Heat stress tolerance indicators to be used as phenotypes in GWAS analyses: a comparison study in dairy cattle.

M. Ramón[1], M.J. Carabaño[2] & C. Díaz[2]

[1] CERSYRA-IRIAF. Av. Del Vino 10, 13300, Valdepeñas, Ciudad Real, Spain
[2] Dpto. Mejora Genética Animal, INIA, Ctra. A Coruña Km. 7,5, Spain
cdiaz@inia.es (Corresponding Author)

Summary

A correct assessment of the effects that heat load has on animal performance is of main interest not only for the quantification of the losses associated to it, but also for the correct characterization of individual thermo-tolerance. The latter is key for the identification of genome regions associated with animal thermo-tolerance. In the present study, nine phenotypes have been proposed and their suitability in genome-wide studies has been investigated. Two main approaches have been considered: a first one using pseudo-phenotypes (obtained from the solutions of a mixed model including a norm of reaction to individual heat load term but ignoring pedigree relationships) as phenotypes in the GWAS studies, and a second one using deregressed proofs from those pseudo-phenotypes obtained from a single-step approach. An FDR threshold of 10% was used to select relevant SNPs. Results have shown that many of the markers bellow that threshold were associated with phenotypes representing the animal level of production and located within genome regions comprising production genes like DGAT1. However, 35 out of 95 and 42 out of 78 SNPS for fat and protein, respectively, showed a significant association with phenotypes describing production losses. The dgvGWAS approach identified more SNPs than pseudoGWAS approach for fat content, but the behaviour was different for protein, where pseudoGWAS captured a larger number of SNPs. These two traits appear to have similar heritabilities but different genetic architecture what might be behind this different behaviour.

Keywords: heat stress, gwas, dairy cattle

Introduction

Dairy cows from highly selected populations greatly suffer the consequences of exposure to high heat loads, which result in productive decays, reproductive impairment and immunological depression (West, 2003; Hansen, 2009). A correct assessment of the effects that heat load has on animal performance is of main interest not only for the quantification of the losses associated to it, but also for the correct characterization of individual thermo-tolerance. The latter is key for the identification of genome regions associated with animal thermo-tolerance, for example in genome wide association studies (GWAS). The most common approach to characterise individual thermotolerance at a large scale is through the use of routine milk recording information merged with meteorological information that allow quantification of the thermal load that animals endure (Misztal, 1999). This approach has the advantage of using already available information and targeting economically important traits such as production. However, it also has some associated problems. Firstly, the antagonistic relationship between high production levels and heat tolerance represents a challenge since selection based on genetic values decays in milk production data may lead to keep animals

with lower production level (Bohmanova et al., 2015). Secondly, there is not a unique approach to measure heat tolerance from productive data. The most common approach is to use the slope of decay after a predetermined threshold but other approaches based on the use of random regression models and the eigenvalues decomposition of the associated covariance matrices can help to disentangle the productive and non-productive components of heat tolerance (Carabaño et al., 2014, Macciota et al., 2017). Apart from how to define heat tolerance, what dependent variable is most adequate for GWAS studies has been a matter of discussion (Ekine et al., 2014), as it is the way in which all the available information (phenotypes, pedigree and genomic) is optimally used to provide more accurate marker selection (Zhang et al., 2016).

In this study, we will use fat and protein yield information from the official milk recording scheme and official meteorological agencies to define alternative phenotypes for heat tolerance and SNP marker information from dense panels to develop GWAS under two different approaches.

Material and methods

Phenotypes

In the first part of this study, we addressed the characterization of the individual heat tolerance by assessing the effect that HL has on individual's fat and protein yields and as a previous step for the identification of genome regions associated to heat tolerance. For that, test day records of Holsteins cows under official milk recording programmes were provided by the Spanish Friesian Association (CONAFE) and merged with meteorological information for the dates of test day provided Spanish National Agency for Meteorology (AEMET) and the Irrigation Advisory Service (SIAR). As phenotypes, we considered fat and protein contents (Kg/day), while as measures of HL we used the daily average temperature (T) and the daily average temperature-humidity index (THI; NRC, 1971).

Because cold and heat tolerance seem to be have different genetic background (Carabaño et al 2014) and their joint modelling may be complex, data taken under cold conditions (from December through February) were excluded. Moreover, since an individual´s tolerance was defined from fitting norms of reaction to HL for each cow, lactations that did not include 4 or more test day records and lactations that did not include at least one record taken under heat stress (T_{avg}> 25 ºC or THI_{avg}> 70) were excluded from the analyses to ensure that enough information was present to obtain estimates of declines in production associated to high HL. HS effects on individual production were studied using a Bayesian framework, with the Gibbsf90 program (Misztal et. 2008) and using the model,

(1)

, where y_{ijklm} is the test day fat or protein yield, HTD_i is the herd-test day of recording, $LDIM_j$ is the lactation-days in milk combination, $Z(HL_i)_k$ is the k^{th} covariate of the quadratic Legendre polynomial evaluated at the heat load (HL) corresponding to the i^{th} HTD of recording and transformed to a [-1,1] scale, b_k is the k^{th} regression coefficient of the Legendre polynomial, α_{kl} is the k^{th} regression coefficient of the Legendre polynomial corresponding to animal l and e_{ijklm} is the residual term.

Solutions from the model above were then used to define a total of nine pseudo-phenotypes: the intercept term of the regression for each animal (*bo*), representing the production level, the linear (*b1*) and quadratic (*b2*) regression coefficient estimates for each

animal, genetic slopes of the individual curves (slg) at mild (*slg_18* and *slg_64* for T and THI, respectively), moderate (*slg_22* and *slg_68* for T and THI) and high (*slg_26* and *slg_72* for T and THI) HL, and the scores corresponding to the first (*eigen_1*), second (*eigen_2*) and third (*eigen_3*) eigenvectors obtained from the eigen-decomposition of the (co)variance matrix between regressions for the 3 Legendre regression coefficients for each animal. Slopes at a given HL value were obtained from the derivative of the individual curve at a given HL value, which is a function of the estimated regression coefficients for each animal. These pseudo-phenotypes were used as phenotypes in the GWAS analyses.

Genotypes

Genotypes from 3,165 Holstein bulls obtained with the Illumina BovineSNP50 BeadChip were provided by CONAFE. For the association study, only autosomal markers were considered. Markers with a call rate below 90%, with a MAF < 0.05 or not satisfying the Hardy-Weinberg equilibrium (HEW; p>0.0001) were discarded. After QC, a total of 3,165 and 44,081 markers remain for the association analysis.

Genome-wide association studies (GWAS)

Two different approaches were considered when conducting the GWAS analyses: (i) pseudophenotypes for cows obtained from solutions to (1) were used to obtain, from a weighted average of the values of their daughters, the bulls phenotypes for GWAS (**pseudoGWAS**) (ii) pseudophenotypes of cows were used as dependent variables in a single-step genomic evaluation approach (Legarra et al., 2014), where genomic and pedigree information are mixed to predict the genetic value of animals with data and their relatives. After deregression of the estimated genetic values, the deregressed proofs were used as phenotypes for bulls in the GWAS (**dgvGWAS**). Both EBVs from single-step GBLUP and deregressed proofs were obtained using BLUPf90 family of programmes. GWAS studies were conducted using the GenABEL R package (GenABEL project developers, 2013) using a mixed-effects model that included information on genomic relationships between genotyped bulls to account for possible subpopulation structures and to address the polygenic contribution to each one of the pseudo-phenotypes. For each of the two analyses procedures, trait (fat or protein) and meteorological measure (T and THI), comparisons between markers showing a significant association with each one of the nine pseudo-phenotypes were made. Furthermore, genes containing or next to those markers were gathered using the cow assembly Bos_taurus_UMD_3.1.1/bosTau8.

Results and discussion

Figure 1 shows the matrices of SNPs correspondence between the nine pseudo-phenotypes for each trait (daily fat and protein yields), meteorological (daily average temperature and THI) and GWAS procedure (pseudoGWAS vs. dgvGWAS; see materials and methods). The total number of distinct SNPs that showed FDR below the 10% threshold were 95 for fat and 78 for protein content, regardless the climatic variable and the type of phenotype used. The number of SNPs only involved on the phenotypes related with production level (b_o and *eigen_1)* were 60 and 36, for fat and protein, respectively. DGAT1 was contained in one of the regions for fat. Nevertheless, our major interest was to identify indicators of heat stress independent of production level therefore, we paid attention to slopes of decay and 2nd and 3rd eigenvalues. Hence, we found 35 SNPs for fat and 42 SNPs for protein out of the total. Two

SNPs were common to both, fat and protein indicators. A relevant number of genes are located surrounding these SNPs. An overview, has allowed us finding a number of ribosomal genes in different chromosome with a well-known co-chaperone function. In addition, some genes involved in feed efficiency and appetite have been identified. In general, dgvGWAS approach identified more SNPs than pseudoGWAS approach for fat content, but the behaviour was different for protein, where pseudoGWAS captured a larger number of SNPs bellow the 10% FDR threshold. Initially, these two traits seem to have similar heritability but a different genetic architecture as it is suggested by the major effect of DGAT1 on fat content (Grisart et al., 2022).

Acknowledgments

This work has been funded by S2013/ABI-2913 (MEDGAN, Comunidad Autónoma de Madrid,Spain) and RTA2015-(MINECO, Spain).

List of References

Bohmanova J., I. Misztal, S. Tsuruta, H.D. Norman & T.J. Lawlor, 2005. National genetic evaluations of milk yield for heat tolerance of United States Holsteins. Interbull Bulletin 33: 160–162.

Carabaño M.J., K. Bachagha, M. Ramón & C. Díaz, 2014. Modeling heat stress effect on Holstein cows under hot and dry conditions: selection tools. J. Dairy Sci. 97: 1–16.

Ekine C.C., S. J. Rowe, S. C. Bishop & D. J. de Koning, 2014. Why Breeding Values Estimated Using Familial Data Should Not Be Used for Genome-Wide Association Studies. G3: GENES, GENOMES, GENETICS. 4(2): 341-347

GenABEL project developers, 2013. GenABEL: genome-wide SNP association analysis. R package version 1.8-0. https://CRAN.R-project.org/package=GenABEL

Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M, Snell R, 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res., 12:222–231.

Hansen, P.J., 2009. Effects of heat stress on mammalian reproduction. Phil. Trans. R. Soc. B. 364: 3341–3350.

Legarra A., O. F. Christensen, I. Aguilar & I. Misztal, 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166: 54-65.

Macciotta N.P.P, S. Biffani, U. Bernabucci, N. Lacetera, A. Vitali, P. Ajmone-Marsan, &A. Nardone, 2017. Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. J. Dairy Sci. 100: 4383-4697.

Misztal I., 2008. Reliable computing in estimation of variance components. J. Anim. Breed. Genet. 125: 363-370.

NATIONAL RESEARCH COUNCIL. 1971. A guide to environmental research on animals. National Academy of Science, Washington, DC.

Zhang X, D. Lourenco, I. Aguilar, A. Legarra & I. Misztal, 2016. Weighting strategies for single-step genomic BLUP: An alternative approach for accurate calculation of GEBV and GWAS. Front. Genet. 7: 15

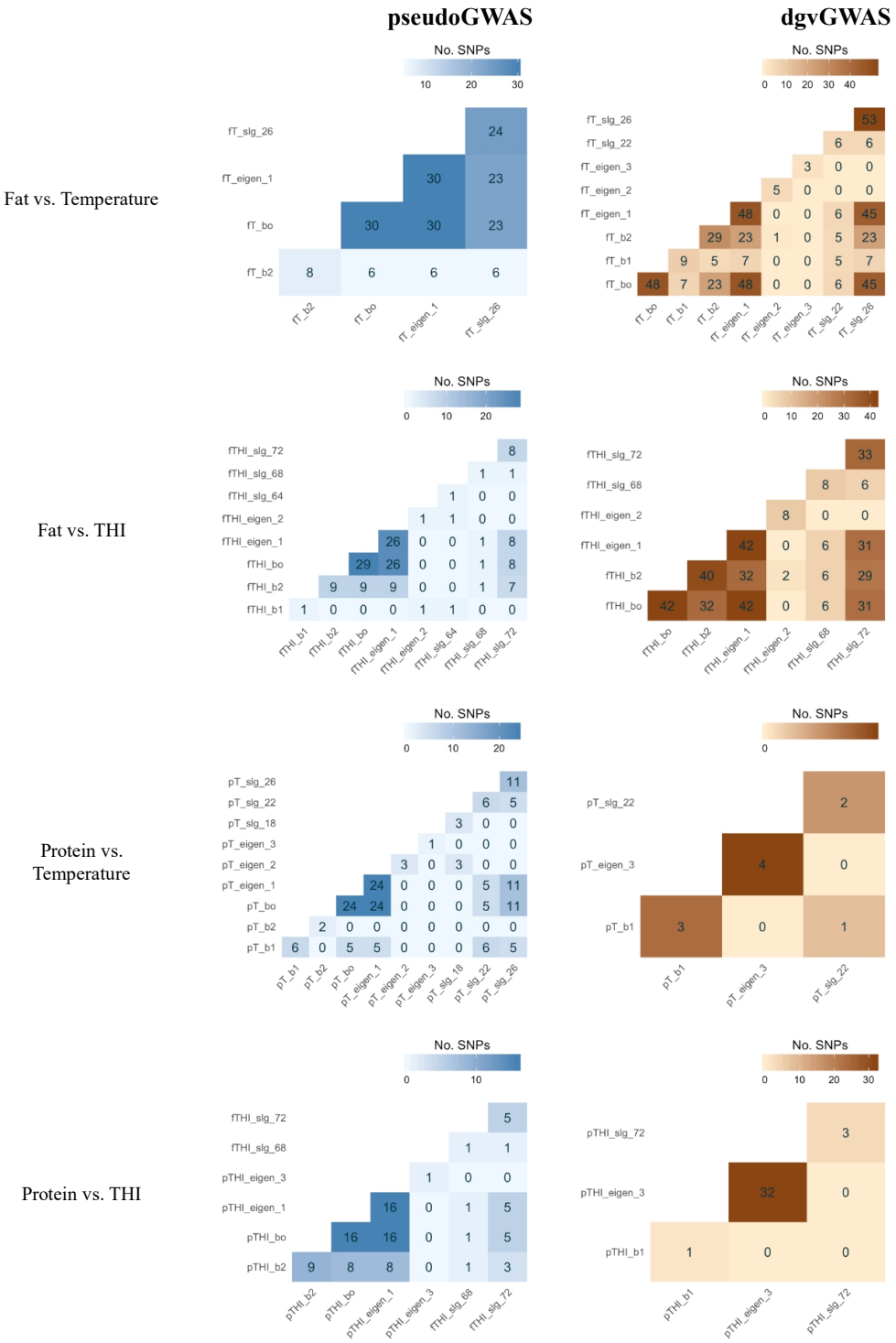

Figure 1. Matrices showing SNPs correspondence between Heat Stress phenotypes for each of the analyses considered (pseudoGWAS and dgvGWAS). Daily fat (f) and protein (p) contents, daily average temperature (T) and THI, and the nine pseudo-phenotypes considered in this study: intercept (bo), linear (b1) and quadratic (b2) coefficients from linear regression, first three eigenvalues (eigen_1, eigen_2 and eigen_3) and genetic slopes at different temperatures (slg_18, slg_22 and slg_26) and THIs (slg_64, slg_68 and slg_72) are represented.