Revista Colombiana de Ciencias Pecuarias
ISSN: 0120-0690
Facultad de Ciencias Agrarias, Universidad de Antioquia

Zepeda Batista, José L; Carrillo Díaz, María I; Saavedra Jiménez, Luis A
Sources of bias in genetic association studies in cattle: A review
Revista Colombiana de Ciencias Pecuarias, vol.
31, no. 4, October-December, 2018, pp. 256-266
Facultad de Ciencias Agrarias, Universidad de Antioquia

DOI: 10.17533/udea.rccp.v31n4a02

Available in: http://www.redalyc.org/articulo.oa?id=295058168003

Literature review

doi: 10.17533/udea.rccp.v31n4a02

Sources of bias in genetic association studies in cattle: A review¤

Fuentes de sesgo en estudios de asociación genética en ganado bovino: Revisión de literatura

Fontes de viés nos estudos de associação genética em bovinos: Revisão da literatura

José L Zepeda Batista[1]

María I Carrillo Díaz[2*]

Luis A Saavedra Jiménez[1]

*Corresponding author: María Isabel Carrillo Díaz. Facultad de Medicina Veterinaria y Zootecnia, Universidad de Colima, km. 40,Autopista Colima-Manzanillo, Tecomán, Colima, CP 28100, México. Tel.:/Fax: +52 313 322 94 07 Ext. 52300. E-mail: mcarrillo13@ucol.mx.

[1]*Posgrado en Producción Animal, Universidad Autónoma Chapingo, Chapingo, Estado de México, México.*

[2]*Facultad de Medicina Veterinaria y Zootecnia, Universidad de Colima, Tecomán, Colima, México.*

(Received: June 9, 2017; accepted: April 20, 2018)

Abstract

Background: Genetic association studies have been increasingly used in cattle breeding programs. However, inconsistent results -such as positive, negative, or absence of association- across studies restrain reproducibility and proper implementation, propitiating the occurrence of bias.

Objective: To identify and classify potential sources of bias and determine possible strategies to avoid it in genetic association studies in cattle.

Source of bias in genetic association studies: Genetic and genomic sources of bias include effects associated with the gene loci governing expression. Sampling-related and statistical biases are related with factors such as stratification and database size.

Strategies to correct bias in genetic association studies: Correction strategies differ in nature. Genetic and genomic strategies are based on determining the appropriate approach to obtain and report the genetic information. Sampling-related and statistical strategies are based on grouping individuals with certain traits that lead to a reduction in heterogeneity.

Conclusion: It is necessary to consider the methodology used in previous studies to establish a hierarchy of sources of bias and facilitate decisions on the use of tools to reduce inconsistencies in the results of future studies.

Keywords: *association estimates, genetic bias, genetic improvement, sampling-related bias, statistical bias.*

Resumen

Antecedentes: Los estudios de asociación genética son cada vez más usados en los programas de mejoramiento genético. Sin embargo, resultados inconsistentes de los estudios -como positivos, negativos o ausencia de asociación- restringen la reproducibilidad y su aplicación adecuada, propiciando la aparición de sesgos.

Objetivo: Identificar y clasificar las fuentes potenciales de sesgo y determinar posibles estrategias para evitarlo en estudios de asociación genética en ganado.

Fuentes de sesgo en estudios de asociación genética: Las fuentes genéticas y genómicas de sesgo incluyen los efectos asociados con la expresión que gobierna los loci. Los sesgos

estadísticos y de muestreo están relacionados con factores como la estratificación y el tamaño de la base de datos.

Estrategias para corregir sesgos en estudios de asociación genética: Las estrategias de corrección difieren en naturaleza. Las estrategias genéticas y genómicas se basan en determinar el enfoque apropiado para obtener la información genética. Las estrategias estadísticas y relacionadas con el muestreo se basan en la agrupación de individuos con ciertos rasgos que conducen a una reducción de la heterogeneidad.

Conclusión. Se deben considerar las metodologías utilizadas en estudios previos para jerarquizar las fuentes de sesgo y facilitar las decisiones sobre el uso de herramientas para reducir inconsistencias en resultados futuros.

Palabras clave: *estimados de asociación, mejoramiento genético, sesgo de muestreo, sesgo estadístico, sesgo genético.*

Resumo

Antecedentes: Nos programas de criação de bovinos, os estudos de associação genética têm sido cada vez mais utilizados. No entanto, resultados inconsistentes, como positivos, negativos ou ausência de associação entre os estudos, restringem a reprodutibilidade e sua adequada implementação, propiciando o aparecimento de viés.

Objetivo: Identificar e classificar potenciais fontes de viés e determinar estratégias possíveis para evitá-lo nos estudos de associação genética em bovinos.

Fonte de viés em estudos de associação genética: Fontes genéticas e genômicas do viés incluem os efeitos associados aos genes que relacionam a expressão. Os vícios estatísticos e de amostragem estão relacionados a fatores como a estratificação e o tamanho do banco de dados.

Estratégias para corrigir os viéses nos estudos de associação genética: As estratégias de correção diferem na natureza. As estratégias genéticas e genômicas são baseadas na determinação da abordagem apropriada para obter e relatar a informação genética. As estratégias estatísticas e de amostragem baseiam-se no agrupamento de indivíduos com certos traços que levam a uma redução na heterogeneidade.

Conclusão: É necessário considerar a metodologia utilizada em estudos anteriores para estabelecer uma hierarquia de fontes de viés e facilitar decisões sobre o uso de ferramentas para reduzir inconsistências nos resultados de estudos futuros.

Palavras-chave: *estimativas de associação, melhoria genética, viés de amostragem, viés estatístico, viés genético.*

Introduction

Genetics association studies (GAS) aim to detect associations between one or more genetic polymorphism and a quantitative or discrete trait by testing for a correlation between a specific trait and a genetic variation (Lewis and Knight, 2012). The number of genetic association studies have increased, and their assessment has become a powerful approach to identify common and rare variants underlying complex diseases (Wu *et al*., 2012), discovering causative mutations (Schwarzenbacher *et al.*, 2016), or identification of quantitative trait loci (QTLs; Jahuey *et al*., 2016) on a population. Nevertheless, inconsistencies in GAS due to the combination of factors contribute to spurious or not consistently results (Table 1).

The inconsistencies found in GAS suggest that many original results could be false-positive (type I errors), especially in studies with systematic differences between sample and population, affecting their representativeness (Shingarpure and Xing, 2014). Thus, factors like paternity misidentification, stratification, and population structure are crucial in establishing sample size and its representativeness (Pyo and Wan, 2012). Other important source of inconsistencies in GAS are undetectable small genetic effects (false-negative, type II errors) (Lee, 2015). In this regard, poor design quality of the database usually means high *p*-values and low recognition of genetic associations (Ioannidis, 2005), especially when genotypes have low frequencies in the population or the study deals with low heritability traits (Satkoski *et al*., 2011).

Table 1. Results of genetic association studies between CSN3 gene with milk yield in dairy cattle.

	Study		
	Gustavsson et al. (2014)	**Duifhuis-Rivera et al. (2014)**	**Deb et al. (2014)**
Sampled animals	400	202	200
Reported effect	Positive	Absence	Positive
Best genotype*	AA	N/D	AB

*Best genotype: genotype reported with the best performance for milk yield ; N/D gen-trait association absent.

This lack of reproducibility tends to produce genetic associations of no value for genetic improvement. Ioannidis (2005) defined bias as the combination of design, data, analysis, and presentation factors resulting in findings that otherwise should not be produced. However, most reviews on bias in GAS have focused in the analysis of the genetic factors or address other factors as part of the genetic issues.

Bovine breed(s) considered in the study has been addressed as a genetic source of bias due to intra- and inter-racial diversity in genetic population (Lenstra *et al*., 2014), especially in the presence of crossbred animals (Dickerson, 1993). Besides, contemporary group factor has been confounded with the pure environment effect as it affects results due to the influence of interaction between genotype and environment (Ramírez-Valverde *et al*., 2008). Additionally, genomic factors of bias are associated with the gene loci governing expression, and are confused with environmental or residual variance (Burgueño *et al*., 2012), especially if those factors have an epigenetic nature such as genomic imprinting (Manolio *et al*., 2009), or influences more than one marker like the linkage disequilibrium, pleiotropy or polygenic effect (Pereira *et al*., 2016).

Lastly, even when the statistical model used in GAS is not usually confounded or assessed as a genetic factor of bias, its importance as a possible source of bias is remarkable since there are models that can work with just few markers at the same time (Pärna *et al*., 2012) and methods to determine the associations of thousands of markers at once. The variability resulting from the use of so different assessment methods could then be confounded with genetic or sampling factors of bias. Thus, it is necessary to classify bias in GAS according to its nature to better understand and reduce possible spurious results. Therefore, the objective of this study was to identify and classify potential sources of bias and determine possible strategies to avoid it in genetic association studies.

Sources of bias in genetic association studies

Different approaches, based on related or non- related individuals, have been used to carry out GAS (Table 2). The literature reports that some widely cited associations cannot be replicated due to inaccuracies in the approaches used to determine them (Sagoo *et al*., 2009). In this sense, inconsistencies in GAS could be attributable to factors such as genetic, genomic,

sampling-related, or statistical, which influence production traits, and contribute to the risk of false- positive results (Pärna *et al.*, 2012)

Genetic factors

The breed(s) used in the study could be a source of bias due to intra- and inter-racial bovine genetic population diversity (Figure 1). Besides, the presence of crossbred populations confers changes in the behavior of offspring, relative to that of the parents. Modifications can be evaluated by direct, maternal effects and heterosis of breeds and their crosses, with enough precision to predict the expected behavior of several breeding alternatives and mating systems (Dickerson, 1993). On this regard, Trail *et al*. (1984) reported direct and maternal effects on economic production traits in crossbred Boran cattle showing differences due to paternal or maternal breed.

Contemporary group (CG) is another genetic factor of bias, affecting results by the influence between genotype and environment interaction. Contemporary group as a fixed effect reduces bias in genetic comparisons, while the variance of the prediction error is reduced when CG is considered random (Ramírez-Valverde *et al*., 2008).

Genomic factors

Genomic factors of bias are associated with the gene loci governing expression and are confused with environmental or residual variance (Burgueño *et al*., 2012). Genomic imprinting bias in GAS is related with production traits due to their nature as epigenetic factors (Manolio *et al*., 2009). Han *et al.* (2013) mentioned that maternal effects could be confused with genomic imprinting because they produce the same parent-of-origin patterns of phenotypic variation, leading to an over- or underestimation in GAS of traits that include maternal effects. Su *et al.* (2012) reported a 3.5% bias decrease in genetic association values when additive, dominance, and epistatic effects are included in the analysis model compared to models previously reported that only included the additive effect.

Table 2. Former and current approaches used in genetic association studies.

	Approach	Advantages	Disadvantages
Family Based	TDT[1] FB-GWAS[2]	Quality control; robustness to population stratification; ability to perform genotyping quality control	Less power than pop-based GWAS; computationally demanding; not practical for late-onset diseases
Population based	Candidate polymorphism & Candidate gene	Determine if a given SNP or set of SNPs influences the trait directly; involve multiple SNPs within a single gene; capture information of the underlying genetic variability	SNPs may not serve as the true trait-causing variants; multiple SNPs measurements are needed to know a precise location on the genome
	Fine mapping	Set out to identify with a high level of precision the location of a trait-causing variant; determine the position on the genome of the causative mutation	High computational needs; specific software requirements; need for candidate gene studies to validate findings from GWAS
	Genome-wide	Identify associations between SNPs and a trait; involves the characterization of larger number of SNPs	

[1]TDT: transmission disequilibrium test; [2]FB-GWAS: family based genome-wide association study (Benyamin *et al*., 2009; Foulkes, 2009).

The type of markers used in GAS is a potential source of bias due to its effect on the analysis power to determine the linkage disequilibrium (LD) level of the data (Goode and Jarvik, 2005). Additionally, Rosenberg *et al.* (2010) reported mean information content (IC) differences between microsatellites and biallelic markers across the genome, with a better performance from the second one (Figure 2). Moreover, according with Kinghorn *et al*. (2010) correct choice of markers could increase the performance of quantitative genotyping.

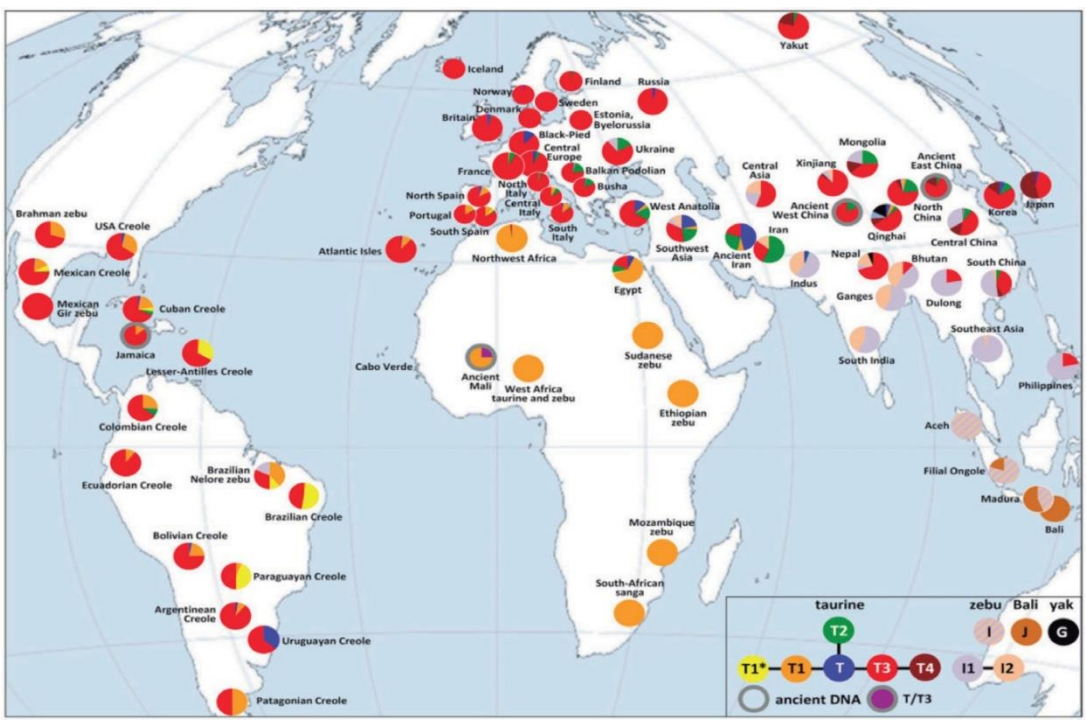

Figure 1. Diversity and distribution of major *Bos taurus* and *Bos indicus* haplogroups (taken from Lenstra *et al*., 2014).

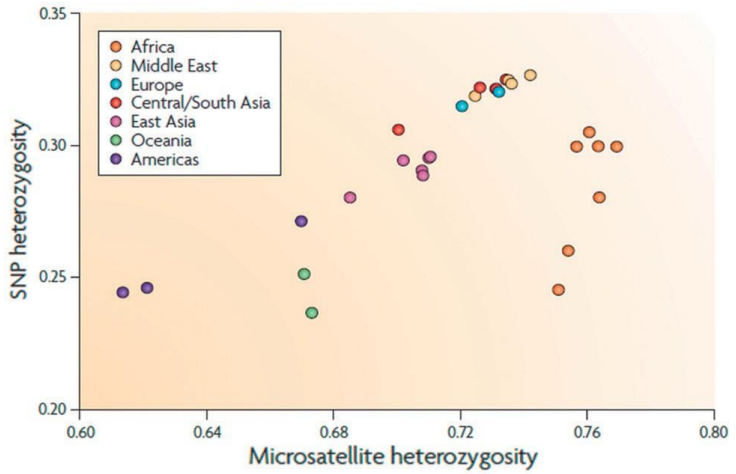

Figure 2. Information content variability for haplotype level in Europeans (taken from Rosenberg *et al*., 2010).

Monomorphism bias is based on the presence of uninformative markers in GAS (De *et al*., 2014). Thus, appearance of possible loss of power related with use of inadequate type of marker can occur. Another important genomic factor of bias is the minor allele frequency (MAF), it shows different behavior according to its effect size (Figure 3) and it is related with the Hardy-Weinberg proportions (HWP) potential bias. Therefore, MAF bias could occur if GAS use low density, monomorphic, or incorrect type of markers (Eynard *et al*., 2015).

Pleiotropic and polygenic effects are other important genetic sources of bias due to the influence over more than one economic trait in cattle (Figure 4).

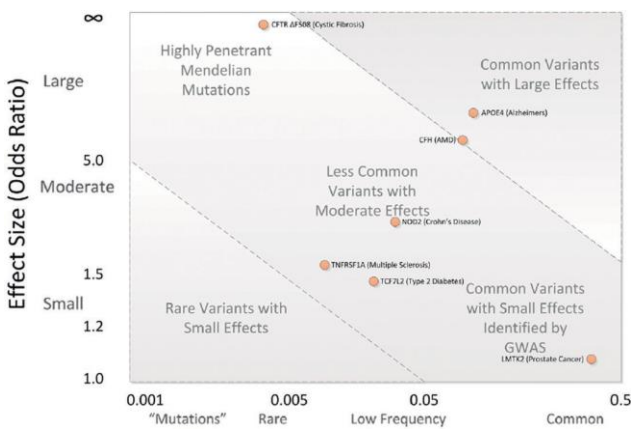

Figure 3. Types of MAF according to its effect size (taken from Bush and Moore, 2012).

Pleiotropic genes, such as *PLAG1*, operate like satellite regulators of the growth pathway while polygenic effect influences the estimation of genetic values. Segregation factor potential bias

is related with the monomorphic and type of marker factors of bias and highly influences the linkage disequilibrium (LD) in the population (Bush and Moore, 2012). Since, LD describes the degree to which an allele of one SNP is inherited or correlated with the allele of another SNP within a population (De *et al*., 2014), recombination events and type of markers to detect them are critical for the development of this factor bias.

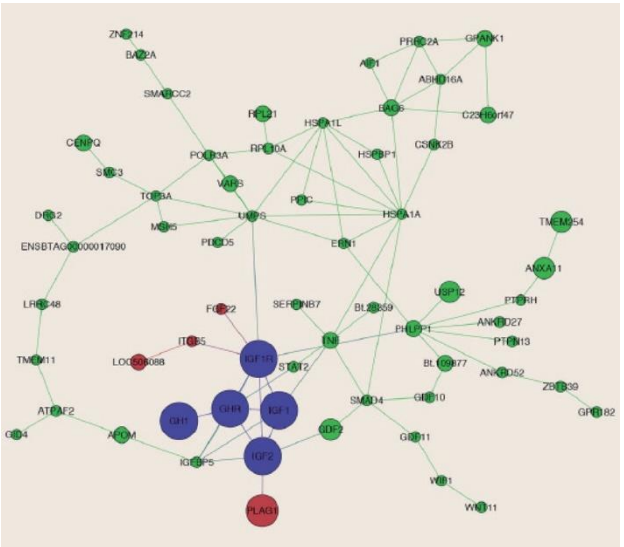

Figure 4. Network of candidate pleiotropic genes for carcass traits in Nellore cattle (taken from Pereira *et al*., 2016).

Genomic factors also include heritability bias, which is related with the gap between the phenotypic variance explained by GWAS results and those estimated by classical heritability. Zaitlen and Kraft (2012) mentioned that "missing heritability" could be due to presence of rare variants, epistatic and gene-environment interactions, or structural variation, that are not well captured by current GWAS or their analysis methods.

Sampling-related factors

Sample selection is another source of bias. It is defined asany systematic difference between the sample and the population affecting its representativeness (Shringarpure and Xing, 2014), leading to inaccurate estimation of relationships between variables (Figure 5). According to Pyo and Wan (2012), a larger sample size is required to achieve enough statistical power and to improve the ability of prediction. On the other hand, small sample size increases false negative rates and reduces the reliability of a study.

Paternity misidentification, stratification, and population structure are also factors related to sample size and its representativeness. On this regard, Visscher *et al.* (2002) determined a proportional selection response decrease of 2 to 3% for each 10% of paternity misidentification rate. Additionally, Sifuentes-Rincón *et al.* (2006) reported differences of 47% in the genetic values between simulated- and uncertain- paternity populations. Similarly, stratification bias could lead to spurious association that have no value as a tool for genetic improvement. In this sense, Zaitlen and Kraft (2012) mentioned that stratification bias arises when there is a difference in the phenotypic variance between the population.

Statistical factors

Statistical factors of bias are those related with the model and the nature of data used. According to Pyo and Wang (2012), the observed signal for association is considered statistically significant when the p-value is lower than a present threshold value (e.g., 0.05) to reject a null hypothesis of genetic association. Poor design quality of the database usually means high p-values and lower recognition of genetic associations (Ioannidis, 2005), especially if some of the genotypes have low frequency in the population or traits with low heritability (Satkoski *et al*., 2011).

Odd ratios can be a statistical factor of bias (Figure 6) when they are wrongly used as a weighted average to quantify genetic effects in GAS (Su and Lee, 2016). Due to their non-collapsible nature and tendency towards being null, a quantitative difference between conditional and marginal odd ratios in the absence of confounding is a mathematical oddity, not a reflection of bias (Groenwold *et al*., 2011).

Another factor that could cause bias is collinearity, which refers to the non-independence of predictor variables, usually in a regression-type analysis (Dormann *et al*., 2013). Yoo *et al.* (2014) mentioned that collinearity inflates the variance of regression parameters with a potential misidentification of relevant predictors in a statistical model. Dias *et al.* (2011) reported multicollinearity in genetic effects related with weaning weight in a Brazilian cattle population. They reported 9.8% of bias in the sum squared deviations, with variance inflation factors of 16 and 5.3 when using least square and ridge regression methodologies, respectively.

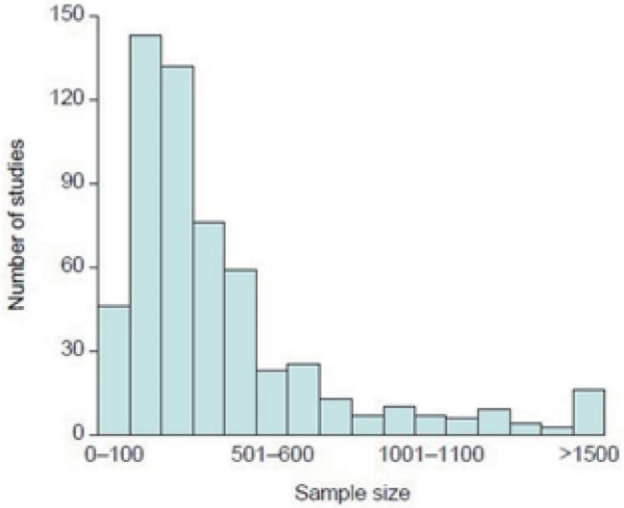

Figure 5. Sample size used in genetic association studies showing type I errors (taken from Ioannidis, 2005).

The presence of collinearity could lead to collider bias (*i.e.*, the reversal paradox), an artificial association created between exposures (A and B) when a shared outcome (X) is included in the model as a covariate (Day *et al*., 2016). Day *et al.* (2016) identified over 200 spurious GAS, when the shared outcome was included as a covariate in the model used to analyze the data.

One of the most important sources of bias in GAS is the statistical model chosen due to the differences within obtained results (Figure 7). The first models used in GAS included only fixed effects, causing bias when random effects were ignored (Miciński *et al*., 2007). On the other hand, mixed models can differentiate between the effects of random error and those from systematic error (Pärna *et al*., 2012). In the same way, Maximum likelihood (ML) is another procedure used in GAS with potential of bias. Kučerová *et al.* (2006) determined that ML can estimate genetic associations of casein genes and reported mean differences in protein concentration between 42 and 73% across κ-casein genotypes (AA, AB, AE, BB, and BE). However, when estimating a higher number of associations (*e.g.*, in genome-wide association studies), the power of mixed models and ML is reduced.

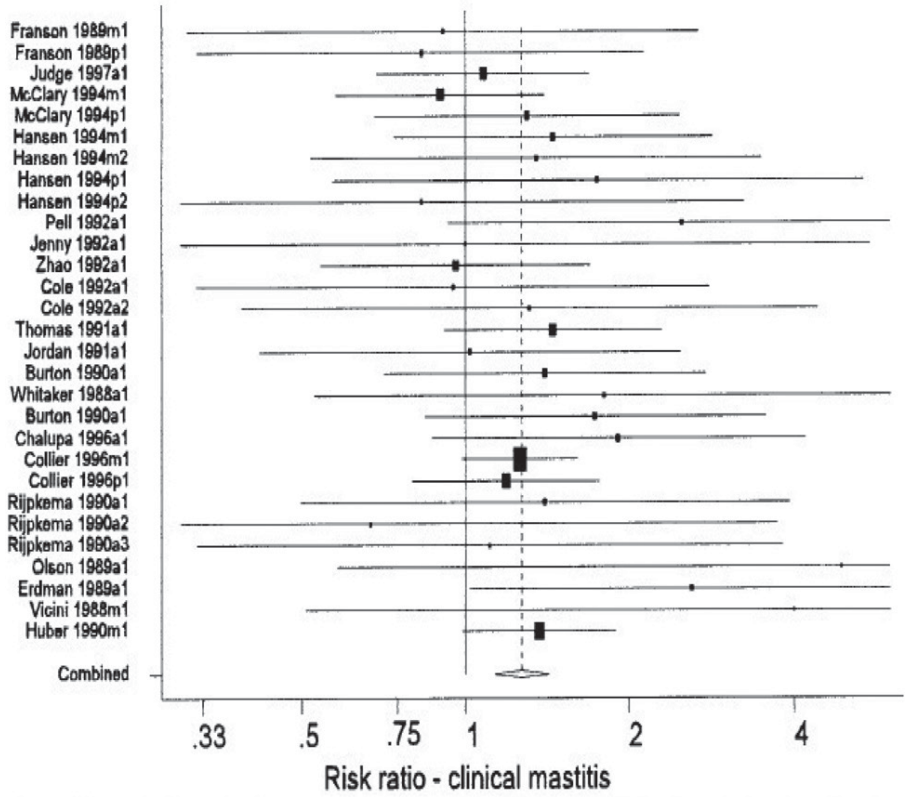

Figure 6. Forest plot of the effects or recombinant bovine somatotropin on the risk ratio of clinical mastitis (taken from Dohoo *et al*., 2003).

Extensive GAS need methods to determine the associations of thousands of markers at once. On this regard, De los Campos *et al.* (2009) reported Bayesian regression models (BM) able to adjust for the effects of thousands of markers simultaneously. Tenesa *et al.* (2003) observed that the differences between estimates obtained with ML and BM were small (about 5%), and both estimation procedures yielded essentially the same results. On the other hand, there are non-Bayesian models (NBM) that use information of genotyped and non-genotyped animals to perform genomic predictions (*e.g*. single-step genomic model) (Ma *et al*., 2015). However, due to its ability to estimate genetic association, even with markers lacking information, BM and NBM are under the influence of sample size and require a pedigree as complete as possible (Sahana *et al*., 2010).

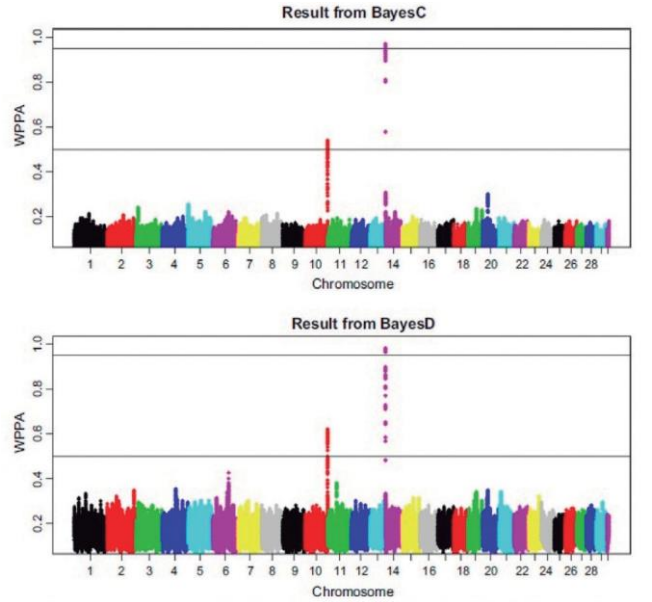

Figure 7. Probabilities of association obtained with two different Bayesian-based methods (taken from Bennewitz *et al*., 2017).

Strategies to correct biases in GAS

The aim of bias correction in GAS methodologies focuses on bias reduction, rather than its elimination (Pärna *et al*., 2012). Thus, it is possible to group bias correction into genetic-genomic, statistical, and methodological strategies.

Genetic-genomic strategies

Strategies of genetic-genomic bias correction rest on two aspects: source and conditions of genetic information. The source of genetic information in GAS refers to the approach used to obtain and report genetic information (i.e., single and multi-loci genotype or haplotype). Instead of analyzing the effects of individual alleles, some researchers estimate the effects of haplotypes defined by genes associated with the traits under study (Zhou et al., 2013), while other authors use multi-loci genotypes for the same purpose (Jaiswal et al., 2016).

The use of haplotypes and multi-loci genotypes can reduce bias arising from the way several genes are combined, the polygenic effect of the studied traits, and the position of the analyzed loci within the genome. However, unlike multi-loci genotypes, it has been argued that haplotypes have similar effects on different breeds (Andrés *et al*., 2007). As a result, a common

approach to analyzing the effects of haplotype has been to determine the most likely configuration for each and assume that this allocation of haplotypes is known without error when subsequent statistical analyses are performed. However, precise haplotype construction could be difficult, and often leads to biased estimates and reduced analytical power in GAS (Andrés *et al.*, 2007). In addition, when multiple loci are genotyped, haplotypes are unknown because there is no information about linkage phase of alleles at different loci (Sahana *et al*., 2010). Sahana *et al.* (2010) observed a high rate of type I error when using haplotypes as a fixed effect in genetic association models. Zhang *et al.* (2016) concluded that when there is a lack of tools available to reconstruct haplotypes, the best alternative is to use multi-loci genotypes regardless of whether phase adjustment information is available.

Other factors affecting the reliability of results are the number of markers used for reconstruction and the way that haplotypes and multi-loci genotypes are included in GAS models. For reconstruction, the best results have been obtained using 2 to 5 markers (Abdallah *et al*., 2004). In this sense, the main benefit of using haplotypes or multi-loci genotypes is their ability to explain most of the additive, dominance, and epistasis effects on the loci studied (Zhao *et al*., 2012). With respect to inclusion methods, incorporating haplotype as a random effect conveys better performance compared with models that include it as a fixed effect in terms of power, control of type I error, and precision (Boleckova *et al*., 2012). Hence, some of the probable HWP bias in these studies can be avoided, especially if the nature of the alleles being studied is considered. Kent *et al.* (2007) concluded that due to the risk of wrong associations, it is best to use common genetic variants greater than 10% as rare alleles generate biases in their association values and equally affect the values of common alleles. Therefore, the conditions needed to establish the use of haplotypes, genotypes, or both in GAS are of utmost importance for devising strategies to correct bias of genetic information.

Sampling-related and statistical strategies

Methodological strategies used to avoid sampling bias are based on grouping individuals or samples that share the same features in order to reduce heterogeneity and increase representativeness of results (Gustavsson *et al*., 2014). On the other hand, the use of previously reported information becomes important when establishing a methodological bias reduction strategy. Published information enables to use features and results previously validated, and helps to avoid the risk of bias related with transferring results among breeds (Poulsen *et al*., 2015).

Methodological strategies to reduce bias associated with statistical source are based on reviews, as well as the use of estimates and other literature results to determine the best models and features for the studied phenomenon (Brito *et al*., 2011). Commonly used association methods are based on family structure (pedigree) and case-control studies with unrelated individuals (De los Campos *et al*., 2009). However, case-control studies are the most viable to study genetic association because studies based on family structure involve extended testing periods (Kent *et al*., 2007). The presence of type I errors due to the subjective nature of the estimates (underlying assumptions) could address the risk of under- or overestimation of studied traits (Zoche-Golob *et al*., 2015). Therefore, the best strategy to reduce statistical bias lies in all aspects related to the predictive power of the approaches since it depends on all elements of bias that might arise.

In conclusion, it is necessary to consider the methodology used in previous GAS to establish a hierarchy of sources of bias and to facilitate better decisions on the use of tools to reduce inconsistencies in the results of future studies.

Acknowledgments

Authors thank CONACY for financial support of the first and third authors during their doctoral studies.

Conflicts of interest

The authors declare they have no conflicts of interest with regard to the work presented in this report.

References

Abdallah JM, Mangin B, Goffinet B, Cierco-Ayrolles C, Perez-Enciso M. A comparison between methods for linkage disequilibrium fine mapping of quantitative trait loci. Genet Res 2004; 83:41-47.

Andrés AM, Clark AG, Shimmin L, Boerwinkle E, Sing CF, Hixson JE. Understanding the accuracy of statistical haplotype inference with sequence data of known phase. Genet Epidemiol 2007; 31:659-671.

Bennewitz J, Edel C, Fries R, Meuwissen THE, Wellmann R. Application of a Bayesian dominance model improves power in quantitative trait genome-wide association analysis. Genet Sel Evol 2017; 49:7.

Benyamin B, Visscher PM, McRae A. Family-based genome-wide association studies. Pharmacogenomics 2009; 10(2):181-190.

Boleckova J, Christensen OF, Sørensen P, Sahana G. Strategies for haplotype-based association mapping in a complex pedigreed population. Czech J Anim Sci 2012; 57:1-9.

Brito LF, Silva FG, Melo ALP, Caetano GC, Torres RA, Rodrigues MT, Menezes GRO. Genetic and environmental factors that influence production and quality of milk of Alpine and Saanen goats. Genet Mol Res 2011; 10:3794-3802.

Burgueño J, De los Campos G, Weigel K, Crossa J. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Science 2012; 52:707–719.

Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 2012; 8(12): e1002822.

Day FR, Loh PR, Scott RA, Ong KK, Perry JRB. A robust example of collider bias in a genetic association study. Am J Hum Genet 2016; 98:392–393.

De R, Bush WS, Moore JH. Bioinformatics challenges in genome- wide association studies (GWAS). In: Ronald Trent editor. Clinical Bioinformatics, Methods in Molecular Biology. New York: Springer Science+Bussiness Media; 2014. p.63-81.

De los Campos, G., Naya H., Gianola D., Crossa J., Legarra A., Manfredi E., Weigel K., Cotes J. M. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182:375-385.

Deb R, Singh U, Kumar S, Singh R, Sengar G, Sharma A. Genetic polymorphism and association of kappa-casein gene with milk production traits among Frieswal (HF x Sahiwal) cross breed of Indian origin. Iran J Vet Res 2014; 15(4):406-408.

Dias RAP, Petrini J, Sterman JB, Pereira J, Santos R, Lopes AL, Barreto G. Multicollinearity in genetic effects for weaning weight in a beef cattle composite population. Livest Sci 2011; 142:188–194.

Dickerson GE. Evaluation of breeds and crosses of domestic animals. Rome (IT): Animal Production and Health Paper FAO; 1993.

Dohoo IR, DesCoteaux L, Leslie K, Fredeen A, Shewfelt W, Preston A, Dowling P. A meta-analysis review of the effects of recombinant bovine somatotropin 2. Effects on animal health, reproductive performance, and culling. Can J Vet Res 2003; 67:252-264.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, García JR, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013; 36:027–046.

Duifhuis-Rivera T, Lemus-Flores C, Ayala-Valdovinos MÁ, Sánchez-Chiprés DR, Galindo-García J, Mejía-Martínez K, González-Covarrubias E. Polymorphisms in beta and kappa casein are not associated with milk production in two highly technified populations of Holstein cattle in México. J Anim Plant Sci 2014; 24:1316-1321.

Eynard SE, Windig JJ, Leroy G, Van Binsbergen R, Calus MPL. The effect of rare alleles on estimated genomic relationships from whole genome sequence data. BMC Genetics 2015; 16(24):1-12.

Foulkes AS. Applied statistical genetics with R for population- based association studies. Cham (Swiss): Springer International Publishing AG; 2009.

Goode EL, Jarvik GP. Assessment and implications of linkage disequilibrium in genome-wide single-nucleotide polymorphism and microsatellite panels. Genet Epidemiol 2005; 29(Suppl.1): S72-S76.

Groenwold RHH, Moons KGM, Peelen LM, Knol MJ, Hoes AW. Reporting of treatment effects from randomized trials: A plea for multivariable risk ratios. Contemp Clin Trials 2011; 32:399-402.

Gustavsson F, Buitenhuis AJ, Johansson M, Bertelsen HP, Glantz M, Poulsen NA, Lindmark-Månsson H, Stålhammar H, Larsen LB, Bendixen C, Paulsson M, Andrén A. Effects of breed and casein genetic variants on protein profile in milk from Swedish Red, Danish Holstein, and Danish Jersey cows. J Dairy Sci 2014; 97:3866-3877.

Han M, Hu YQ, Lin S. Joint detection of association, imprinting and maternal effects using all children and their parents. Eur J Hum Genet 2013; 21:1449–1456.

Ioannidis JP. Why most published research findings are false. Plos Med 2005; 2(8):0696-0701 e124.

Jaiswal V, Gahlaut V, Meher PK, Mir RR, Jaiswal JP, Rao AR, Bayan HS, Gupta PK. Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (*T. aestivum L.*). Plos One 2016; 11(7):1-25 e0159343.

Jahuey-Martínez FJ, Parra-Bracamonte GM, Sifuentes-Rincón AM, Martínez-González JC, Gondro C, García-Pérez CA, López- Bustamantes LA. Genomewide association analysis of growth traits in Charolais beef cattle. J Anim Sci 2016; 94:4570-4582.

Kent JW, Dyer TD, Göring HH, Blangero J. Type I error rates in association versus joint linkage/association tests in related individuals. Genet Epidemiol 2007; 31:173-177.

Kinghorn BP, Bastiaansen JWM, Ciobanu DC, Van Der Steer HAM. Quantitative genotyping to estimate genetic contributions to pooled samples and genetic merit of the contributing entities. Acta Agric Scand A 2010; 60:3-12.

Kučerová J, Matějiček A, Jandurová OM, Sørensen P, Němcová E, Štípková M, Kott V, Bouška J, Frelich J. Milk protein genes CSN1S1, CSN2, CSN3, LGB and their relation to genetic values of milk production parameters in Czech Fleckvieh. Czech J Anim Sci 2006; 51:241-247.

Lee YH. Meta-analysis of genetic association studies. Ann Lab Med 2015; 35:283-287.

Lenstra JA, Ajmone-Marsan P, Beja-Pereira A, Bollongino R, Bradley DG, Colli L, De Gaetano A, Edwards CJ, Felius M, Ferretti L, Ginja C, Hristov P, Kantanen J, Lewis CM, Knight J. Introduction to genetic association studies. Cold Spring Harbor Protoc 2012; 3:297-306.

Lirón JP, Magee DA, Negrini R, Radoslavov GA. Meta-analysis of mitochondrial DNA reveals several population bottlenecks during worldwide migrations of cattle. Diversity 2014; 6:179-187.

Ma P, Lund MS, Nielsen US, Aamand GP, Su G. Single-step genomic model improved reliability and reduced the bias of genomic predictions in Danish Jersey. J Dairy Sci 2015; 98:9026-9034.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos ME, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CE, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature 2009; 461(7265):747–753.

Miciński J, Klupczyński J, Mordas W, Zablotna R. Yield and composition of milk from Jersey cows as dependent on the genetic variants of milk proteins. Pol J Food Nutr Sci 2007; 57:95-99.

Pärna E, Kaart T, Kiiman H, Bulitko T, Viinalass H. In: Chaiyabutr N, editor. Milk Production-Advanced Genetic Traits, Cellular Mechanism, Animal Management and Health. Rijeka: InTech; 2012. p.155-172.

Pereira AGT, Utsunomiya YT, Milanesi M, Torrecilha RBP, Carmo AS, Neves HHR, Carvalheiro R, Ajmone-Marsan P, Sonstegard TS, Sölkner J, Contreras-Castillo CJ, Garcia JF. Pleiotropic genes affecting carcass traits in *Bos indicus* (Nellore) cattle are modulators of growth. Plos One 2016; 11(7):1-13 e0158165.

Poulsen NA, Buitenhuis AJ, Larsen LB. Phenotypic and genetic associations of milk traits with milk coagulation properties. J Dairy Sci 2015; 98:1-9.

Pyo HE, Wan PJ. Sample size and statistical power calculation in genetic association studies. Genomics Inform 2012; 10(2):117-122.

Ramírez-Valverde R, Núñez-Domínguez R, Ruíz-Flores A, García-Muñiz JG, Magaña-Valencia F. Comparison of contemporary group definitions for genetic evaluation of Braunvieh cattle. Téc Pec Mex 2008; 46(4):359-370.

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nat Rev Genet 2010; 11(5):356-366.

Sagoo GS, Little J, Higgins JPT. Systematic reviews of genetic association studies. PLoS Med 2009; 6(3):e1000028.

Sahana G, Guldbrandtsen B, Janss L, Lund MS. Comparison of association mapping methods in a complex pedigreed population. Genet Epidemiol 2010; 34:455–462.

Satkoski JA, Malhi RS, Kanthaswamy S, Johnson J, Garnica WT, Malladi VS, Smith DG. The effect of SNP discovery method and sample size on estimation of population genetic data for Chinese and Indian rhesus macaques (*Macaca mulatta*). Primates 2011; 52:129-138.

Schwarzenbacher H, Burgstaller J, Seefried FR, Wurmser C, Hilbe M, Jung S, Fuerst C, Dinhopl N, Weissenböck H, Fuerst-Waltl B, Dolezal M, Winkler R, Grueter O, Bleu U, Wittek T, Fries R, Pausch H. A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. BMC Genomics 2016; 17: 400.

Shringarpure S, Xing EP. Effects of sample selection bias on the accuracy of population structure and ancestry inference. G3-Genes Genom Genet 2014; 4:901-911.

Sifuentes-Rincón AM, Parra-Bracamonte GM, De la Rosa RXF, Sánchez VA, Serrano MF, Rosales AJ. Importancia de las pruebas de paternidad basadas en microsatélites para la evaluación genética de ganado de carne en empadre múltiple. Tec Pec Mex 2006; 44(3):389-398.

Su YS, Lee WC. False appearance of gene–environment interactions in genetic association studies. Medicine 2016; 95(9):1-8.

Su G, Christensen OF, Ostersen T, Henryon M, Lund MS. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. Plos One, 2012, 7(9):e45293.

Tenesa A, Knott SA, Ward D, Smith D, Williams JL, Visscher PM. Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. J Anim Sci 2003; 81:617-623.

Trail JCM, Gregory KE, Durkin J, Sandford J. Crossbreeding cattle in beef production programmes in Kenya. II. Comparison of purebred Boran and Boran crossed with the Red

Poll and Santa Gertrudis breeds. Trop Anim Hlth Prod 1984; 16:191-200.

Visscher PM, Woolliams JA, Smith D, Williams JL. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. J Dairy Sci 2002; 85:2368– 2375.

Wu C, Li S, Cui Y. Genetic Association Studies: An Information Content Perspective. Curr Genomics 2012; 13(7):566-573.

Yoo W, Mayberry R, Bae S, Singh K, He Q, Lillard J. A study of effects of multicollinearity in the multivariable analysis. Int J Appl Sci Technol 2014; 4(5):9–19.

Zaitlen N, Kraft P. Heritability in the genome-wide association era. Hum Genet 2012; 131(10):1655-1664.

Zhang W, Li J, Guo Y, Zhang L, Xu L, Gao X, Zhu B, Gao H, Ni H, Chen Y. Multi-strategy genome-wide association studies identify the DCAF16-NCAPG region as a susceptibility locus for average daily gain in cattle. Sci Rep 2016; 6:38073.

Zhao H, Rebbeck TR, Mitra N. Analyzing genetic association studies with an extended propensity score approach. Stat Appl Genet Mol Biol 2012; 11(5):1-17.

Zhou L, Ding X, Zhang Q, Wang Y, Lund MS, Su G. Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. Genet Sel Evol 2013; 45:1-7.

Zoche-Golob V, Heuwieser W, Krömker V. Investigation of the association between the test day milk fat-protein ratio and clinical mastitis using a Poisson regression approach for analysis of time- to-event data. Prev Vet Med 2015; 121:64-73.