity using genomic data have been published, but an increase in accuracy of EBV for residual variance was observed in all of them. The second is to use another way to measure uniformity such as standard deviation according to the data structure.

The opportunity to change uniformity through selection observed in our study and in others has been confirmed by results of selection experiments in laboratory species. However, its application in livestock must be further examined since multiple traits are selected simultaneously and potential correlated responses are expected.

In terms of genomic control, regions only controlling the mean, others the uniformity and others both were observed. These findings were mainly due to the use of multiple response variables for uniformity of yearling weight. The agreement

among functions of the potential genes in each region for each response variable highlights the importance of these mechanisms to determine uniformity. Taken together, our findings and previous studies suggest that: i) residual variance may be partially determined by regions or genes that also control the mean of the traits; and ii) the limited number of SNPs or regions associated reflects the impact of the small genetic variance related to uniformity, regardless of the response variable.

Additionally, although here we focused on a growth trait, other traits can be investigated in terms of uniformity mainly if an optimum intermediate is desired. In beef cattle, reproduction, birth weight and meat quality traits such as marbling, fat thickness and fatty acid profile fit into this situation; however, so far, only studies on these traits (days from calving to first service and fatty acids in milk) in dairy cattle were developed.

APPENDICES

Appendix A – Additional File 1

Table S1.1. Complete gene list in 1-Mb windows harboring SNPs showing a strong association (BF > 20) with $\ln\_\sigma^2_{\hat{e}}$

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000011682	*GLI2*	GLI family zinc finger 2	2	72977209	73168370
ENSBTAG00000007765	***TFCP2L1***	Transcription factor CP2 like 1	2	73379885	73441557
ENSBTAG00000048035			2	73424379	73424446
ENSBTAG00000019781	*CLASP1*	Cytoplasmic linker associated protein 1	2	73490617	73723487
ENSBTAG00000043135	*U4atac*		2	73664455	73664580
ENSBTAG00000019387	*NIFK*	Nucleolar protein interacting with the FHA domain of MKI67	2	73804850	73815299
ENSBTAG00000006059	*TSN*	Translin	2	73827082	73833492
ENSBTAG00000011590	*GLIPR1*	GLI pathogenesis related 1	5	4974615	5021025
ENSBTAG00000011591	*KRR1*	KRR1, small subunit processome component homolog	5	5020415	5034909
ENSBTAG00000034985	***PHLDA1***	Pleckstrin homology like domain family A member 1	5	5557939	5559917
ENSBTAG00000020908	*NAP1L1*	Nucleosome assembly protein 1-like 1	5	5575920	5612121
ENSBTAG00000043236	*U6*		5	5805440	5805536
ENSBTAG00000004425	*BBS10*	Bardet-Biedl syndrome 10	5	5813257	5816003
ENSBTAG00000004427	*OSBPL8*	Oxysterol binding protein like 8	5	5823987	5903115
ENSBTAG00000005966			8	67233123	67242351
ENSBTAG00000012855	***LPL***	Lipoprotein lipase	8	67481089	67511227
ENSBTAG00000043702	*U4*		8	67635228	67635310
ENSBTAG00000020788	*SLC18A1*	Solute carrier family 18 member A1	8	67724946	67781207
ENSBTAG00000018646	*ATP6V1B2*	ATPase H+ transporting V1 subunit B2	8	67798291	67824441
ENSBTAG00000000841	*LZTS1*	Leucine zipper putative tumor suppressor 1	8	67861051	67867745
ENSBTAG00000036968	*U3*		10	74049336	74049548
ENSBTAG00000020935	*HIF1A*	Hypoxia inducible factor 1 alpha subunit	10	74095881	74139364
ENSBTAG00000020937	*SNAPC1*	Small nuclear RNA activating complex polypeptide 1	10	74148207	74178126
ENSBTAG00000018947	***SYT16***	Synaptotagmin 16	10	74384705	74497702
ENSBTAG00000043474	*SNORA7*		10	74432661	74432800
ENSBTAG00000030792			10	74517549	74530270
ENSBTAG00000045741	*U1*		10	74737241	74737404
ENSBTAG00000046304	*U1*		10	74754682	74754845

In bold: gene or nearest gene from the SNP

Table S1.2. Complete gene list in 1-Mb windows harboring SNPs showing a strong association (BF > 20) with $dEBV_v$

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000008894			3	41127032	41127631
ENSBTAG00000000845	***OLFM3***	Olfactomedin 3	3	41420909	41642403
ENSBTAG00000019031	*PLPPR4*	Phospholipid phosphatase related 4	3	44078270	44130952
ENSBTAG00000043198	*U6*		3	44300129	44300234
ENSBTAG00000004591	***PLPPR5***	Phospholipid phosphatase related 5	3	44372996	44473523
ENSBTAG00000031461	*SNX7*	Sorting nexin 7	3	44657522	44775833
ENSBTAG00000029809	***bta-mir-137***		3	45448508	45448609
ENSBTAG00000046112			3	45449362	45449470
ENSBTAG00000031358	*DPYD*	Dihydropyrimidine dehydrogenase	3	45563732	46487165
ENSBTAG00000042940	*U6*		3	45755686	45755792
ENSBTAG00000047071			3	45832959	45833078
ENSBTAG00000020330	***BTG1***	BTG anti-proliferation factor 1	5	22086071	22088781
ENSBTAG00000033600			5	22333869	22334542
ENSBTAG00000033594	*C12orf74*	Chromosome 12 open reading frame 74	5	22719223	22720382
ENSBTAG00000000818	*PLEKHG7*	Pleckstrin homology and RhoGEF domain containing G7	5	22757834	22799117
ENSBTAG00000045179	*U6*		5	22762056	22762156
ENSBTAG00000000421	*EEA1*	Early endosome antigen 1	5	22806703	22884689
ENSBTAG00000029683	*U1*		12	4035320	4035486
ENSBTAG00000044554	*U6*		12	4035644	4035735
ENSBTAG00000045398	*bta-mir-2310*		14	14953690	14953753
ENSBTAG00000006188	*USH2A*	Usherin	16	19573856	20502175
ENSBTAG00000010392	***ESRRG***	Estrogen related receptor gamma	16	20595768	20789340
ENSBTAG00000009653	*TCL1B*	T-cell leukemia/lymphoma 1B	21	62099946	62103587
ENSBTAG00000019580	***TCL1A***	T-cell leukemia/lymphoma 1A	21	62131713	62136163
ENSBTAG00000021717	*BDKRB2*	Bradykinin receptor B2	21	62701850	62743837
ENSBTAG00000007052	*BDKRB1*	Bradykinin receptor B1	21	62760703	62772148
ENSBTAG00000012833	*ATG2B*	Autophagy-related protein 2 homolog B	21	62793771	62869911
ENSBTAG00000019836	*GSKIP*	GSK3B interacting protein	21	62870563	62892534
ENSBTAG00000044547	*SNORA27*		21	62871948	62872045
ENSBTAG00000044757	*SNORA27*		21	62872118	62872222
ENSBTAG00000016854	*AK7*	Adenylate kinase 7	21	62895825	62964704
ENSBTAG00000004054	*PAPOLA*	Poly(A) polymerase alpha	21	62980457	63032069
ENSBTAG00000012335	*UBA7*	Ubiquitin like modifier activating enzyme 7	22	50991902	51000966
ENSBTAG00000012322	*FAM212A*	Family with sequence similarity 212 member A	22	51001140	51002896

Continuation Table S1.2.

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000005728	*CDHR4*	Cadherin related family member 4	22	51005176	51012628
ENSBTAG00000011595	*IP6K1*	Inositol hexakisphosphate kinase 1	22	51014686	51049408
ENSBTAG00000032026	*GMPPB*	GDP-mannose pyrophosphorylase B	22	51052512	51054893
ENSBTAG00000011588	*RNF123*	Ring finger protein 123	22	51054914	51078183
ENSBTAG00000045547	*AMIGO3*	Adhesion molecule with Ig like domain 3	22	51056714	51058249
ENSBTAG00000011585	*MST1*	Macrophage stimulating 1	22	51080758	51085659
ENSBTAG00000011583	*APEH*	Acylaminoacyl-peptide hydrolase	22	51086070	51094554
ENSBTAG00000011581	*BSN*	Bassoon presynaptic cytomatrix protein	22	51104195	51175293
ENSBTAG00000011580	*DAG1*	Dystroglycan 1	22	51187154	51200326
ENSBTAG00000022632	*TCTA*	T-cell leukemia translocation altered	22	51254439	51257322
ENSBTAG00000002321	*AMT*	Aminomethyltransferase	22	51257611	51261937
ENSBTAG00000000387	*NICN1*	Nicolin 1	22	51262430	51266667
ENSBTAG00000004279	*RHOA*	Ras homolog family member A	22	51277867	51323093
ENSBTAG00000044873	*U6*		22	51328304	51328406
ENSBTAG00000011899	*USP4*	Ubiquitin specific peptidase 4	22	51335299	51374742
ENSBTAG00000011894	*C3orf62*	Chromosome 3 open reading frame 62	22	51375767	51379188
ENSBTAG00000009371	*CCDC36*	Coiled-coil domain containing 36	22	51386309	51413444
ENSBTAG00000000982	*C3orf84*	Chromosome 3 open reading frame 84	22	51418820	51426814
ENSBTAG00000000981	*KLHDC8B*	Kelch domain containing 8B	22	51427513	51432321
ENSBTAG00000013932	*CCDC71*	Coiled-coil domain containing 71	22	51435769	51439453
ENSBTAG00000022635	*LAMB2*	Laminin subunit beta 2	22	51456200	51470441
ENSBTAG00000018921	*USP19*	Ubiquitin specific peptidase 19	22	51470661	51482027
ENSBTAG00000018928	*QARS*	Glutaminyl-tRNA synthetase	22	51484253	51491566
ENSBTAG00000018924	*QRICH1*	Glutamine rich 1	22	51492855	51535040
ENSBTAG00000031837	*IMPDH2*	Inosine monophosphate dehydrogenase 2	22	51535326	51540157
ENSBTAG00000018918	*NDUFAF3*	NADH:ubiquinone oxidoreductase complex assembly factor 3	22	51540981	51542102
ENSBTAG00000029957	*bta-mir-191*		22	51543470	51543560
ENSBTAG00000029913	*bta-mir-425*		22	51543952	51544038
ENSBTAG00000018913	*DALRD3*	DALR anticodon binding domain containing 3	22	51544297	51547509
ENSBTAG00000018910	*WDR6*	WD repeat domain 6	22	51547600	51555978
ENSBTAG00000005490	*P4HTM*	Prolyl 4-hydroxylase, transmembrane	22	51556260	51569916
ENSBTAG00000005488	***ARIH2***	Ariadne RBR E3 ubiquitin protein ligase 2	22	51572989	51610467
ENSBTAG00000000191	*SLC25A20*	Solute carrier family 25 member 20	22	51617498	51652558
ENSBTAG00000014205	*PRKAR2A*	Protein kinase cAMP-dependent type II regulatory subunit alpha	22	51658044	51727840

Continuation Table S1.2.

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000021853	*IP6K2*	Inositol hexakisphosphate kinase 2	22	51747919	51771674
ENSBTAG00000021850	*NCKIPSD*	NCK interacting protein with SH3 domain	22	51773520	51785996
ENSBTAG00000021846	*CELSR3*	Cadherin EGF LAG seven-pass G-type receptor 3	22	51793227	51818651
ENSBTAG00000045827			22	51812698	51812783
ENSBTAG00000038381	*SLC26A6*	Solute carrier family 26 member 6	22	51821391	51832167
ENSBTAG00000021390	*TMEM89*	Transmembrane protein 89	22	51834383	51835268
ENSBTAG00000019096	*UQCRC1*	Ubiquinol-cytochrome c reductase core protein I	22	51847305	51856057
ENSBTAG00000045476			22	51875910	51875984
ENSBTAG00000019078	*UCN2*	Urocortin 2	22	51890963	51891301
ENSBTAG00000006752	*PFKFB4*	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	22	51898015	51933996
ENSBTAG00000009183	*SHISA5*	Shisa family member 5 (SHISA5)	22	51944064	51969644
ENSBTAG00000008406	*TREX1*	Three prime repair exonuclease 1	22	51969881	51971755
ENSBTAG00000008399	*ATRIP*	ATR interacting protein	22	51971832	51989358
ENSBTAG00000013413	*CCDC51*	Coiled-coil domain containing 51	22	51994224	52003466
ENSBTAG00000029842	***bta-mir-346***		28	41329055	41329149
ENSBTAG00000012213	*WAPL*	WAPL cohesin release factor	28	41492640	41565133
ENSBTAG00000045482	*U6*		28	41498833	41498922
ENSBTAG00000042382	*U6*		28	41535977	41536083
ENSBTAG00000042523	*U6*		28	41548683	41548785
ENSBTAG00000032800	*OPN4*	Opsin 4	28	41644490	41654402
ENSBTAG00000018707	*LDB3*	LIM domain binding 3	28	41657195	41713873
ENSBTAG00000000231	*BMPR1A*	Bone morphogenetic protein receptor type 1A	28	41817915	41875990
ENSBTAG00000003510	*MMRN2*	Multimerin 2	28	41886524	41901491
ENSBTAG00000003515	*SNCG*	Synuclein gamma	28	41902561	41907124
ENSBTAG00000032774	*ADIRF*	Adipogenesis regulatory factor	28	41912322	41914565
ENSBTAG00000017946	*FAM25A*	Family with sequence similarity 25 member A	28	41921827	41924801
ENSBTAG00000007540	*GLUD1*	Glutamate dehydrogenase 1	28	41941042	41979591

In bold: gene or nearest gene from the SNP

Table S1.3. Complete gene list in 1-Mb windows among the top 20 that explained the largest proportion of genetic variance shared by $dEBV_m$ and $dEBV_v$

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000042772	*U6*		1	91892152	91892224
ENSBTAG00000008894			3	41127032	41127631
ENSBTAG00000000845	*OLFM3*	Olfactomedin 3	3	41420909	41642403
ENSBTAG00000019031	*PLPPR4*	Phospholipid phosphatase related 4	3	44078270	44130952
ENSBTAG00000043198	*U6*		3	44300129	44300234
ENSBTAG00000004591	*PLPPR5*	Phospholipid phosphatase related 5	3	44372996	44473523
ENSBTAG00000031461	*SNX7*	Sorting nexin 7	3	44657522	44775833
ENSBTAG00000017475	*GNAS*	GNAS complex locus	13	58010287	58049012
ENSBTAG00000047223	*GNAS*	GNAS complex locus	13	58019524	58049012
ENSBTAG00000037319	*bta-mir-296*		13	58091936	58092013
ENSBTAG00000016724	*NPEPL1*	Aminopeptidase like 1	13	58207245	58222951
ENSBTAG00000003872	*STX16*	Syntaxin 16	13	58237976	58260726
ENSBTAG00000038838			13	58308600	58308740
ENSBTAG00000020397	*APCDD1L*	APC down-regulated 1 like	13	58439175	58446547
ENSBTAG00000017424	*VAPB*	VAMP associated protein B and C	13	58463228	58510710
ENSBTAG00000018053	*RAB22A*	RAB22A, member RAS oncogene family	13	58537701	58585721
ENSBTAG00000013574	*ANKRD60*	Ankyrin repeat domain 60	13	58663305	58673817
ENSBTAG00000016926	*C20orf85*	Chromosome 13 open reading frame, human C20orf85	13	58709492	58725285
ENSBTAG00000026306			14	23067159	23068103
ENSBTAG00000000878	*RB1CC1*	RB1 inducible coiled-coil 1	14	23147992	23177073
ENSBTAG00000042577	*U6*		14	23254313	23254419
ENSBTAG00000016159	*NPBWR1*	Neuropeptides B/W receptor 1	14	23263815	23264810
ENSBTAG00000000914	*OPRK1*	Opioid receptor kappa 1	14	23373836	23395443
ENSBTAG00000003450	*ATP6V1H*	ATPase H+ transporting V1 subunit H	14	23511301	23558678
ENSBTAG00000003454	*RGS20*	Regulator of G-protein signaling 20	14	23561303	23613020
ENSBTAG00000003460	*TCEA1*	Transcription elongation factor A1	14	23620858	23637730
ENSBTAG00000004243	*LYPLA1*	Lysophospholipase I	14	23651477	23668800
ENSBTAG00000001174	*MRPL15*	Mitochondrial ribosomal protein L15	14	23675383	23680928
ENSBTAG00000033627			14	23724626	23725771
ENSBTAG00000022539	*POLR2K*	RNA polymerase II subunit K	14	23735755	23737474
ENSBTAG00000005748	*SOX17*	SRY-box 17	14	23884842	23886642
ENSBTAG00000011203	*RP1*	RP1, axonemal microtubule associated	14	23990193	23999338
ENSBTAG00000047303			14	24019278	24100855
ENSBTAG00000044050	*XKR4*	XK related 4	14	24295567	24610955

Continuation Table S1.3.

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000005893	*TMEM68*	Transmembrane protein 68	14	24711327	24747118
ENSBTAG00000005898	*TGS1*	Trimethylguanosine synthase 1	14	24747219	24772713
ENSBTAG00000020034	*LYN*	LYN proto-oncogene, Src family tyrosine kinase	14	24847257	24920713
ENSBTAG00000019147	*RPS20*	Ribosomal protein S20	14	24955079	24956324
ENSBTAG00000045097	*snoU54*		14	24955769	24955835
ENSBTAG00000028889	*U1*		14	24970516	24970679
ENSBTAG00000019145	*MOS*	MOS proto-oncogene, serine/threonine kinase	14	24975950	24976948
ENSBTAG00000004022	*PLAG1*	PLAG1 zinc finger	14	25007291	25009296
ENSBTAG00000033284	*CHCHD7*	Coiled-coil-helix-coiled-coil-helix domain containing 7	14	25052885	25058779
ENSBTAG00000039031			14	25067486	25067823
ENSBTAG00000018570	*SDR16C5*	Short chain dehydrogenase/reductase family 16C, member 5	14	25105062	25117554
ENSBTAG00000040321	*SRD16C6*	Short chain dehydrogenase/reductase family 16C, member 6	14	25153583	25179651
ENSBTAG00000004924	*PENK*	Proenkephalin	14	25218586	25222991
ENSBTAG00000043923	*U6*		14	25492090	25492184
ENSBTAG00000015637	*IMPAD1*	Inositol monophosphatase domain containing 1	14	25544907	25560879
ENSBTAG00000006188	*USH2A*	Usherin	16	19573856	20502175
ENSBTAG00000010392	*ESRRG*	Estrogen related receptor gamma	16	20595768	20789340

Table S1.4. Complete gene list in 1-Mb windows among the top 20 that explained the largest proportion of genetic variance shared by $dEBV_v$ and $dEBV_{v\_r0}$

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000029683	*U1*		12	4035486	4035320
ENSBTAG00000044554	*U6*		12	4035735	4035644

Table S1.5. Complete gene list in 1-Mb windows among the top 20 that explained the largest proportion of genetic variance shared by $dEBV_v$ and $ln\_\sigma^2_{\hat{e}}$

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000012335	*UBA7*	Ubiquitin like modifier activating enzyme 7	22	50991902	51000966
ENSBTAG00000012322	*FAM212A*	Family with sequence similarity 212 member A	22	51001140	51002896
ENSBTAG00000005728	*CDHR4*	Cadherin related family member 4	22	51005176	51012628
ENSBTAG00000011595	*IP6K1*	Inositol hexakisphosphate kinase 1	22	51014686	51049408
ENSBTAG00000032026	*GMPPB*	GDP-mannose pyrophosphorylase B	22	51052512	51054893
ENSBTAG00000011588	*RNF123*	Ring finger protein 123	22	51054914	51078183
ENSBTAG00000045547	*AMIGO3*	Adhesion molecule with Ig like domain 3	22	51056714	51058249

Continuation Table S1.5.

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000011585	*MST1*	Macrophage stimulating 1	22	51080758	51085659
ENSBTAG00000011583	*APEH*	Acylaminoacyl-peptide hydrolase	22	51086070	51094554
ENSBTAG00000011581	*BSN*	Bassoon presynaptic cytomatrix protein	22	51104195	51175293
ENSBTAG00000011580	*DAG1*	Dystroglycan 1	22	51187154	51200326
ENSBTAG00000022632	*TCTA*	T-cell leukemia translocation altered	22	51254439	51257322
ENSBTAG00000002321	*AMT*	Aminomethyltransferase	22	51257611	51261937
ENSBTAG00000000387	*NICN1*	Nicolin 1	22	51262430	51266667
ENSBTAG00000004279	*RHOA*	Ras homolog family member A	22	51277867	51323093
ENSBTAG00000044873	*U6*		22	51328304	51328406
ENSBTAG00000011899	*USP4*	Ubiquitin specific peptidase 4	22	51335299	51374742
ENSBTAG00000011894	*C3orf62*	Chromosome 3 open reading frame 62	22	51375767	51379188
ENSBTAG00000009371	*CCDC36*	Coiled-coil domain containing 36	22	51386309	51413444
ENSBTAG00000000982	*C3orf84*	Chromosome 3 open reading frame 84	22	51418820	51426814
ENSBTAG00000000981	*KLHDC8B*	Kelch domain containing 8B	22	51427513	51432321
ENSBTAG00000013932	*CCDC71*	Coiled-coil domain containing 71	22	51435769	51439453
ENSBTAG00000022635	*LAMB2*	Laminin subunit beta 2	22	51456200	51470441
ENSBTAG00000018921	*USP19*	Ubiquitin specific peptidase 19	22	51470661	51482027
ENSBTAG00000018928	*QARS*	Glutaminyl-tRNA synthetase	22	51484253	51491566
ENSBTAG00000018924	*QRICH1*	Glutamine rich 1	22	51492855	51535040
ENSBTAG00000031837	*IMPDH2*	Inosine monophosphate dehydrogenase 2	22	51535326	51540157
ENSBTAG00000018918	*NDUFAF3*	NADH:ubiquinone oxidoreductase complex assembly factor 3	22	51540981	51542102
ENSBTAG00000029957	*bta-mir-191*		22	51543470	51543560
ENSBTAG00000029913	*bta-mir-425*		22	51543952	51544038
ENSBTAG00000018913	*DALRD3*	DALR anticodon binding domain containing 3	22	51544297	51547509
ENSBTAG00000018910	*WDR6*	WD repeat domain 6	22	51547600	51555978
ENSBTAG00000005490	*P4HTM*	Prolyl 4-hydroxylase, transmembrane	22	51556260	51569916
ENSBTAG00000005488	*ARIH2*	Ariadne RBR E3 ubiquitin protein ligase 2	22	51572989	51610467
ENSBTAG00000000191	*SLC25A20*	Solute carrier family 25 member 20	22	51617498	51652558
ENSBTAG00000014205	*PRKAR2A*	Protein kinase cAMP-dependent type II regulatory subunit alpha	22	51658044	51727840
ENSBTAG00000021853	*IP6K2*	Inositol hexakisphosphate kinase 2	22	51747919	51771674
ENSBTAG00000021850	*NCKIPSD*	NCK interacting protein with SH3 domain	22	51773520	51785996
ENSBTAG00000021846	*CELSR3*	Cadherin EGF LAG seven-pass G-type receptor 3	22	51793227	51818651
ENSBTAG00000045827			22	51812698	51812783
ENSBTAG00000038381	*SLC26A6*	Solute carrier family 26 member 6	22	51821391	51832167

Continuation Table S1.5.

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000021390	*TMEM89*	Transmembrane protein 89	22	51834383	51835268
ENSBTAG00000019096	*UQCRC1*	Ubiquinol-cytochrome c reductase core protein I	22	51847305	51856057
ENSBTAG00000045476			22	51875910	51875984
ENSBTAG00000019078	*UCN2*	Urocortin 2	22	51890963	51891301
ENSBTAG00000006752	*PFKFB4*	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	22	51898015	51933996
ENSBTAG00000009183	*SHISA5*	Shisa family member 5 (SHISA5	22	51944064	51969644
ENSBTAG00000008406	*TREX1*	Three prime repair exonuclease 1	22	51969881	51971755
ENSBTAG00000008399	*ATRIP*	ATR interacting protein	22	51971832	51989358
ENSBTAG00000013413	*CCDC51*	Coiled-coil domain containing 51	22	51994224	52003466

Table S1.6. Complete gene list in 1-Mb windows among the top 20 that explained the largest proportion of genetic variance shared by $dEBV_{v\_r0}$ and $ln\_\sigma^2_e$

Ensembl ID	Gene symbol	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSBTAG00000006307	*HSD17B11*	Hydroxysteroid 17-beta dehydrogenase 11	6	104056137	104006456
ENSBTAG00000011354	*NUDT9*	Nudix hydrolase 9	6	104136963	104099623
ENSBTAG00000004094	*SPARCL1*	SPARC like 1	6	104202396	104149824
ENSBTAG00000045714	*DSPP*	Dentin sialophosphoprotein	6	104278722	104273700
ENSBTAG00000019132	*DMP1*	Dentin matrix acidic phosphoprotein 1	6	104320863	104304075
ENSBTAG00000017829			6	104377801	104341538
ENSBTAG00000006508	*MAN2B2*	Epididymis-specific alpha-mannosidase	6	104453232	104408001
ENSBTAG00000020598	*PPP2R2C*	Protein phosphatase 2 regulatory subunit Bgamma	6	104657218	104598245
ENSBTAG00000046671	*WFS1*	Wolframin ER transmembrane glycoprotein	6	104696495	104673701
ENSBTAG00000011170	*JAKMIP1*	Janus kinase and microtubule interacting protein 1	6	104944120	104789270
ENSBTAG00000021729			6	105015210	104991391
ENSBTAG00000035855			9	7904947	7903535
ENSBTAG00000015335	*ADGRB3*	Adhesion G protein-coupled receptor B3	9	8455993	7914952

Appendix B – Additional File 2

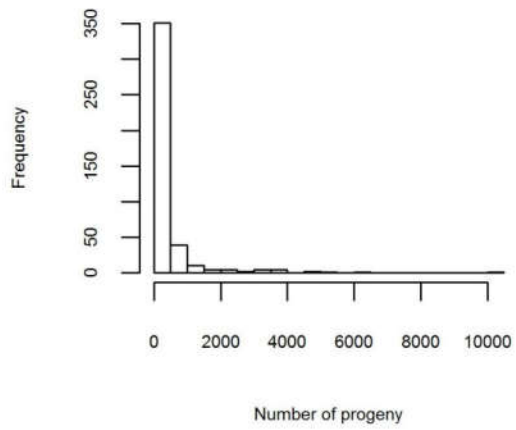

Figure 3a. Distribution of number of progeny per sire.

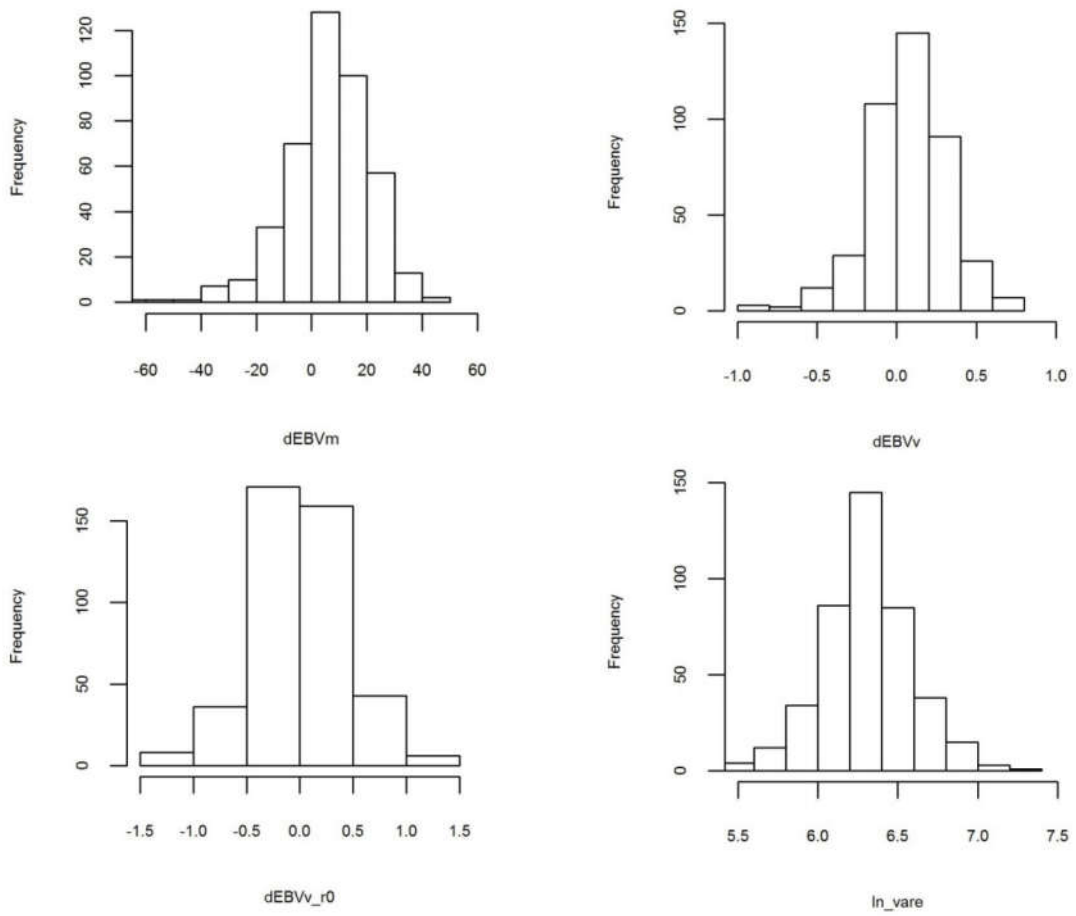

Figure 3b. Distribution of the four responses variables ($dEBV_m$, $dEBV_v$, $dEBV_{v\_r0}$ and $ln\_\sigma^2_{\hat{e}}$) used in this study.

Appendix C – Additional File 3

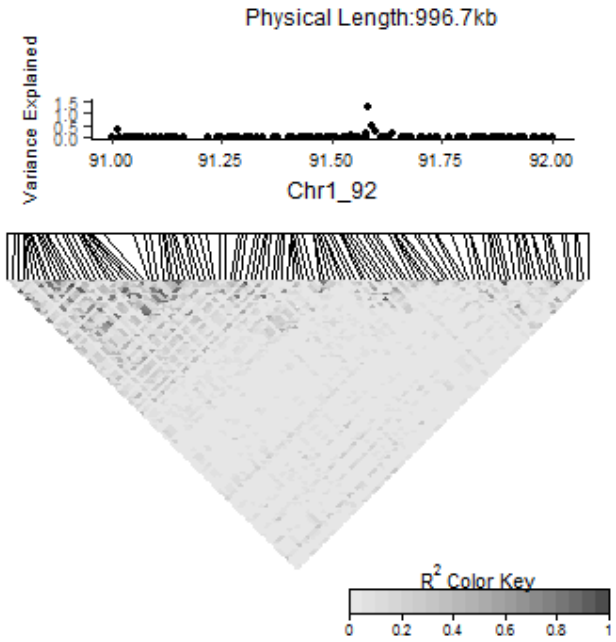

Figure 4a. Manhattan plot of the variance explained by each SNP in the window Chr1_92 for $dEBV_m$ and linkage disequilibrium (LD; r^2)

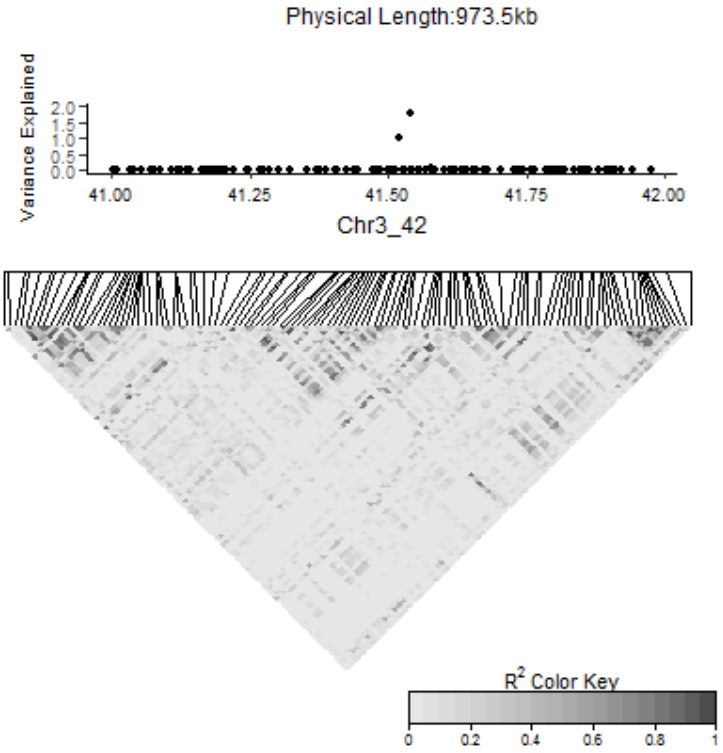

Figure 4b. Manhattan plot of the variance explained by each SNP in the window Chr3_42 for $dEBV_m$ and linkage disequilibrium (LD; r^2)

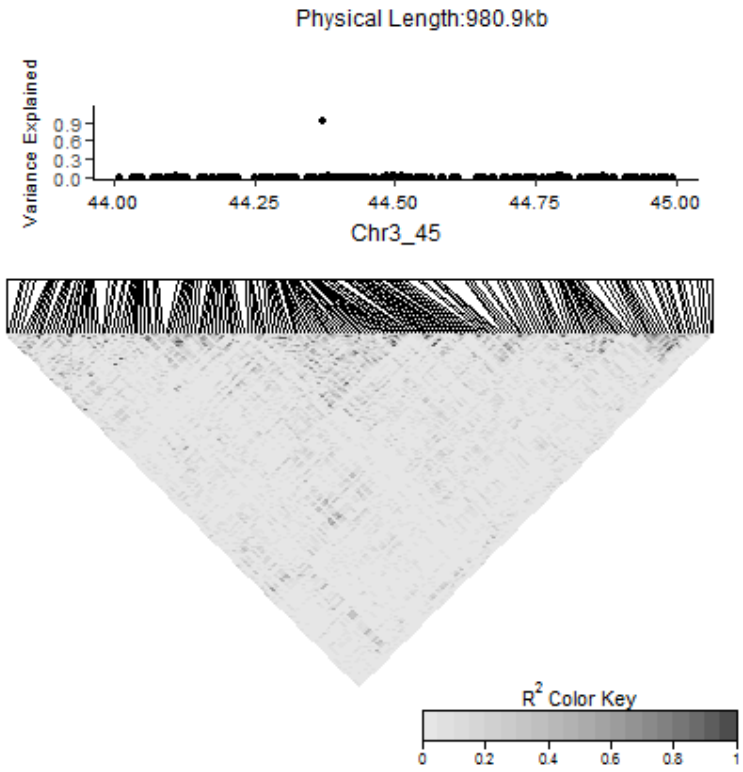

Figure 4c. Manhattan plot of the variance explained by each SNP in the window Chr3_45 for $dEBV_m$ and linkage disequilibrium (LD; r^2)

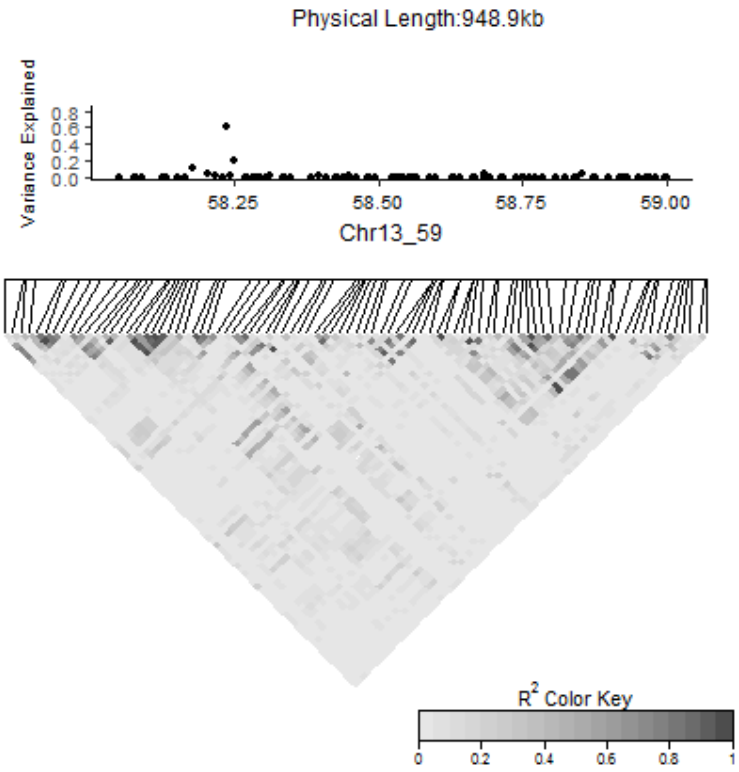

Figure 4d. Manhattan plot of the variance explained by each SNP in the window Chr13_59 for $dEBV_m$ and linkage disequilibrium (LD; r^2)

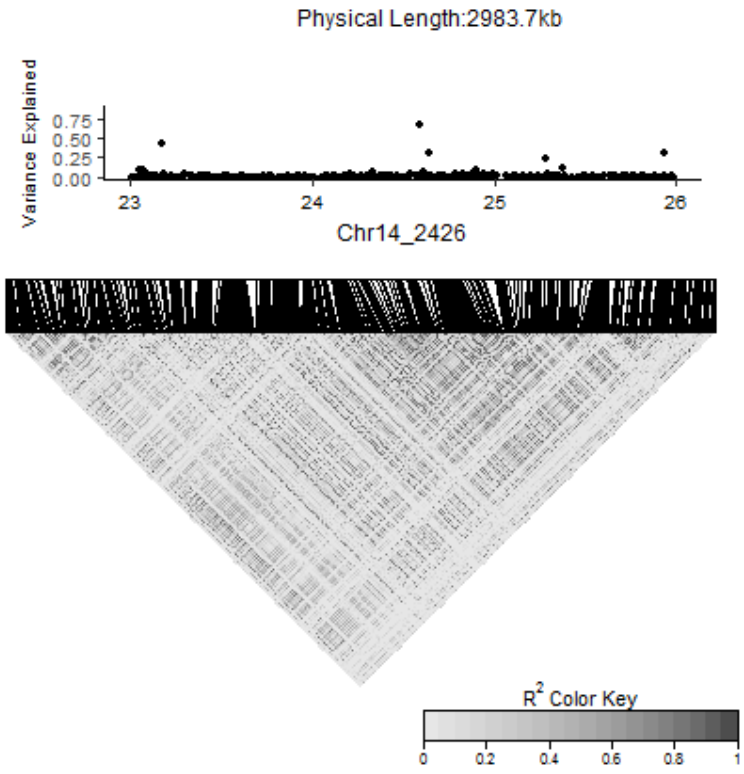

Figure 4e. Manhattan plot of the variance explained by each SNP in the window Chr14_24:26 for $dEBV_m$ and linkage disequilibrium (LD; r^2)

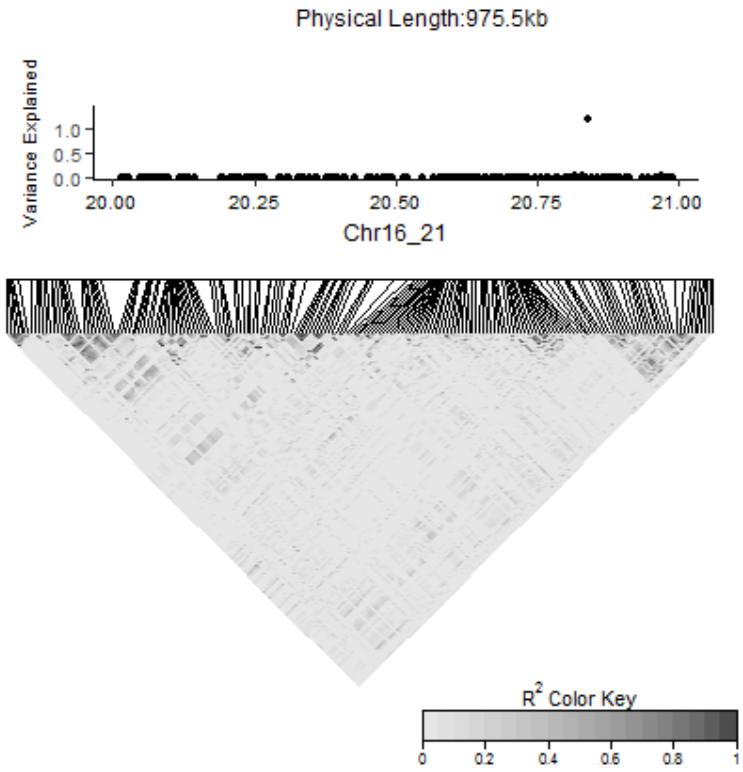

Figure 4f. Manhattan plot of the variance explained by each SNP in the window Chr16_21 for $dEBV_m$ and linkage disequilibrium (LD; r^2)

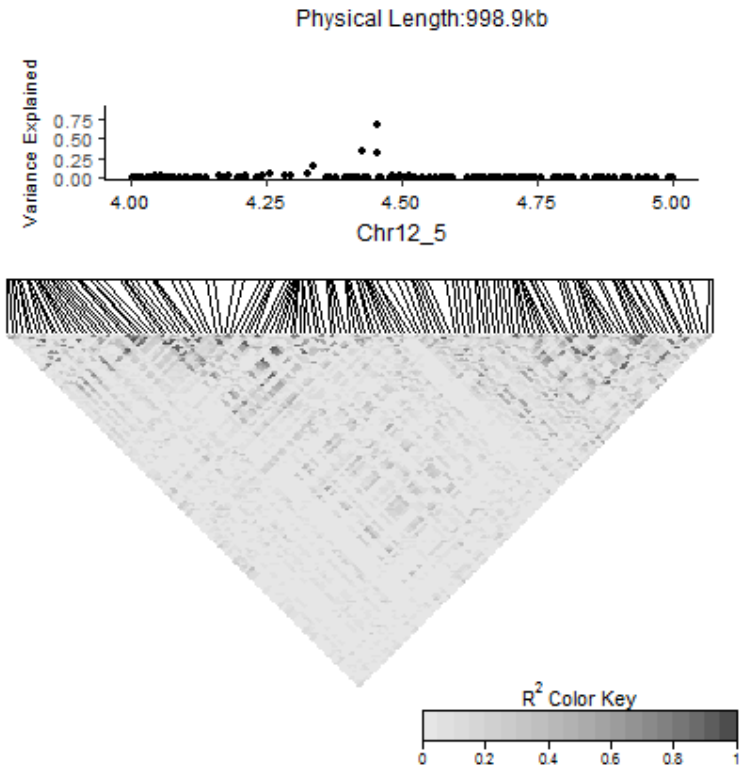

Figure 4g. Manhattan plot of the variance explained by each SNP in the window Chr12_5 for $dEBV_v$ and linkage disequilibrium (LD; r^2)

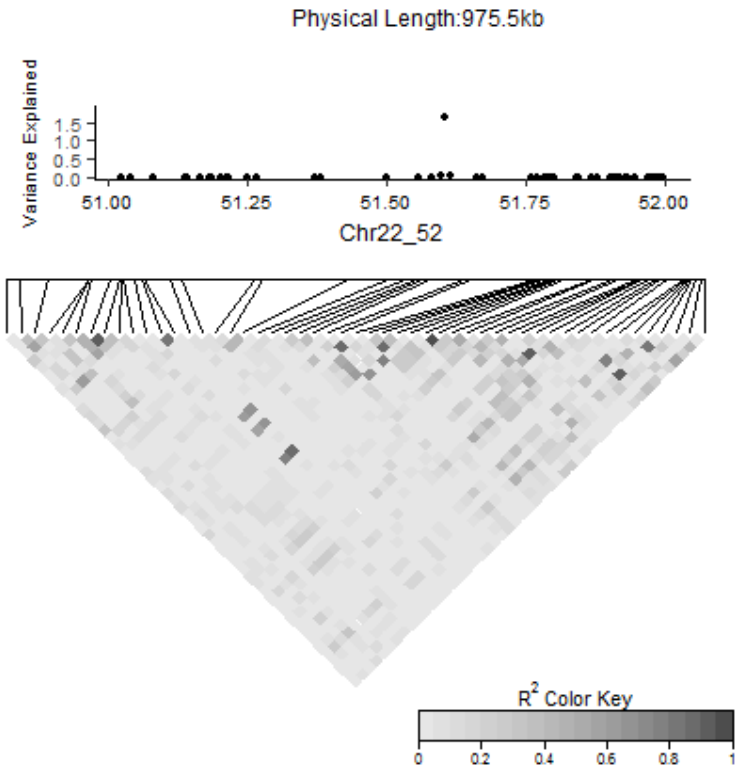

Figure 4h. Manhattan plot of the variance explained by each SNP in the window Chr22_52 for $dEBV_v$ and linkage disequilibrium (LD; r^2)

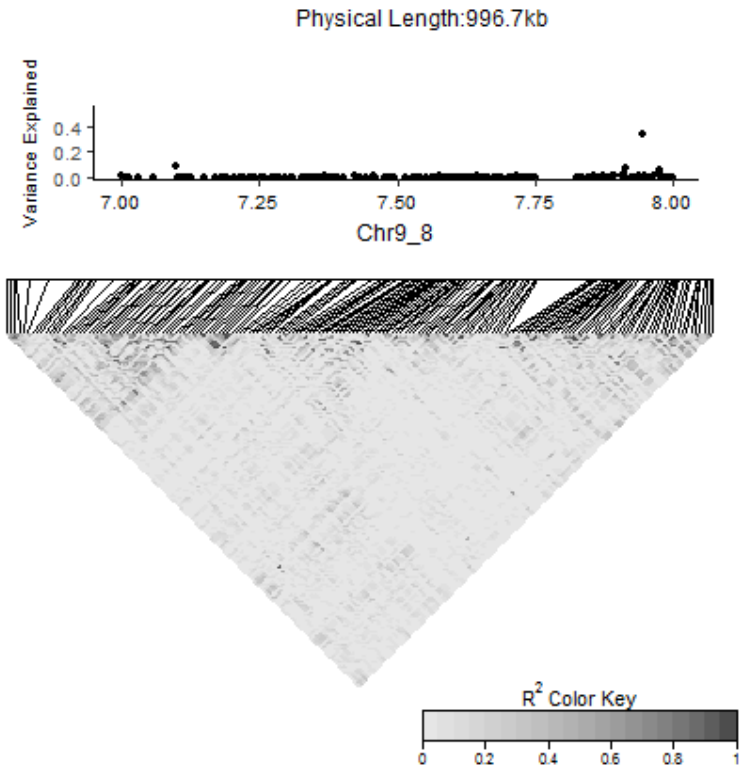

Figure 4i. Manhattan plot of the variance explained by each SNP in the window Chr9_8 for ln_$\sigma^2_{\hat{e}}$ and linkage disequilibrium (LD; r^2)

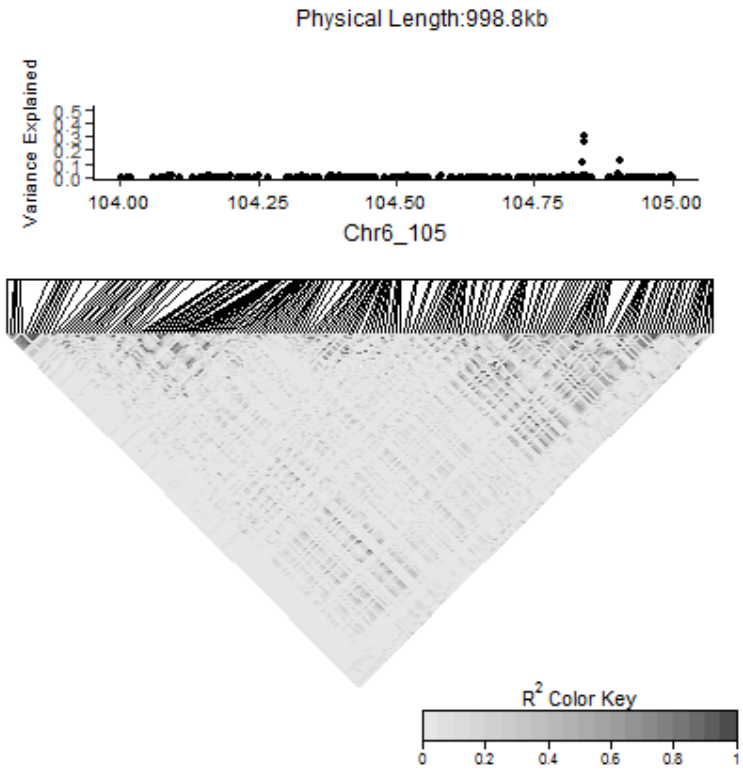

Figure 4j. Manhattan plot of the variance explained by each SNP in the window Chr6_105 for ln_$\sigma^2_{\hat{e}}$ and linkage disequilibrium (LD; r^2)

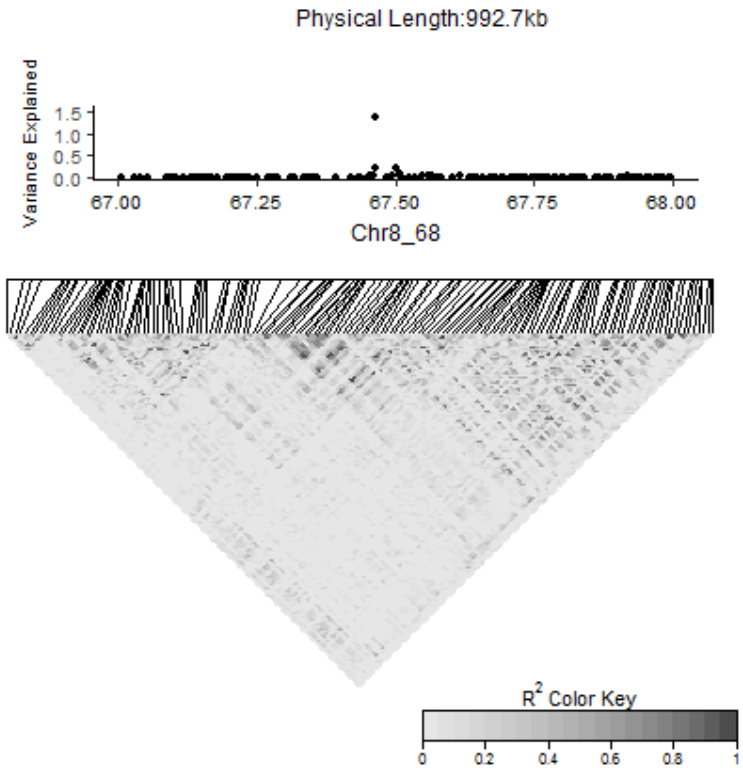

Figure 4k. Manhattan plot of the variance explained by each SNP in the window Chr8_68 for ln_$\sigma^2_{\hat{e}}$ and linkage disequilibrium (LD; r^2)

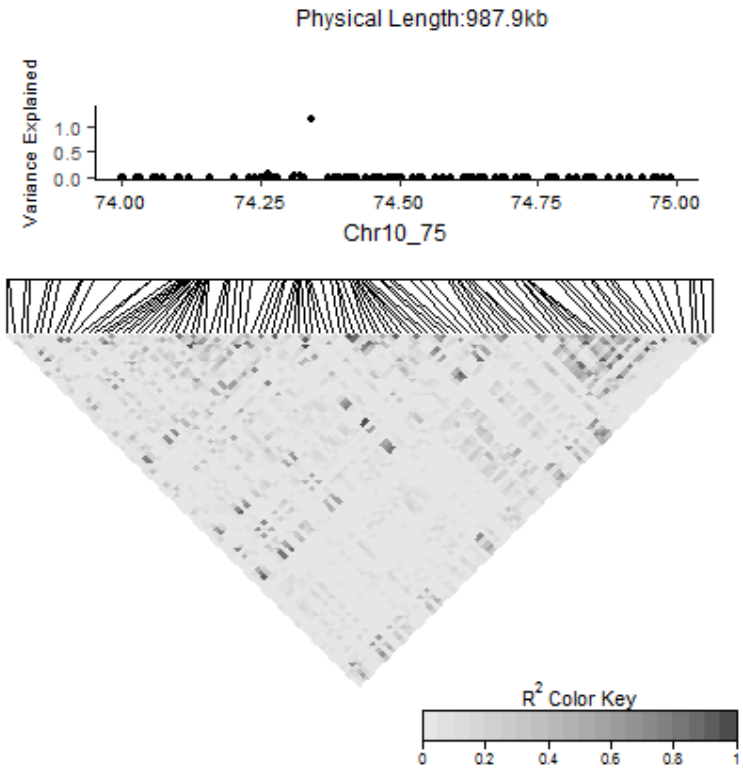

Figure 4l. Manhattan plot of the variance explained by each SNP in the window Chr10_75 for ln_$\sigma^2_{\hat{e}}$ and linkage disequilibrium (LD; r^2)

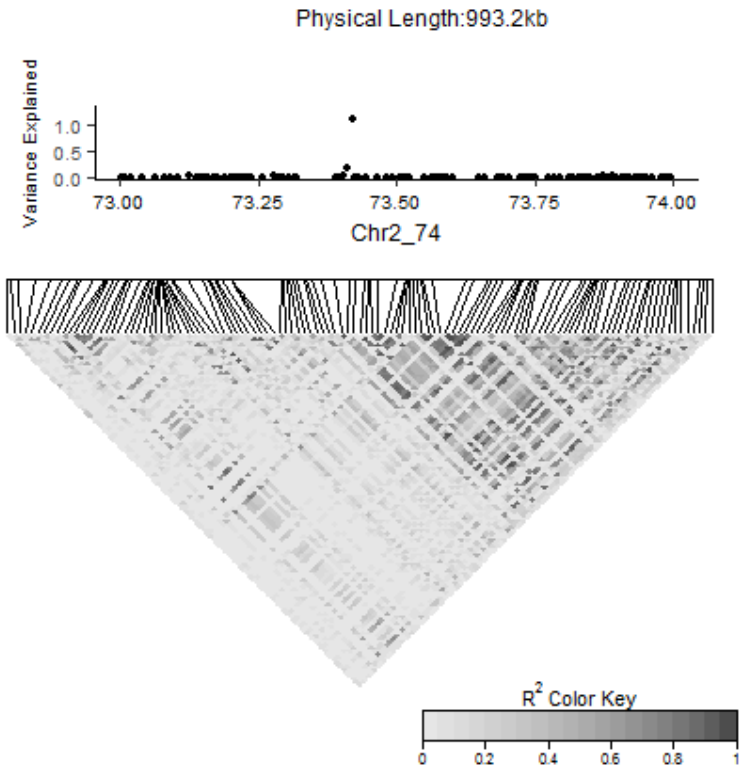

Figure 4m. Manhattan plot of the variance explained by each SNP in the window Chr2_74 for ln_$\sigma^2_{\hat{e}}$ and linkage disequilibrium (LD; r^2)

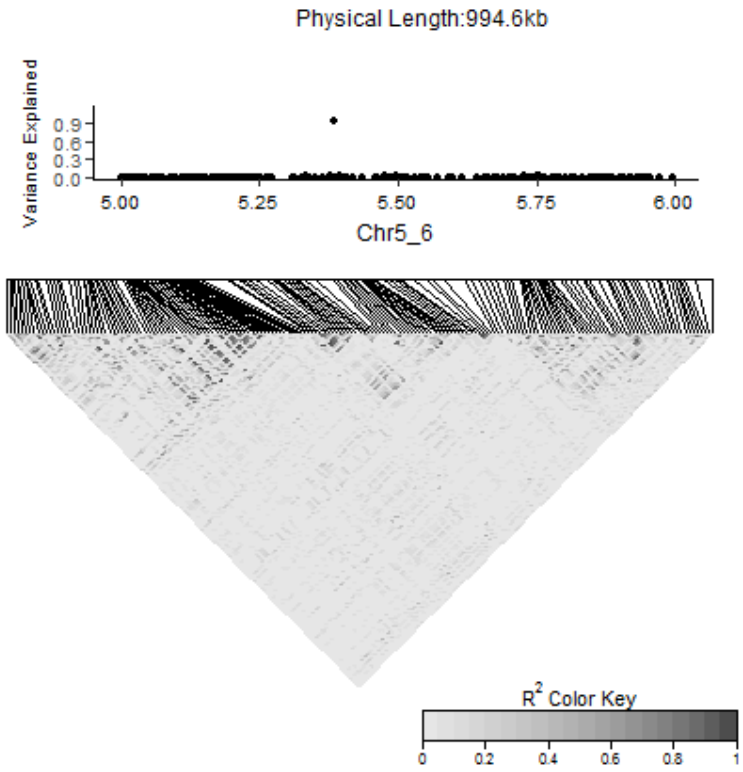

Figure 4n. Manhattan plot of the variance explained by each SNP in the window Chr5_6 for ln_$\sigma^2_{\hat{e}}$ and linkage disequilibrium (LD; r^2)

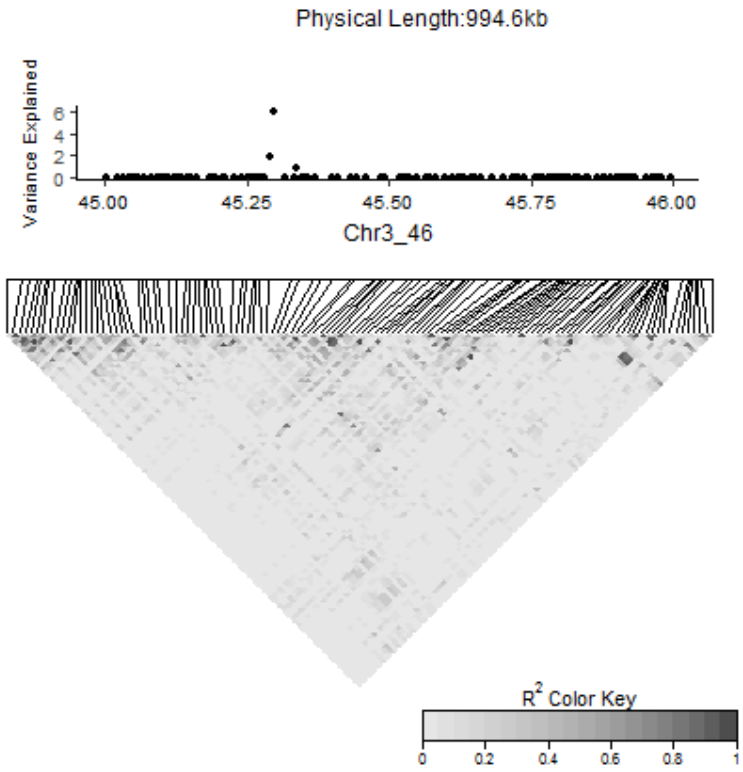

Figure 4p. Manhattan plot of the variance explained by each SNP in the window Chr3_46 for $dEBV_v$ and linkage disequilibrium (LD; r^2)

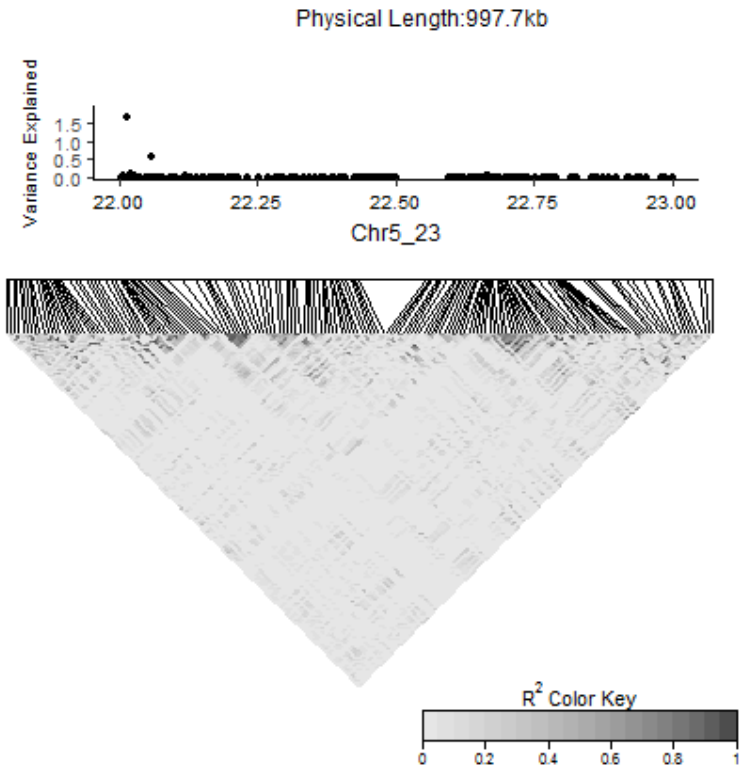

Figure 4q. Manhattan plot of the variance explained by each SNP in the window Chr5_23 for $dEBV_v$ and linkage disequilibrium (LD; r^2)

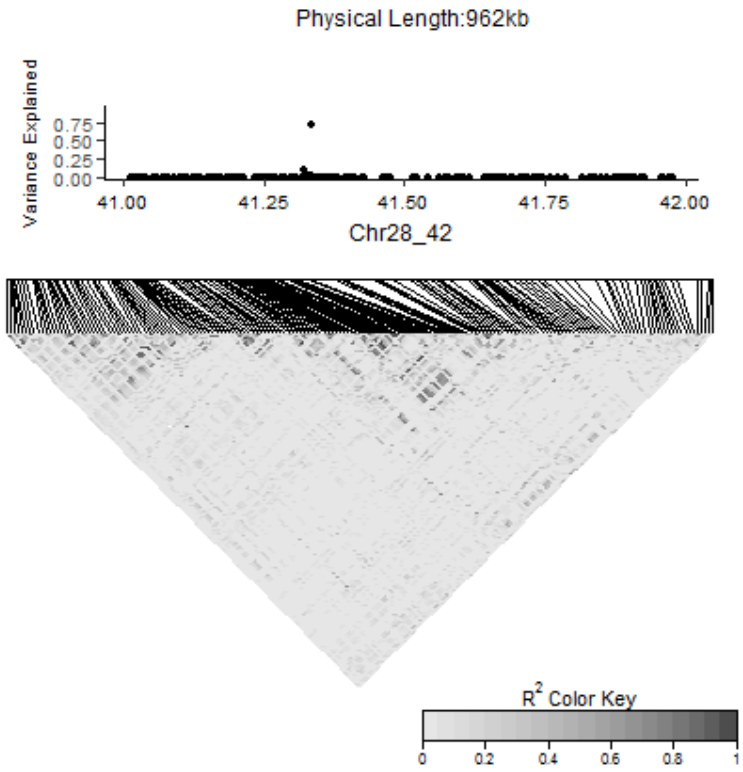

Figure 4r. Manhattan plot of the variance explained by each SNP in the window Chr28_42 for $dEBV_v$ and linkage disequilibrium (LD; r^2)

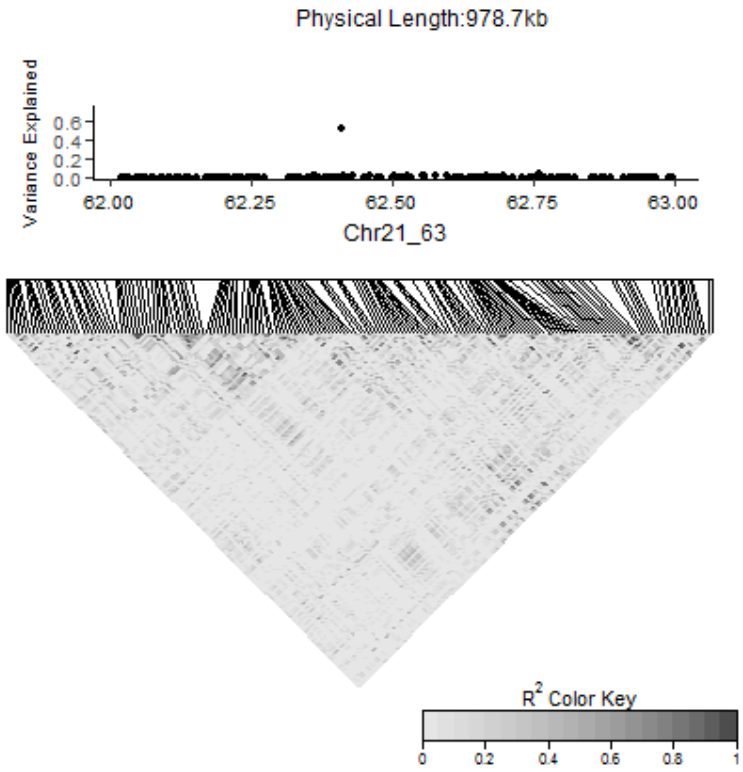

Figure 4s. Manhattan plot of the variance explained by each SNP in the window Chr21_63 for $dEBV_v$ and linkage disequilibrium (LD; r^2)

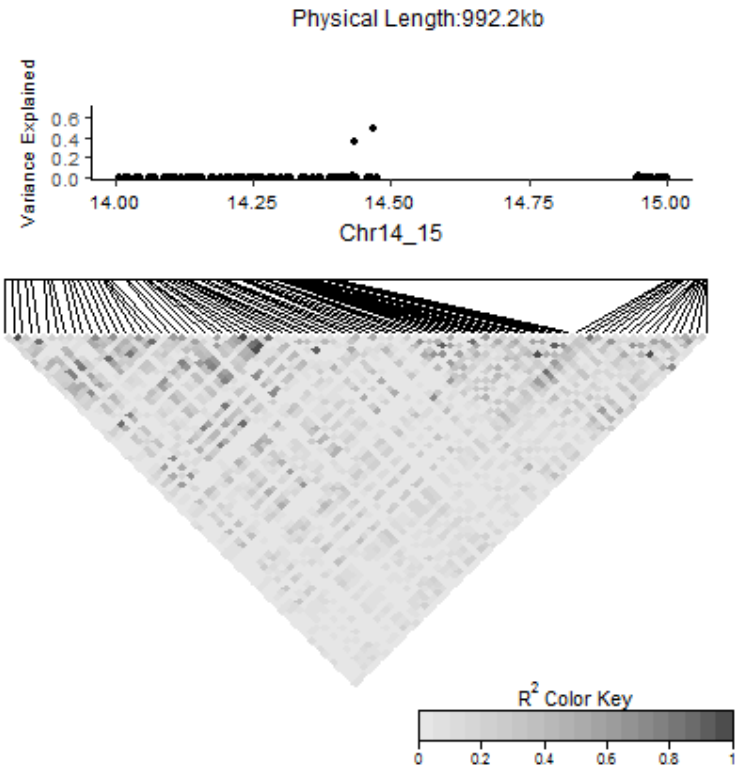

Figure 4t. Manhattan plot of the variance explained by each SNP in the window Chr14_15 for $dEBV_v$ and linkage disequilibrium (LD; r^2)