Association analysis of loci implied in "buffering" epistasis

Andrés Legarra[*], Zulma G. Vitezica[*], Marina Naval-Sánchez[**], John Henshall[†], Fernanda Raidan[**], Yutao Li[**], Karin Meyer[‡], Nicholas J. Hudson[§], Laercio R. Porto-Neto[**] and Antonio Reverter[**]

[*]INRA/INPT, UMR 1388 GenPhySE, F-31326 Castanet-Tolosan, France.

[**]CSIRO Agriculture & Food, 306 Carmody Rd., St. Lucia, Brisbane, QLD 4067, Australia.

[†]Cobb-Vantress Inc., Siloam Springs, Arkansas 72761-1030, USA.

[‡]Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia.

[§]School of Agriculture and Food Sciences, The University of Queensland, Gatton, QLD 4343, Australia

18 Running title: Epistasis in association studies

19

20 Key words: Epistasis, genome-wide association studies, beef cattle

21

22 Corresponding Authors: Andres Legarra andres.legarra@inra.fr

23

24

25

ABSTRACT

The existence of buffering mechanisms is an emerging property of biological networks, and this results in the possible existence of "buffering" loci, that would allow buildup of robustness through evolution. So far, there are no explicit methods to find loci implied in buffering mechanisms. However, buffering can be seen as interaction with genetic background. Here we develop this idea into a tractable model for quantitative genetics, in which the buffering effect of one locus with many other loci is condensed into a single (statistical) effect, multiplicative on the total (statistical) additive genetic effect. This allows easier interpretation of the results, and it also simplifies the problem of detecting epistasis from quadratic to linear in the number of loci. Armed with this formulation, we construct a linear model for genome-wide association studies that estimates, and declares significance, of multiplicative epistatic effects at single loci. The model has the form of a variance components, norm reaction model and likelihood ratio tests are used for significance. This model is a generalization and explanation of previous ones. We then test our model using bovine data: Brahman and Tropical Composite animals, phenotyped for body weight at yearling and genotyped up to ~770,000 Single Nucleotide Polymorphisms (SNP). After association analysis and based on False Discovery Rate rules, we find a number of loci with buffering action in one, the other, or both breeds; these loci do not have significant statistical additive effect. Most of these loci have been reported in previous studies, either with an additive effect, or as footprints of selection. We identify epistatic SNPs present in or near genes encoding for proteins that are functionally enriched for peptide activity and transcription factors reported in the context of signatures of selection in multi-breed cattle population studies. These include loci known to be associated with coat color, fertility and adaptation to tropical environments. In these populations we found loci that have a non-significant statistical additive effect but a significant epistatic effect. We argue that the

discovery and study of loci associated with buffering effects allows attacking the difficult problems, among others, of release of maintenance variance in artificial and natural selection, of quick adaptation to the environment, and of opposite signs of marker effects in different backgrounds. We conclude that our method and our results generate promising new perspectives for research in evolutionary and quantitative genetics based on the study of loci that buffer effect of other loci.

INTRODUCTION

Epistasis is a biological phenomenon that is increasingly being given attention in genetics. One of the biological phenomena in which epistasis is likely implied is "buffering" (Visser *et al.* 2003; Flatt 2005), a mechanism that would allow buildup of robustness through evolution (see (Flatt 2005) for examples). A known example is chaperones (Visser *et al.* 2003; Kitano 2004). Loci implied in buffering would mitigate heritable perturbations. For instance, for a trait with intermediate optima, too high total genotypic values would not be expressed. The existence of buffering mechanisms is an emerging propriety of networks (Mackay 2014) and therefore, because biochemical and gene networks are pervasive in nature, buffering loci must exist. Moreover, the existence of segregating (not fixed) buffering epistatic loci would explain several phenomena that are not well understood: environmental robustness, release of additive variance after disturbing events, (Visser *et al.* 2003; Flatt 2005), maintenance of genetic variance in selected populations previously under stabilizing selection (Gimelfarb 1989), and opposite signs of GWAS associations in different populations (Huang *et al.* 2012).

There are so far no explicit methods to detect loci implied in buffering mechanisms. However, buffering can be understood as interaction with genetic background (Visser *et al.*

2003), and methods to detect epistasis against genetic backgrounds have been proposed (Jannink 2007). There is, in addition, increasing evidence of dependency of gene substitution effects in genetic background (Hansen 2013). More recently, a method that implicitly detects loci in interaction with genetic backgrounds has been presented (Crawford *et al.* 2017). However, in neither of these cases the connection with buffering mechanisms has been explicitly shown or put forward.

In this work, we present for the first time, to our knowledge, a formal quantitative genetics framework for the phenomenon of "buffering epistasis", derive efficient methods for genome-wide association studies, and perform association analyses for buffering loci in two real tropical cattle populations. Next, we present the biological discoveries following these analyses, showing that significant hits in the association analysis for buffering epistasis have been reported previously either as having an additive effect or as harboring selection signatures for tropical adaptation.

MATERIAL AND METHODS

Biometrical model.

First, we start our development with a rather general model for epistasis. Then, we show how part of the epistatic variability in this general model can be reformulated as a sum of terms of a buffering action of the "buffering" epistatic locus towards the effect of all other loci, whereas the rest of the epistatic variation is ignored. Then we reformulate the sum of buffering actions in terms of a single buffering effect and overall additive genetic value. Armed with this formulation, we describe two methods of analysis, one exact and one approximate. We generally follow existing notations (Mäki-Tanila and Hill 2014).

Consider the total genotypic effect of one individual as the sum of additive (statistical) effects of all loci (α_j at locus *j*), plus all possible additive by additive (statistical) interactions ($(\alpha\alpha)^{ij}$ for the pairwise interaction between loci *i* and *j*):

$$g = \mu + \sum_j z_j \alpha_j + \sum_i \sum_j z_i z_j (\alpha\alpha)_{ij}$$

Note that these statistical pairwise interactions capture the effect of *functional* pairwise and higher-order interactions (Mäki-Tanila and Hill 2014). Gene content indicators z_i are centered with respect to the population, so $E(g) = \mu$, $E(\sum_i z_i \alpha_i) = 0$ and $E\left(\sum_i \sum_j z_i z_j (\alpha\alpha)_{ij}\right) = 0$ across the population. The model so composed has *n* additive and n^2 additive by additive effects, and they are orthogonal by construction (Cockerham 1954; Álvarez-Castro and Carlborg 2007). Consider Example in Table 1. Allele B is a "shrinker" or "bufferer" allele whereas allele b is a "magnifier" allele.

Table 1. Effect of buffering locus B/b on additive locus c/C

	CC	**Cc**	**cc**
BB	0	0.35	0.70
Bb	0	1.05	2.10
bb	0	3.15	6.30

A buffering effect can be seen as a multiplication of the additive effect of locus *j* (locus C/c in Table 1) by a quantity determined by its interaction with (buffering) locus *i* (locus B/b in Table 1), in other words, $(\alpha\alpha)^{ij} = k_{ij}\alpha_j$. In Table 1, $k_{ij} = 3$. For instance, the

effect of allele "c" in a genetic background "bb" is to increase 3.15 units per one unit of gene content, whereas in a "BB" background the increase is only of 0.35 units.

In our work, we assume that k_{ij} (the buffering effect of locus *i* on locus *j*) can be approximated, for all pairs of locus *i* with other loci *j*, by a locus-specific k_i value. In other words, for a given locus i, the n interactions $(\alpha\alpha)^{ij}$ can be approximated by $(\alpha\alpha)^{ij} \approx k_i\alpha_j$. Alternatively, the value k_i can be seen as the regression coefficient of the equation $(\alpha\alpha)^{ij} = k_i\alpha_j + \epsilon_{ij}$. This simplification allows a reduction in number of parameters (from n^2 interactions to n buffering effects), and, more important, allows focusing on individual buffering loci instead of pairs of loci. In other words, we will find loci which tend to buffer in the same manner across genome.

Thus, for the purpose of detecting buffering epistatic loci, we model the total genotypic value (ignoring remaining epistatic actions) of an individual as

$$g = \mu + \sum_j z_j\alpha_j + \sum_i \sum_j z_i z_j k_i \alpha_j$$

Which after some algebra and because $u = \sum_j z_j\alpha_j$, becomes

$$g = \mu + u + \sum_i z_i k_i u$$

Where u stands for the (statistical) additive genetic value, also known as breeding value in animal genetics. For GWAS purposes, we will use a model considering only the *i*-th loci:

$$g = \mu + u + z_i k_i u \quad (1)$$

Interestingly, the term $z_i k_i u$ can be seen as an interaction of locus *i* with all genome, as described by (Visser *et al.* 2003; Flatt 2005). This model was presented by Jannink (2007) as modelling interaction of locus with all genome, but without stating that it is in fact modelling buffering epistatic action. The same model was presented by Crawford et al.

(2017)) without formalizing the kind of epistatic action, and (in particular) lacking an orthogonal model.

A feature of model in (1) is that because $E(u) = 0$, the observed additive effect of locus *i* changes sign in the extremes of the distribution of u. Imagine for instance that the buffering effect of locus *i* (B/b) is $k_i = -0.2$ and $p = freq(B) = 0.6$. For an individual with $u = 20$ and carrier of BB genotype, $z_i = 2 - 2p = 0.8$, the epistatic effect is negative: $k_i z_i u = -0.2 \times 0.8 \times 20 = -3.2$, and the total genotypic value is $g = 20 - 3.2 = 16.8$. Similarly, for an individual with $u = 0$, the epistatic locus has no effect (there is no buffering); for an individual with $u = -20$, the epistatic effect is positive and *increases* total genotypic value. In all cases, carriers of the "BB" genotype are regressed towards 0.

Plugging (1) into a linear model, a GWAS model would be

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{u} + k_i \boldsymbol{Z}_i^* \boldsymbol{u} + \boldsymbol{e} \quad (2)$$

Where $\boldsymbol{y}$ are quantitative phenotypes of interest, $\boldsymbol{b}$ is the fixed effects vector (e.g. herd-sex-year contemporary group), $\boldsymbol{X}$ is a design matrix relating records to fixed effects, $\boldsymbol{Z}_i^*$ is a matrix whose diagonal contains $\boldsymbol{z}_i$, the coding of the different genotypes at locus *i*. Note that the term $k_i \boldsymbol{Z}_i^* \boldsymbol{u} = k_i \sum_{j=1,n} \boldsymbol{z}_i \odot \boldsymbol{z}_j \alpha_j$ is equivalent to Eq. 1 in Crawford et al. (2017) and their terms α_j are equivalent to our terms $k_i \alpha_j$. However, they do not present their model in terms of buffering, and their matrices x (z in our notation) are not centered, which leads to lack of orthogonality of their model (Álvarez-Castro and Carlborg 2007; Vitezica *et al.* 2017).

Our model in (2), is not usable because both the terms k_i and $\boldsymbol{u}$ implied in the regression are unknown. However, the epistatic component $\boldsymbol{\phi} = k_i \boldsymbol{Z}_i^* \boldsymbol{u}$ defines a covariance matrix for $\boldsymbol{\phi}$ in the *i*-th locus:

$$Cov(\boldsymbol{\phi}) = \boldsymbol{Z}_i^* \boldsymbol{G} \sigma_u^2 k_i^2 \boldsymbol{Z}_i^{*\prime}$$

which suggests a linear model with the form $\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{u} + \boldsymbol{\phi} + \boldsymbol{e}$, with covariance as above (Jannink 2007; Crawford *et al.* 2017). Unfortunately, GWAS tests with this formulation imply computing and inverting $Cov(\boldsymbol{\phi})$ matrix at each locus (which is computing intensive) and can result in lack of convergence (Crawford *et al.*, 2017). We instead propose an equivalent formulation that uses

$$\boldsymbol{g} = \boldsymbol{u} + \boldsymbol{\phi} = \boldsymbol{u} + \boldsymbol{Z}_i^* \boldsymbol{u}^i$$

where $\boldsymbol{u}^i = k_i \boldsymbol{u}$. We have therefore defined a random effect, $\boldsymbol{u}^i$, which multiplies real values given by the covariable $\boldsymbol{Z}_i^*$. This is known as a reaction norm or random regression model (Laird and Ware 1982; Schaeffer 2004). Using this formulation, there are two additive genetic traits in this model: a general additive trait $\boldsymbol{u}$ with variance $\boldsymbol{G}\sigma_u^2$, where G is a relationship matrix (Wright 1922; VanRaden 2008) and a transformation of the buffering action of locus *i* into another additive trait: $\boldsymbol{u}^i = \boldsymbol{u}k_i$, $Var(\boldsymbol{u}^i) = \boldsymbol{G}\sigma_u^2 k_i^2$. The joint covariance matrix is:

$$Var\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{u}^i\end{pmatrix} = \begin{pmatrix}\sigma_u^2 & k_i\sigma_u^2\\ k_i\sigma_u^2 & k_i^2\sigma_u^2\end{pmatrix} \otimes \boldsymbol{G} = \boldsymbol{G}_0 \otimes \boldsymbol{G}$$

where $\begin{pmatrix}\sigma_u^2 & k_i\sigma_u^2\\ k_i\sigma_u^2 & k_i^2\sigma_u^2\end{pmatrix} = \boldsymbol{G}_0$ is a non-full rank matrix because $\boldsymbol{u}^i = \boldsymbol{u}k_i$, and $\otimes$ indicates the Kronecker product. Thus the final linear model, considering the epistatic interaction of locus *i* with all other loci, is

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{u} + \boldsymbol{Z}_i^* \boldsymbol{u}^i + \boldsymbol{e}$$

with $Var\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{u}^i\end{pmatrix} = \begin{pmatrix}\sigma_u^2 & k_i\sigma_u^2\\ k_i\sigma_u^2 & k_i^2\sigma_u^2\end{pmatrix} \otimes \boldsymbol{G} = \boldsymbol{G}_0 \otimes \boldsymbol{G}$. This model can be used in an exact method as described below.

Exact maximum likelihood method

The exact method proceeds by likelihood ratio test of the two-alternative hypothesis, using random regression with the Model H_1:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{z}_i\alpha_i + \boldsymbol{u} + \boldsymbol{Z}_i^*\boldsymbol{u}^i + \boldsymbol{e} \quad (3)$$

and a simpler model excluding random regression with the Model H_0:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{z}_i\alpha_i + \boldsymbol{u} + \boldsymbol{e} \quad (4)$$

Where the actual parameter being tested is $k_i^2\sigma_u^2 \neq 0$. The regression on gene content $\boldsymbol{z}_i\alpha_i$ corrects for eventual statistical additive (not epistatic) effects of locus *i*. Parameters are estimated by REML (Patterson and Thompson 1971).

After fitting the two models, the likelihood ratio test of the competing models is distributed as a mixture of 0 and 1 degrees of freedom chi-square, from which *P*-values can be obtained. In addition, from the estimated covariance matrix $\boldsymbol{G}_0$ estimated under H_1, the estimated buffering epistatic effect can be obtained as:

$$\hat{k}_i = \widehat{\boldsymbol{G}}_0[2,2]/\widehat{\boldsymbol{G}}_0[1,2]. \quad (5)$$

Contrary to Crawford *et al.* (2017) matrix $\boldsymbol{G}$ has to be computed and inverted only once, because inclusion of the i-th locus has zero influence on the result (Gianola *et al.* 2016) and the matrix $\boldsymbol{Z}_i^*\boldsymbol{G}\sigma_u^2\boldsymbol{Z}_i^{*\prime}$ is never explicitly computed. This results in great savings of computing time. Matrix $\boldsymbol{G}_0$ is a non-full rank matrix, which slows down convergence. An easy solution is to use a reduced rank model fitting one principal component (Meyer and Kirkpatrick 2005) as implemented in Wombat (Meyer 2007). Convergence takes a few iterations in this case compared to hundreds using a standard REML algorithm. We conceived two approximate methods that will not be reported here but whose description can be found at (Reverter *et al.* 2018).

201

202 ***Animals, phenotypes and genotypes***

203 Animals, phenotypes and genotypes used in this study were a subset of those used in
204 Raidan *et al.* (2018). In brief, we used data of 2,111 Brahman (BB) and 2,550 Tropical
205 Composite (TC) cows and bulls genotyped using either the BovineSNP50 (Matukumalli *et al.*
206 2009)) or the BovineHD (Illumina Inc., San Diego, CA) that includes more than 770,000
207 SNP. Animals that were genotyped with the lower density array had their genotypes imputed
208 to higher density as described previously by Bolormaa *et al.* (2014). SNPs were mapped to
209 the ARS-UCD1.2 bovine genome assembly. After selecting autosomal SNP with minor allele
210 frequency (MAF) > 1%, we retained 651,253 SNPs for BB and 689,818 for TC. We used
211 body weight at yearling (YWT) as the phenotype of interest. The average, minimum and
212 maximum YWT (kg) were 227.7, 115 and 353 kg for BB; and 247.07, 120.5 and 394.5 kg for
213 TC. Moreover, the average, minimum and maximum age at YWT was 360, 302 and 416 days
214 for BB; and 361, 319 and 403 days for TC.

215

216 ***False discovery rate (FDR)***

217 Following Bolormaa et al. (2014) and with equivalent original derivations from Storey
218 (2002), FDR was calculated as:

$$\text{FDR} = \frac{P\left(1-\frac{A}{T}\right)}{\left(\frac{A}{T}\right)(1-P)}$$

219 Where P is the P-value tested, A is the number of SNP that were significant at the P-
220 value tested, and T is the total number of SNP tested (T = 651,253 and 689,818 for BB and
221 TC, respectively).

222

Implementation

We implemented the analyses using shell scripts to manipulate the data and Wombat (Meyer 2007) for the REML analyses. Markers were analyzed in parallel in the exact analyses in the Genotoul Toulouse bioinfo platform; wall clock computing time was approximately 4 days for the exact analysis run in parallel, and a few minutes for the fast approximate one (see Appendix).

Data Availability Statement

The minimal raw data is the SNP genotype data, phenotypes and metadata (eg. sex, herd, year) for 4,661 cattle and 729,068 SNP genotypes. These raw data are part of the Beef CRC project (http://www.beefcrc.com/) and are co-owned with Meat and Livestock Australia, and can be made available upon reasonable request and subject to the agreement of the owners. Any parties seeking access to the raw data should contact Dr. Antonio Reverter (toni.reverter-gomez@csiro.au or +61732142392).

RESULTS

The exact method took ~4 day in a parallel cluster running ~100 process simultaneously. Computation time for a single marker consist of roughly 1 minute. Note that if a medium density chip had been used (50,000 markers instead of ~600,000), computational times divide by an order of magnitude. There is considerable room for improvement of the computational methods, because most of the time is spent reading and manipulating text files.

RESULTS – Brahman (BB) and Tropical Composite (TC) Populations

Table 2 presents the number of significant SNP and FDR at various *P*-value thresholds. At any given *P*-value, the number of significant SNP was lower in BB than in TC.

As a result, the FDR was lower in TC than in BB for a given *P*-value. For instance, at *P*-value < 0.0001 the FDR was 9.83% and 4.38% for BB and TC, respectively. The higher number of epistatic SNPs identified in the TC compared to the BB population was attributed to the distinct allele frequencies observed in the two populations. Across all SNPs, the first (reference) allele was found to be either mostly absent (reference allele frequency near 0) or nearly fixated (reference allele frequency near 1) in the BB population, while intermediate allele frequencies (highly-polymorphic and hence more informative SNP) were predominant in the TC population (Figure 2 in (Reverter *et al.* 2017)). In other words, the distribution of allele frequencies in BB are U-shaped.

Figure 1 shows the Manhattan plots for the GWAS for additive and epistatic effects in the BB and TC populations. The most likely candidate gene in the most significantly associated regions are also given in Figure 1. It can already be noted that additive and epistatic gene effects are mutually orthogonal.

In the BB population, the strongest significance for epistatic effect corresponded to SNP BovineHD1700017822 mapped to 60,216,894 bp of BTA17 at 11,627 bp of the coding region of *TBX5* (T-box 5 transcription factor) and with an estimated epistatic effect of -0.533 (-Log10(*P*-value) = 6.849). The corresponding human chromosome segment is involved in ulnar mammary syndrome (Klopocki *et al.* 2006), and a recent large meta-GWAS study reveal TBX5 as a candidate gene for mammary gland morphology in Fleckvieh cattle (Pausch *et al.* 2016).

Following *TBX5*, we found the second strongest signal for epistasis in BB to SNP BovineHD2000011094 (estimated epistatic effect of -0.282 and -Log10(*P*-value) = 6.669) mapped to 38.97 Mb of BTA20 and 2.6 kb downstream of prolactin receptor (*PRLR*). *PRLR* is in a region captured by selection signatures for adaptation in beef cattle (Boitard *et al.* 2016)

and mutations on this gene have been found to have a major genetic effect on hair length and coat structure characteristics of cattle (Littlejohn *et al.* 2014; Porto-Neto *et al.* 2018).

The third strongest signal corresponded to SNP BovineHD2300014569 mapped to 50.10 Mb of BTA23 in the coding region of *PRPF4B* (*pre-mRNA processing factor 4B*) with an estimated effect of 0.324 (-Log10(*P*-value) = 6.309). With no reported function in the context of bovine breeding and genetics, *PRPF4B* is an essential kinase induced by estrogen (Lahsaee *et al.* 2016) and its loss promotes sustained growth factor signaling (Corkery *et al.* 2018). Quite strikingly, loci on the coding region of *SPEN* (SNP BovineHD1600014616, epistatic effect = 0.213, -Log10(*P*-value) = 2.137) and *GHR* (SNP BovineHD2000009203, epistatic effect = 0.782, -Log10(*P*-value) = 2.234) were found to be significantly epistatic in our study. *SPEN* is an estrogen receptor cofactor and a key regulator of fat deposition and energy balance (Hazegh *et al.* 2017). Furthermore, a SNP-based co-association gene network by our group previously identified *ESRRG* and *PPARG* as key regulators of age at puberty in Brahman cows (Fortes *et al.* 2013).

In the TC population, we found the strongest signal in SNP BovineHD0100028404 (epistatic effect = 0.294, -Log10(*P*-value) = 9.261) mapped to 98.71 Mb of BTA1 in the coding region of *LOC100139843* (*mCG140927-like*) with limited information known about its function, but quite strikingly, recently reported to be associated with age at puberty in Angus bulls (Fernández *et al.* 2016). We found the second and third strongest signal in the coding region of *ZNF521* (SNP BovineHD2400008618, mapped to BTA24:31,439,030 with and estimated epistatic effect = -0.260, -Log10(*P*-value) = 7.432) and *AGTR1* (SNP BovineHD0100034098 mapped to BTA1: 119,483,491 with and estimated epistatic effect = -0.194, -Log10(*P*-value) = 3.204), respectively. The loci on *ZNF521* has been found to associate with female fertility in Nordic Red cattle, consisting of three different populations from Finland, Sweden and Denmark (Höglund *et al.* 2015). Whereas the role in bovine

297 fertility of *AGTR1* (*angiotensin II receptor type 1*) has long been documented (Portela *et al.*
298 2008; Marey *et al.* 2016) including its differential expression at the level of the oviduct
299 between *Bos taurus* and *Bos indicus* cattle (Fontes *et al.* 2018).

300 The relationship between epistatic and additive effect in each population is illustrated
301 in Figure 2. It can be seen that effects are empirically orthogonal as expected. Note that unlike
302 additive effects, epistatic effects have no units as they are defined as a multiplier on additive
303 effects. Figure 3 shows the relationship between the epistatic effects in both populations, BB
304 and TC. Significant simultaneously in both populations were 42 SNPs of which 24 where
305 located within 50 kb of the coding region of known genes and these are listed in Table 2.
306 Porto-Neto et al. (2014) showed that LD dropped below 0.2 at distances of 50 kb.

307 Among those listed in Table2, prominent genes for their reported role in mammalian
308 fertility including bovine are: *ALDH1A1 (aldehyde dehydrogenase 1 family member A1),*
309 *PCSK5 (proprotein convertase subtilisin/kexin type 5),* and *TSHR (thyroid stimulating*
310 *hormone receptor), and IL21 (Interleukin-21).*

311 The role of *ALDH1A1* during bovine ovarian development has recently been
312 stablished (Hatzirodos *et al.* 2019; Hummitzsch *et al.* 2019). Antenos et al. (2011) reported
313 the role of PCSK5 in mouse ovarian follicle development. Similarly, *TSHR* is a well-known
314 gene for its function regulating growth, fat metabolism and fertility. Dias et al. (2017)
315 identified a candidate QTL in *TSHR* affecting puberty in five cattle breeds across the taurine
316 and Indicine lineages: Brangus, Brahman, Nellore, Angus and Holstein. Also, one of the most
317 prominent selective sweeps found in all domestic chickens occurred at the locus for *TSHR*
318 (Rubin *et al.* 2010). Finally, the immune system response gene *IL21* has been shown to harbor
319 selection signatures among divergently selected subpopulations of Polish Red cattle (Gurgul
320 *et al.* 2019), and among goats and sheep indigenous to a hot arid environment (Kim *et al.*
321 2016).

DISCUSSION

In this study, we present a rigorous formulation of buffering epistasis due to single loci and how it relates to previous works (Jannink 2007; Crawford *et al.* 2017). Then we present an exact test for association analysis of single locus in buffering epistatic interaction based on bi-variate random regression restricted maximum likelihood (REML) analyses.

Methodology for detection of locus by genetic background interactions has been developed at least two times, (sadly) in apparent isolation from each other (Jannink 2007; Crawford *et al.* 2017). Jannink (2007) first and formally derived the linear model involving (additive) genetic background and one locus showing interaction with the background. He showed orthogonality and derived a variance component useful for genome-wide association studies (GWAS), but he did not realize that the model discovered a particular kind of epistasis. Later, the same model was re-derived by Crawford et al. (2017) who, however, did not build an orthogonal model (something that may lead to spurious effect estimates), and could not use maximum likelihood estimates. Neither of the two authors explicitly estimated the effect of the epistatic loci. Here, we complete and interpret their work.

Then we applied our method to two datasets in cattle. Taking together, our findings in these datasets are striking. Some of our significant genes have been reported as having a statistical *additive* effect in other studies, whereas in ours, they have a statistical *epistatic* effect but not an additive one. As shown by theory, a shift in the mean of the genetic values towards one of the extremes changes the statistical effect from epistatic into additive, whereas the functional effect is always epistatic. For instance, a gene with epistatic buffering effect $k_i = 0.1$ in a population with $\bar{u} = 0$ will have a statistically null additive effect, whereas the

same gene in a population with $\bar{u} > 0$ will have a positive substitution effect, and a negative one in a population with $\bar{u} < 0$. This explains indeed why some effects change sign when observed in different genetic backgrounds (Magwire *et al.* 2010).

A similar argument explains why we find genes found under selection in a multi-breed comparison study. These genes are bound to have little variation in their coding region and/or no additive effect in any given individual breed, because they are fixed or near fixation. However, inside our populations TC and BB, they are identified as relevant because they have opposite effects in the extremes of the polygenic background, and they would get alternative fixation of alleles if the populations were selected towards either extreme. In other words, our epistatic association analysis can be seen as a selection signature analysis *before* selection of the extremes.

It is somehow difficult to grasp the meaning of the estimate of buffering epistatic effects, *e.g.* $\hat{k}_i = 0.53$ (adimensional) in TBX5 in the BB population. For instance, genetic variation for human height is approximately 39 cm^2 in European populations (Visscher 2008). Imagine an individual with large total additive genetic value, *i.e.* 4 standard deviations (25 cm taller than the mean, *i.e.* 190 cm for a woman and 203 cm for a man). If such an individual is carrier of the "bufferer" copies *and* the allele has a frequency of 0.8 (we will assume that the bufferer allele is fairly frequent as protective), her/his height changes by $z_i k_i u = (2 - 2 \times 0.8) \times -0.53 \times 25$, i.e. 5.3 centimeters shorter. Still, the estimate of -0.53 is probably too high – due to the winner's curse or Beavis effect (Xu 2003).

The possibility of detection and further functional analysis of these epistatic loci implied in buffering mechanisms is of particular relevance in quantitative genetics. As described in the Introduction, these loci must exist (Mackay 2014), be selected (Flatt 2005) and they would explain perplexing phenomena: release or maintenance of additive variance

(Visser *et al.* 2003; Flatt 2005), (Gimelfarb 1989), "conversion" of epistatic into additive variance (Mackay 2014) and opposite signs of GWAS associations in different populations (Huang *et al.* 2012) (Magwire *et al.* 2010).

Our model is based on interaction with genetic background, and there is increasing evidence of dependency of gene substitution on genetic background (Hansen 2013). This implies, potentially, large changes in gene effects and selection dynamics even if genetic variance is nearly fully additive at each step (Hansen 2013). Indeed, it has been argued that epistasis provides the basis of rapid adaptation to new environments (Wright 1931; Mackay 2014).

For evolutionary research and plant and animal breeding, there is growing interest in understanding the extent and mechanisms of epistasis in biology because of the intriguing prospect of it being an untapped future source of additive variation that may be exploited by nature and by breeding programs to evolve phenotypes as well as influencing genetic heterogeneity. Indeed, the role of epistasis in determining which mutations ultimately succeed or fail in a population under selection remains a central challenge in biology. Buskirk et al. (2017) demonstrated the power of experimental evolution to identify epistatic interactions. In multi-generation selection programs continued response has been seen, classically for over 100 generations in the Illinois maize kernel content lines , and there have been large and still continuing genetic improvements in livestock populations, notably in broiler chickens (Hill 2016). Whilst the obvious source of continued response is *de novo* mutation, some of the additive variation being utilized may have derived from existing mutations whose behavior changes from epistatic to additive in response to changes in the remainder of the genome, consistent with the argument of Carlborg et al. (2006). Paixão and Barton (2016) argued that 'epistasis sustains additive genetic variance for longer: Alleles that were initially deleterious or near-neutral may acquire favorable effects as the genetic background changes,

395 "converting" epistatic variance into additive, and so prolonging the response to selection'.
396 Similarly, Hill (2017) concluded that 'It seems better to concentrate on utilizing additive
397 variance, and hope for a bonus from converting epistatic variance'.

398 To conclude, our method and our results show promising new perspectives for
399 research in evolutionary and quantitative genetics based on the study of loci that buffer effect
400 of other loci.

401

402 **ACKNOWLEDGMENTS**

403 Excellent advice and review was provided by Sonja Dominik and Andrew George.
404 Work supported by INRA SELGEN metaprogram (projects EpiSel and OptiMaGics) and
405 INRA-CSIRO Linkage Action and SMARTER. Project partly supported by Toulouse Midi-
406 Pyrénées bioinformatics platform.

407

408 **AUTHORS' CONTRIBUTIONS**

409 AR, JH, KM, ZV and AL derived the equations, wrote the programs to do the
410 analyses, performed analyses, and assisted drafting the manuscript. L P-N and M N-S
411 provided sequence-level data, and designed and performed the functional analyses. FR, YL
412 and NH provided valuable insights throughout the analysis and writing process.

413

414

415 **APPENDIX**

416 ***Technical details for the REML analyses***

417 Because the covariance matrix in the alternative model is not full rank,

418 $\begin{pmatrix} \sigma_u^2 & k_i\sigma_u^2 \\ k_i\sigma_u^2 & k_i^2\sigma_u^2 \end{pmatrix} = \boldsymbol{G}_0$, the reduced rank method of (Meyer and Kirkpatrick 2005) was used

(explicitly stating rank equal to 1), and the iterative algorithm was a mixture of PX and AI (we refer the reader to the Wombat manual for details). Unlike other software programs, Wombat does not accept centered genotypes as covariates for the random regression model – only integers $m = \{1,2,3\}$ for each genotype that are transformed into reals (centered and scaled) internally. Thus, although the likelihood is in the correct scale, the estimated $\widehat{\boldsymbol{G}_0}$ on output is not. In a model with random regressions, $\boldsymbol{y} = \cdots + \boldsymbol{Pu} + \cdots$ the covariance across two points is a function of $\boldsymbol{PKP'}$ where $\boldsymbol{K} = Var(\boldsymbol{u})$ is a matrix of covariances and $\boldsymbol{P}$ are covariates. Accordingly, to put back $\widehat{\boldsymbol{G}_0}$ in the right scale we need to construct, for each locus i, $\boldsymbol{P} = \begin{bmatrix} 1 & \bar{m}_i \\ 0 & 1 \end{bmatrix}$ where $\bar{m}_i = 1 + 2p_i$ is the mean of the regressors $m = \{1,2,3\}$. Back transformation to the regular scale is carried out using $\widehat{\boldsymbol{G}_0} \Leftarrow \boldsymbol{P}\widehat{\boldsymbol{G}_0}\boldsymbol{P'}$ and from here the correct value $\widehat{k}_i = \widehat{\boldsymbol{G}}_0[2,2]/\widehat{\boldsymbol{G}}_0[1,2]$.

REFERENCES

Álvarez-Castro J. M., and Ö. Carlborg, 2007 A Unified Model for Functional and Statistical Epistasis and Its Application in Quantitative Trait Loci Analysis. Genetics 176: 1151–1167. https://doi.org/10.1534/genetics.106.067348

Antenos M., L. Lei, M. Xu, A. Malipatil, S. Kiesewetter, *et al.*, 2011 Role of PCSK5 expression in mouse ovarian follicle development: identification of the inhibin α- and β-subunits as candidate substrates. PLoS ONE 6: e17348. https://doi.org/10.1371/journal.pone.0017348

Boitard S., M. Boussaha, A. Capitan, D. Rocha, and B. Servin, 2016 Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. Genetics 203: 433–450. https://doi.org/10.1534/genetics.115.181594

443 Bolormaa S., J. E. Pryce, A. Reverter, Y. Zhang, W. Barendse, *et al.*, 2014 A Multi-Trait, Meta-analysis
444 for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef
445 Cattle. PLOS Genetics 10: e1004198. https://doi.org/10.1371/journal.pgen.1004198

446 Buskirk S. W., R. E. Peace, and G. I. Lang, 2017 Hitchhiking and epistasis give rise to cohort dynamics
447 in adapting populations. PNAS 114: 8330–8335. https://doi.org/10.1073/pnas.1702314114

448 Carlborg Ö., L. Jacobsson, P. Åhgren, P. Siegel, and L. Andersson, 2006 Epistasis and the release of
449 genetic variation during long-term selection. Nature Genetics 38: 418–420.
450 https://doi.org/10.1038/ng1761

451 Cockerham C. C., 1954 An extension of the concept of partitioning hereditary variance for analysis of
452 covariances among relatives when epistasis is present. Genetics 39: 859.

453 Corkery D. P., L. E. Clarke, S. Gebremeskel, J. Salsman, J. Pinder, *et al.*, 2018 Loss of PRP4K drives
454 anoikis resistance in part by dysregulation of epidermal growth factor receptor endosomal
455 trafficking. Oncogene 37: 174–184. https://doi.org/10.1038/onc.2017.318

456 Crawford L., P. Zeng, S. Mukherjee, and X. Zhou, 2017 Detecting epistasis with the marginal epistasis
457 test in genetic mapping studies of quantitative traits. PLOS Genetics 13: e1006869.
458 https://doi.org/10.1371/journal.pgen.1006869

459 Dias M. M., A. Cánovas, C. Mantilla-Rojas, D. G. Riley, P. Luna-Nevarez, *et al.*, 2017 SNP detection
460 using RNA-sequences of candidate genes associated with puberty in cattle. Genet. Mol. Res.
461 16. https://doi.org/10.4238/gmr16019522

462 Fernández M. E., A. Prando, A. Rogberg-Muñoz, P. Peral-García, A. Baldo, *et al.*, 2016 Association of a
463 region of bovine chromosome 1 (BTA1) with age at puberty in Angus bulls. Reprod. Fertil.
464 Dev. 28: 1618–1621. https://doi.org/10.1071/RD14511

Flatt T., 2005 The Evolutionary Genetics of Canalization. The Quarterly Review of Biology 80: 287–316. https://doi.org/10.1086/432265

Fontes P. K., R. L. Ereno, A. R. Peixoto, R. F. Carvalho, W. R. Scarano, *et al.*, 2018 Can the antral follicular count modulate the gene expression of bovine oviducts in Aberdeen Angus and Nelore heifers? PLoS ONE 13: e0202017. https://doi.org/10.1371/journal.pone.0202017

Fortes M. R. S., K. Kemper, S. Sasazaki, A. Reverter, J. E. Pryce, *et al.*, 2013 Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. Animal Genetics 44: 636–647. https://doi.org/10.1111/age.12075

Gianola D., M. I. Fariello, H. Naya, and C.-C. Schön, 2016 Genome-Wide Association Studies with a Genomic Relationship Matrix: A Case Study with Wheat and Arabidopsis. G3: Genes, Genomes, Genetics 6: 3241–3256. https://doi.org/10.1534/g3.116.034256

Gimelfarb A., 1989 Genotypic variation for a quantitative character maintained under stabilizing selection without mutations: epistasis. Genetics 123: 217–227.

Gurgul A., I. Jasielczuk, E. Semik-Gurgul, T. Szmatoła, A. Majewska, *et al.*, 2019 Diversifying selection signatures among divergently selected subpopulations of Polish Red cattle. J. Appl. Genet. 60: 87–95. https://doi.org/10.1007/s13353-019-00484-0

Hansen T. F., 2013 Why Epistasis Is Important for Selection and Adaptation. Evolution 67: 3501–3511. https://doi.org/10.1111/evo.12214

Hatzirodos N., K. Hummitzsch, H. F. Irving-Rodgers, J. Breen, V. E. A. Perry, *et al.*, 2019 Transcript abundance of stromal and thecal cell related genes during bovine ovarian development. PLoS ONE 14: e0213575. https://doi.org/10.1371/journal.pone.0213575

Hazegh K. E., T. Nemkov, A. D'Alessandro, J. D. Diller, J. Monks, *et al.*, 2017 An autonomous metabolic role for Spen. PLOS Genetics 13: e1006859. https://doi.org/10.1371/journal.pgen.1006859

Hill W. G., 2016 Is Continued Genetic Improvement of Livestock Sustainable? Genetics 202: 877–881. https://doi.org/10.1534/genetics.115.186650

Hill W. G., 2017 "Conversion" of epistatic into additive genetic variance in finite populations and possible impact on long-term selection response. Journal of Animal Breeding and Genetics 134: 196–201. https://doi.org/10.1111/jbg.12270

Höglund J. K., B. Buitenhuis, B. Guldbrandtsen, M. S. Lund, and G. Sahana, 2015 Genome-wide association study for female fertility in Nordic Red cattle. BMC Genetics 16: 110. https://doi.org/10.1186/s12863-015-0269-x

Huang W., S. Richards, M. A. Carbone, D. Zhu, R. R. H. Anholt, *et al.*, 2012 Epistasis dominates the genetic architecture of Drosophila quantitative traits. PNAS 109: 15553–15559. https://doi.org/10.1073/pnas.1213423109

Hummitzsch K., N. Hatzirodos, H. F. Irving-Rodgers, M. D. Hartanti, V. E. A. Perry, *et al.*, 2019 Morphometric analyses and gene expression related to germ cells, gonadal ridge epithelial-like cells and granulosa cells during development of the bovine fetal ovary. PLoS ONE 14: e0214130. https://doi.org/10.1371/journal.pone.0214130

Jannink J.-L., 2007 Identifying Quantitative Trait Locus by Genetic Background Interactions in Association Studies. Genetics 176: 553–561. https://doi.org/10.1534/genetics.106.062992

Kim E.-S., A. R. Elbeltagy, A. M. Aboul-Naga, B. Rischkowsky, B. Sayre, *et al.*, 2016 Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity (Edinb) 116: 255–264. https://doi.org/10.1038/hdy.2015.94

508 Kitano H., 2004 Biological robustness. Nature Reviews Genetics 5: 826–837.
509 https://doi.org/10.1038/nrg1471

510 Klopocki E., L. M. Neumann, H. Tönnies, H.-H. Ropers, S. Mundlos, *et al.*, 2006 Ulnar-mammary
511 syndrome with dysmorphic facies and mental retardation caused by a novel 1.28 Mb
512 deletion encompassing the TBX3 gene. Eur. J. Hum. Genet. 14: 1274–1279.
513 https://doi.org/10.1038/sj.ejhg.5201696

514 Lahsaee S., D. P. Corkery, L. E. Anthes, A. Holly, and G. Dellaire, 2016 Estrogen receptor alpha (ESR1)-
515 signaling regulates the expression of the taxane-response biomarker PRP4K. Experimental
516 Cell Research 340: 125–131. https://doi.org/10.1016/j.yexcr.2015.12.013

517 Laird N. M., and J. H. Ware, 1982 Random-effects models for longitudinal data. Biometrics 38: 963–
518 974.

519 Littlejohn M. D., K. M. Henty, K. Tiplady, T. Johnson, C. Harland, *et al.*, 2014 Functionally reciprocal
520 mutations of the prolactin signalling pathway define hairy and slick cattle. Nature
521 Communications 5: 5861. https://doi.org/10.1038/ncomms6861

522 Mackay T. F. C., 2014 Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-Gene
523 Interactions. Nat Rev Genet 15: 22–33. https://doi.org/10.1038/nrg3627

524 Magwire M. M., A. Yamamoto, M. A. Carbone, N. V. Roshina, A. V. Symonenko, *et al.*, 2010
525 Quantitative and Molecular Genetic Analyses of Mutations Increasing Drosophila Life Span.
526 PLOS Genetics 6: e1001037. https://doi.org/10.1371/journal.pgen.1001037

527 Mäki-Tanila A., and W. G. Hill, 2014 Influence of Gene Interaction on Complex Trait Variation with
528 Multilocus Models. Genetics 198: 355–367. https://doi.org/10.1534/genetics.114.165282

529 Marey M. A., M. S. Yousef, J. Liu, K. Morita, M. Sasaki, *et al.*, 2016 Angiotensin II increases sperm
530 phagocytosis by neutrophils in vitro: A possible physiological role in the bovine oviduct. Mol.
531 Reprod. Dev. 83: 630–639. https://doi.org/10.1002/mrd.22672

532 Matukumalli L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, *et al.*, 2009 Development and
533 characterization of a high density SNP genotyping assay for cattle. PLoS ONE 4: e5350.
534 https://doi.org/10.1371/journal.pone.0005350

535 Meyer K., and M. Kirkpatrick, 2005 Restricted maximum likelihood estimation of genetic principal
536 components and smoothed covariance matrices. Genet. Sel. Evol. 37: 1–30.
537 https://doi.org/10.1051/gse:2004034

538 Meyer K., 2007 WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted
539 maximum likelihood (REML). J Zhejiang Univ Sci B 8: 815–821.
540 https://doi.org/10.1631/jzus.2007.B0815

541 Paixão T., and N. H. Barton, 2016 The effect of gene interactions on the long-term response to
542 selection. Proc. Natl. Acad. Sci. U.S.A. 113: 4422–4427.
543 https://doi.org/10.1073/pnas.1518830113

544 Patterson H., and R. Thompson, 1971 Recovery of inter-block information when block sizes are
545 unequal. Biometrika 58: 545.

546 Pausch H., R. Emmerling, H. Schwarzenbacher, and R. Fries, 2016 A multi-trait meta-analysis with
547 imputed sequence variants reveals twelve QTL for mammary gland morphology in Fleckvieh
548 cattle. Genet. Sel. Evol. 48: 14. https://doi.org/10.1186/s12711-016-0190-4

549 Portela V. M., P. B. D. Gonçalves, A. M. Veiga, E. Nicola, J. Buratini, *et al.*, 2008 Regulation of
550 angiotensin type 2 receptor in bovine granulosa cells. Endocrinology 149: 5004–5011.
551 https://doi.org/10.1210/en.2007-1767

552 Porto-Neto L. R., J. W. Kijas, and A. Reverter, 2014 The extent of linkage disequilibrium in beef cattle
553 breeds using high-density SNP genotypes. Genet. Sel. Evol. 46: 22.
554 https://doi.org/10.1186/1297-9686-46-22

555 Porto-Neto L. R., D. M. Bickhart, A. J. Landaeta-Hernandez, Y. T. Utsunomiya, M. Pagan, *et al.*, 2018
556 Convergent Evolution of Slick Coat in Cattle through Truncation Mutations in the Prolactin
557 Receptor. Front. Genet. 9. https://doi.org/10.3389/fgene.2018.00057

558 Raidan F. S. S., L. R. Porto-Neto, Y. Li, S. A. Lehnert, and A. Reverter, 2018 Weighting genomic and
559 genealogical information for genetic parameter estimation and breeding value prediction in
560 tropical beef cattle. J Anim Sci 96: 612–617. https://doi.org/10.1093/jas/skx027

561 Reverter A., L. R. Porto-Neto, M. R. S. Fortes, P. Kasarapu, M. a. R. de Cara, *et al.*, 2017 Genomic
562 inbreeding depression for climatic adaptation of tropical beef cattle. J. Anim. Sci. 95: 3809–
563 3821. https://doi.org/10.2527/jas2017.1643

564 Reverter A., J. Henshall, L. Porto-Neto, F. Raidan, Y. Li, *et al.*, 2018 A rapid method for the
565 identification of epistatic 'dormant' SNPs. Proceedings of the World Congress on Genetics
566 Applied to Livestock Production Methods and Tools-GWAS: 260.

567 Rubin C.-J., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood, *et al.*, 2010 Whole-genome
568 resequencing reveals loci under selection during chicken domestication. Nature 464: 587–
569 591. https://doi.org/10.1038/nature08832

570 Schaeffer L. R., 2004 Application of random regression models in animal breeding. Livestock
571 Production Science 86: 35–45. https://doi.org/10.1016/S0301-6226(03)00151-9

572 Storey J. D., 2002 A direct approach to false discovery rates. Journal of the Royal Statistical Society:
573 Series B (Statistical Methodology) 64: 479–498. https://doi.org/10.1111/1467-9868.00346

574 VanRaden P. M., 2008 Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91: 4414–
575 4423.

576 Visscher P. M., 2008 Sizing up human height variation. Nature Genetics 40: 489–490.
577 https://doi.org/10.1038/ng0508-489

578 Visser J. A. G. M. de, J. Hermisson, G. P. Wagner, L. A. Meyers, H. Bagheri-Chaichian, *et al.*, 2003
579 Perspective: Evolution and Detection of Genetic Robustness. Evolution 57: 1959–1972.
580 https://doi.org/10.1111/j.0014-3820.2003.tb00377.x

581 Vitezica Z. G., A. Legarra, M. A. Toro, and L. Varona, 2017 Orthogonal Estimates of Variances for
582 Additive, Dominance and Epistatic Effects in Populations. Genetics genetics.116.199406.
583 https://doi.org/10.1534/genetics.116.199406

584 Wright S., 1922 Coefficients of inbreeding and relationship. Am. Nat. 56: 330–338.

585 Wright S., 1931 Evolution in Mendelian Populations. Genetics 16: 97–159.

586 Xu S., 2003 Theoretical Basis of the Beavis Effect. Genetics 165: 2259–2268.

587

588

589

590 **TABLES**

591

592 **Table 2**. Number of significant epistatic SNP (N) and false discovery rate (FDR) at

593 decreasing levels of *P*-value for the Brahman and Tropical Composite populations.

***P*-value**	**Brahman**		**Tropical Composite**	
	N	**FDR, %**	**N**	**FDR, %**
< 0.05	66,789	46.06	87,545	36.21
< 0.01	20,542	31.01	30,803	21.61
< 0.005	12,503	25.67	19,831	16.98
< 0.001	3,679	17.62	6,972	9.80
<0.0005	2,186	14.85	4,484	7.64
<0.0001	662	9.83	1,571	4.38
<0.00005	384	8.47	1,028	3.35
<0.00001	91	7.15	342	2.02

594

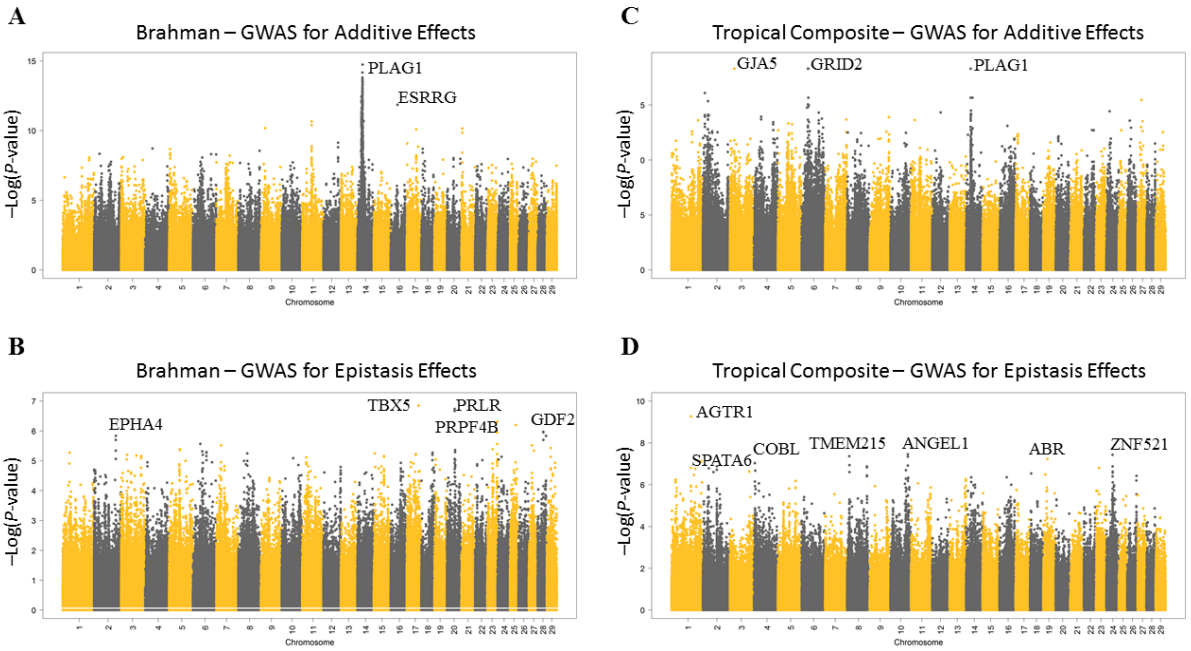
Table 3. Annotation of SNP with significant (P<0.001) epistatic effect in both populations (Brahman and Tropical Composite) including genome position, distance to nearest gene, gene, reference allele frequency (RAF), estimated effect and significance (-Log(P)).

SNP	Chr	Bp	Distance	Gene	Brahman			Tropical Composite		
					RAF	Effect	-Log(P)	RAF	Effect	-Log(P)
BovineHD0200001651	2	5,747,611	38,487	NEMP2	0.0263	-0.5406	3.0404	0.0729	0.3258	3.5760
BovineHD0200001652	2	5,748,515	37,583	NEMP2	0.0263	-0.5406	3.0404	0.0727	0.3237	3.5065
BovineHD0200001654	2	5,757,112	28,986	NEMP2	0.9737	0.5433	3.0404	0.9272	-0.3234	3.5297
ARS-BFGL-NGS-100083	8	11,133,497	0	SCARA3	0.9528	-0.5113	3.4369	0.6920	0.1975	4.1983
BovineHD0800014800	8	49,143,207	36,501	ALDH1A1	0.9820	0.7474	3.5992	0.7831	0.2002	3.9915
BovineHD0800015819	8	52,427,816	8,480	PCSK5	0.0452	-0.4416	3.3439	0.2349	-0.1963	3.6686
Hapmap48089-BTA-81187	8	52,524,895	0	GCNT1	0.9403	0.3760	3.0638	0.6829	0.1562	3.2041
BovineHD1000026980	10	92,312,649	0	TSHR	0.6054	0.2160	3.5761	0.4686	-0.1712	4.0605
BovineHD1000026985	10	92,331,573	0	TSHR	0.3946	-0.2140	3.5761	0.5317	0.1779	4.1064
BovineHD1000030828	10	92,335,071	0	TSHR	0.4216	-0.2017	3.2507	0.7002	0.1623	3.2973
BovineHD1200025989	12	84,830,432	15,785	5S_rRNA	0.7141	0.2066	3.2041	0.6998	-0.1603	3.0872
BovineHD1400011237	14	37,055,691	0	ENSBTAG00000002076	0.1033	-0.3042	3.0404	0.5288	-0.1442	3.0872
BovineHD1400011280	14	37,211,716	0	ENSBTAG00000050514	0.9095	0.3733	3.9915	0.6543	0.2204	5.9710
BovineHD1400013234	14	44,638,868	5,792	PMP2	0.6080	0.1951	3.0169	0.6512	-0.1791	3.9225
BovineHD1600016214	16	56,665,443	0	TNR	0.1409	0.2618	3.0872	0.6231	-0.1635	3.6917
BovineHD1700009805	17	35,070,459	0	IL21	0.9377	0.4452	3.7841	0.7129	-0.1633	3.2973
BovineHD1700009807	17	35,079,284	8,519	IL21	0.0623	-0.4429	3.7610	0.2859	0.1659	3.2274
BovineHD1700009808	17	35,083,771	13,006	IL21	0.0623	-0.4429	3.7610	0.2859	0.1659	3.2274
BovineHD1700009811	17	35,094,499	23,734	IL21	0.9377	0.4452	3.7841	0.7131	-0.1626	3.2740
BovineHD1900012991	19	45,742,542	18,774	ENSBTAG00000053203	0.2399	0.2629	3.5297	0.3921	-0.1908	4.5416
BovineHD4100015365	21	65,988,493	33,080	bta-mir-656	0.0559	0.4343	3.1340	0.0502	-0.4564	5.9259
BovineHD2500009619	25	34,418,988	0	SSC4D	0.1843	0.3028	3.6455	0.4378	-0.1702	3.8764
BovineHD2500009621	25	34,427,964	0	SSC4D	0.1831	0.3168	3.8764	0.3959	-0.2030	4.9974
BovineHD2800005959	28	22,571,976	0	CTNNA3	0.1338	0.3147	3.1807	0.3500	0.1516	3.2973

599

600 # FIGURES

601

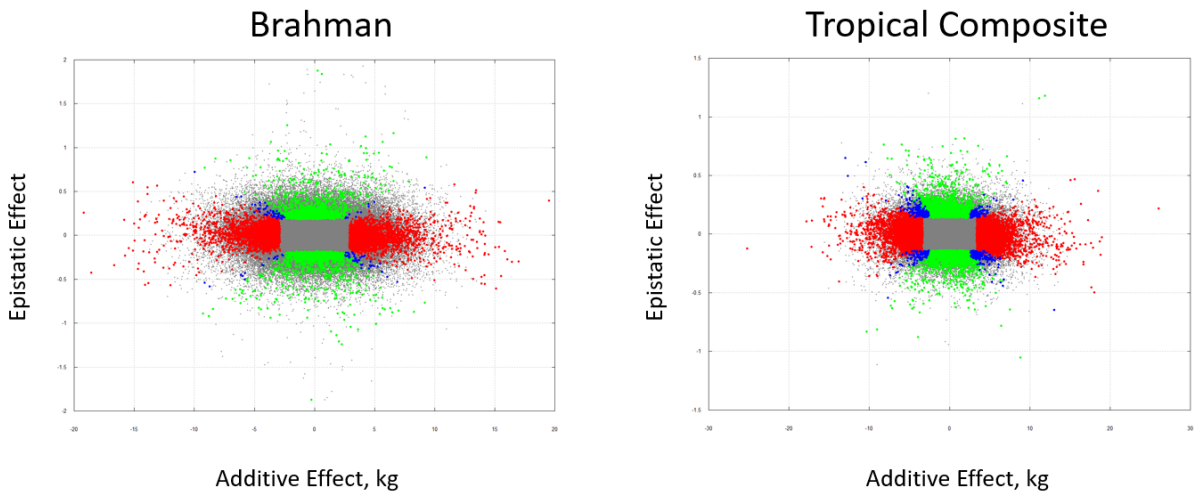


602

603 **Figure 1**. Genome-wide additive and epistasis association: Manhattan plots of the additive (A
604 and C) and epistatic (B and D) association of SNPs across the 29 bovine autosomal
605 chromosomes for the Brahman (A and B) and Tropical Composite (C and D) populations. The
606 most likely candidate genes in the most significantly associated regions are annotated where
607 an obvious candidate could be identified according to the bovine reference genome assembly
608 ARS-UCS1.2. SNPs on odd-numbered chromosomes are in black and those on even-
609 numbered chromosomes are in yellow.

610

611

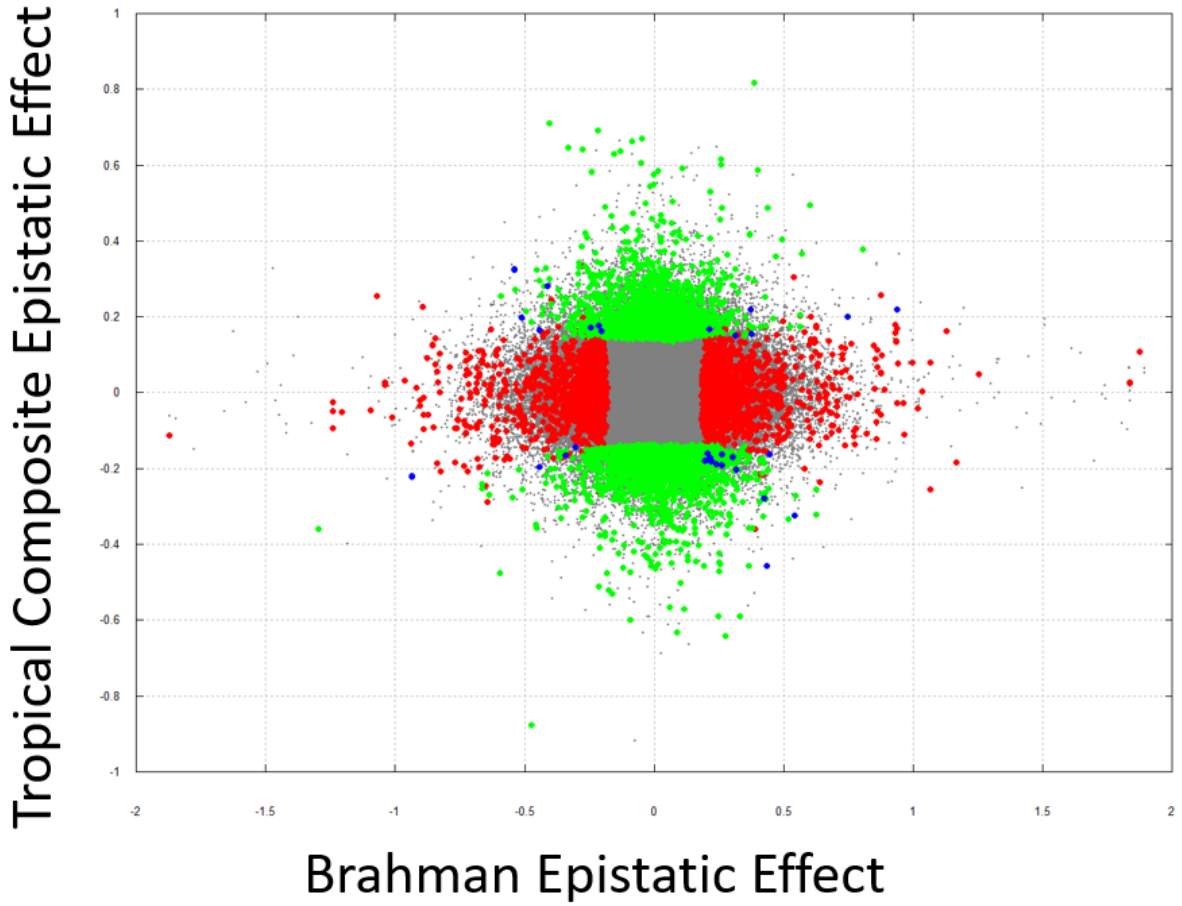


612

613 **Figure 2.** Scatter plot of the relationship between SNP additive (x-axis) and epistatic effects
614 (y-axis) for the Brahman (left panel) and Tropical Composite (right) populations. Red, green
615 and blue indicate significance (P-value < 0.001) for additivity, epistatic and both,
616 respectively.

617

618



619

620 **Figure 3.** Scatter plot of the relationship between SNP epistatic effect in the Brahman (x-axis)
621 and Tropical Composite (y-axis) populations. Red, green and blue indicate significance (P-
622 value < 0.001) in the Brahman, Tropical Composite, and both populations, respectively.

623