Genome-wide association study and functional analysis of feet and leg conformation traits in Nellore cattle

Giovana Vargas,[†] Haroldo H. R. Neves,[‡] Gregório Miguel F. Camargo,^{||} Vânia Cardoso,[‡] Danísio P. Munari,^{S,¶} and Roberto Carneiro^{†,¶,1}

[†]Departamento de Zootecnia, Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal, CEP 14884-900, Jaboticabal, SP, Brazil; [‡]Gensys Associated Consultants, CEP 90680-000, Porto Alegre, RS, Brazil; ^{||}Departamento de Zootecnista, Universidade Federal da Bahia (UFBA), CEP 40170-115, Salvador, BA, Brazil; ^SDepartamento de Ciências Exatas, Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal, CEP 14884-900, Jaboticabal, SP, Brazil; and [¶]Conselho Nacional de Desenvolvimento Científico e Tecnológico, CEP 71605-001, Brasília, Brazil

ABSTRACT: Feet and leg conformation is evaluated as a subset of conformational structure traits in dairy and beef cattle and is related to the feet and leg quality that can compromise the animals' productive performance and longevity. The aim of this study was to perform a genome-wide association study (GWAS) of two traits related to feet and leg conformation in Nellore cattle to identify chromosomal regions related to the expression of these traits. Phenotypic and pedigree data from 104,725 animals and genotypes from 1,435 animals and 407,730 SNPs were used. Feet and leg structure was evaluated as a binary trait (FL1) to identify yearling animals with feet and leg problems or as categorical score (FL2) to assess the overall quality of their feet and leg. The top ten 1-Mb windows that explained the largest proportion of the total genetic variance were identified and

functional enrichment analyses were performed. The 10 windows with large effects obtained for FL1 are located on chromosomes 1, 2, 6, 7, 8, 10, and 14, and together explained 8.96% of the additive genetic variance. For FL2, these windows are located on chromosomes 1, 7, 10, 11, 18, 20, 22, 28, and 29, explaining 8.98% of the additive genetic variance. Several candidate genes were identified, including *DLX2* which is associated with osteogenic differentiation, *IL-1 β* and *IL-1A* associated with some properties of articular cartilage, *PIT1* which plays an important role in bone physiology, and *CTSL* associated with rheumatoid arthritis. The results presented here should contribute to a better understanding of the genetic and physiologic mechanisms regulating both traits, and identifies candidate genes for future investigation of causal mutations.

Key words: beef cattle, *Bos taurus indicus*, candidate genes, weighted single step GBLUP

© The Author(s) 2018. Published by Oxford University Press on behalf of the American Society of Animal Science. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

J. Anim. Sci. 2018.96:1617–1627
doi: 10.1093/jas/sky079

INTRODUCTION

Adequate locomotion is directly associated with production and reproduction efficiency, as well as health and welfare. If an animal has feet

and leg problems, its longevity, welfare, productivity, and reproductive performance will be compromised (Pérez-Cabal and Alenda, 2002). In both beef and dairy cattle, several studies have recorded feet and leg conformation traits based on different indicator traits, such as rear leg rear view, rear leg side view, foot angle, and bone quality. The genetic correlation between conformation and feet and leg health traits (e.g., claw disorders

¹Corresponding author: rcar@fcav.unesp.br

Received December 4, 2017.

Accepted April 16, 2018.

and lameness) ranges from moderate to high magnitude, indicating the feasibility of using them as an indirect selection criteria to also control feet and leg health problems (Häggman et al., 2013; Chapinal et al., 2013; Ødegård et al., 2014).

Genome-wide association studies (GWAS) using SNP markers allow the identification of genomic regions (e.g., QTL; and causal mutations) that affect many economically important traits in livestock species and contributes to a better understanding of its biological mechanisms. The “weighted single-step GBLUP” (wssGBLUP) method proposed by Wang et al. (2012), enables combining pedigree, phenotype and genotype information in a single step, attributing different weights to the markers.

Using GWAS, previous studies have reported QTL and genes associated with feet and leg conformation traits in dairy cattle (Cole et al., 2011; Wu et al., 2013; Abo-Ismael et al., 2017). However, few studies are available for these traits in Nellore cattle, and more research in this area is necessary to find possible solutions that might reduce the incidence of feet and leg problems in the herds. Therefore, the main objective in this study was to identify potential QTLs and candidate genes affecting two feet and leg conformation traits in Nellore cattle using the wssGBLUP approach.

MATERIAL AND METHODS

Animal care and Use Committee approval were not necessary for this study because the data was obtained from an existing database of Nellore cattle.

Phenotypic and Pedigree Data

Phenotypic records for feet and legs and pedigree information were obtained from Nellore cattle from PAINT, the beef cattle breeding program of CRV Lagoa (www.crvlagoa.com.br).

The animals were raised in tropical pasture systems and belonged to herds located in Brazil and Paraguay, totaling 104,725 animals with own records for yearling weight collected from 2000 to 2013. The number of observations, contemporary groups (CG), sires, dams, and the score frequency distribution for FL1 and FL2 are presented in Table 1.

Feet and legs were evaluated by trained technicians of PAINT, which assigned visual scores to the overall structure of feet and legs at two different time points. The same technician was responsible for evaluating all animals of given management group. The recording of the evaluations was performed by using the PAINT software developed by CRV Lagoa (www.rivieratecnologia.com.br). Two different traits were defined: feet and leg evaluated as a binary trait (FL1), measured at yearling (about 550 d of age), to identify whether (FL1 = 1) or not (FL1 = 0) an animal had feet and leg problems; and feet and leg score (FL2), ranging from 1 (less desirable) to 5 (more desirable) was assigned to the top 20% animals according to the selection index adopted by the breeding program. The FL1 and FL2 scores were assigned based on measurement standards and guidelines specified by the Brazilian Association of Zebu Breeders (ABCZ, <http://www.abcz.org.br/>). The selection index considers expected progeny differences (EPD) for the following traits: birth to weaning weight gain, visual scores of conformation, finishing precocity, muscling, navel/prepuce at weaning and yearling, weaning to yearling weight gain, temperament at yearling, and scrotal circumference. These top 20% animals were candidates to receive the Special Certificate of Identification and Production (CEIP), an official certificate that testifies the value of seedstock delivered by breeding programs, i.e., animals that are genetically classified as superior (Horimoto et al., 2007).

Figure 1 illustrates the structure of feet and legs when viewed from the front, hind and side.

Table 1. Summary statistics of feet and legs in Nellore cattle

Traits ^a	Data structure ^b				Score frequency distribution ^c					
	N	NCG	NS	ND	0	1	2	3	4	5
FL1	96,836	2,105	748	73,272	92,469 (95.5)	4,367 (4.5)	-	-	-	-
FL2	14,708	897	340	12,920	-	1,040 (7.1)	3,192 (21.7)	6,088 (41.4)	3,616 (24.6)	772 (5.2)

^aFL1 = feet and leg evaluated as a binary trait (scores assigned to all animals measured at yearling); FL2 = feet and leg scores ranging from 1 (less desirable) to 5 (more desirable), assigned to the top 20% animals according to the selection index applied to this population.

^bN = number of observations; NCG = number of contemporary groups; NS = number of sires; ND = number of dams.

^cThe absolute frequency of each score is given, followed by the relative proportion (in %).

Figure 1. Illustration of the structure of feet and legs when viewed from the front, hind, and side. The figure was created by the authors.

In general, the front legs should be straight when viewed from the front. For a better understanding of the structural appearance of the animal, an imaginary vertical line can be drawn from the shoulder to the middle of the claw. An animal can be classified as “knock-kneed” (when the knee joints lie inside this line), and as “bow-legged” (when the knee joints lie outside this line). An animal classified as “straight-legged” does not have an ideal flexing and shock-absorbing effect. The

“sickle hocked” condition is when the leg joints angle degree is smaller than the ideal. Viewed from behind, the hock joint should be in a straight line. An animal is classified as “cow hocked” when the hocks are rotated inwards and the hooves rotated outwards, and “bow-legged” when the legs are wide at the hocks, but the feet are turned in. The conformation of the feet must be short and steeply angled, high in the heel and claw, with a sole somewhat concave. In general, long or excessively short

claws may indicate too much or not enough pastern angle, leading to an excessive growth or wear.

In the breeding program considered in this study, animals that received score 1 for FL1 could not be candidates to receive the CEIP and therefore were not evaluated for FL2. The recording of FL2 occurred a few days after the animals were evaluated at yearling, when the results of routine genetic evaluations were released and CEIP candidates were defined. Furthermore, animals with an undesirable score for FL2 (FL2 = 1) were not allowed to receive CEIP certificate, a procedure that often results in economic losses because genetically superior animals for growth and carcass traits have to be culled.

Data from CG with fewer than 10 records and/or without variability in the respective trait were removed from further analyses. CG were defined considering the effects of herd, year, and season of birth, management group at weaning, date of measurement at yearling, and management group at yearling. Connectedness among CG was verified using the AMC software (Roso and Schenkel, 2006). Data from disconnected CG, i.e., less than 10 genetic links to the main dataset, were also discarded. After data editing, a total of 96,836 and 14,708 individual records for FL1 and FL2 distributed over 2,105 and 897 CGs, respectively, remained in the dataset for further analyses. The pedigree file contained 188,694 animals tracing back up to five generations.

Genotypes

A total of 1,435 genotypes from 667 yearling animals, 402 sires and 366 dams, were used for GWAS. These animals are part of the genomic reference population of PAINT (www.crvlagoa.com.br) and were chosen among those animals with own performance records and/or progeny evaluated for FL1 or FL2. The sires were genotyped with the Illumina BovineHD chip (HD, Illumina, Inc., San Diego, CA), whereas yearling animals and dams were genotyped with the Illumina BovineSNP50 v2 chip (50K, Illumina, Inc.) containing 777,962 and 54,609 SNPs, respectively. Animals genotyped with 50K were imputed to HD (accuracy of imputation > 0.97, [Carvalho et al., 2014](#)) using the FImpute v2.2 software ([Sargolzaei et al., 2014](#)).

The R software (R Core Team, 2016) was used for quality control (QC) analysis of phenotypic and genotypic datasets. The genotypic QC filtered out SNP markers located in non-autosomal regions, SNPs with unknown or duplicated positions, and markers with a *P*-value for Hardy-Weinberg

equilibrium test smaller than 10^{-5} , SNP call rate lower than 0.95 and a minor allele frequency (MAF) lower than 0.02. The remaining number of SNPs after QC was 407,730. All samples with a call rate lower than 0.9 were also removed from the analyses.

Genome-Wide Association Analyses

For GWAS, single-trait threshold animal models were used for FL1 and FL2. The SNP effects were estimated using the wssGBLUP method proposed by [Wang et al. \(2012\)](#), which enables combining pedigree, phenotypic and genotypic information in a single step. The wssGBLUP method does not require to restrict the use of phenotypes of animals that were not genotyped or to compute pseudo-phenotypes. Furthermore, wssGBLUP is a valuable tool in situations where many animals are phenotyped, but only a reduced proportion of them are also genotyped, which is common in commercial livestock production systems, and may result in increased the accuracy of genomic prediction ([Aguilar et al., 2010](#)) and, as a consequence, increase the precision of QTL detection ([Wang et al., 2012](#)). The wssGBLUP first computes the breeding values then the SNP effects, as described below.

The SNP effect estimates were obtained from the predicted breeding values, which were calculated according to the model:

$$l = X\beta + Z_a a + e,$$

where *l* is a vector of underlying liabilities of FL1 or FL2; β is a vector of systematic effects (CG and a linear effect of yearling age); *a* is a vector of random additive direct genetic effects (breeding values); *e* is a vector of random residual effects, and *X* and *Z_a* are incidence matrices that relate the liabilities in *l* to the effects in β and *a*, respectively. Given that the management groups contained animals of the same sex and that a single technician evaluated all animals of the same management group, there was no need to include both effects (i.e., sex and technician) in the model.

The (co)variance of *a* and *e* were assumed as:

$$\text{var} \begin{bmatrix} a \\ e \end{bmatrix} = \begin{bmatrix} H\sigma_a^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

where σ_a^2 and σ_e^2 are the additive direct and the residual variances, respectively, and *H* is the matrix which combines pedigree and genomic information ([Aguilar et al., 2010](#)), and *I* is an identity matrix.

The underlying liabilities of FL1 and FL2 were defined as:

$$\text{FL1: } \{\text{FL1} = 0 \text{ if } l \leq t_1;$$

$$\text{FL1} = 1 \text{ if } l > t_1\};$$

$$\text{FL2: } \{\text{FL2} = 1 \text{ if } l \leq t_1;$$

$$\text{FL2} = 2 \text{ if } t_1 < l \leq t_2;$$

$$\text{FL2} = 3 \text{ if } t_2 < l \leq t_3;$$

$$\text{FL2} = 4 \text{ if } t_3 < l \leq t_4;$$

$$\text{FL2} = 5 \text{ if } l > t_4\}$$

in which: t_1 to t_{j-1} correspond to thresholds that define, on the underlying scale, the mutually exclusive ordered categories of FL1 and FL2, under the assumptions that $t_1 < t_2 < \dots < t_{j-1}$, and $t_0 = -\infty$ and $t_j = +\infty$ (Gianola and Foulley, 1983). In the analysis of FL1, as σ_e^2 is not estimable, the parameterization $\sigma_e^2 = 1$ was adopted (Gianola and Sorensen, 2002). In the case of FL2, the thresholds t_1 and t_2 were kept fixed at 0 and 1, respectively, so that σ_e^2 could be estimated (Sorensen et al., 1995).

The parameters of the model were estimated by replacing the inverse of the numerator relationship matrix A (A^{-1}) in the regular mixed model equations (Henderson, 1984) by the inverse of matrix H (H^{-1}) (Misztal et al., 2009; Aguilar et al., 2010)—a matrix which combines pedigree and genomic information—using the Gibbs sampling program THRGIBBS1F90 (Tsuruta and Misztal, 2006). The inverse of matrix H can be written as follows:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

where A is the numerator relationship matrix based on pedigree for all animals; A_{22} is the numerator relationship matrix based on pedigree for genotyped animals only; and G is the genomic relationship matrix for genotyped animals. Flat prior distributions were assumed for the variance components and for the fixed and random effects. The default prior distributions were assumed for the parameters of the model, and a single chain with a length of 1,000,000 iterations was generated for each analysis (FL1 and FL2), applying a conservative burn-in period of 100,000 cycles and a thinning interval of 250 cycles. The *coda* package

v.0.19–1 of the R software (R Core Team, 2016) was employed to assess the convergence of the chains using Geweke's (1992) and Heidelberger and Welch (1983) tests, in addition to visual inspection. We also estimated the genomic heritabilities for each trait using the estimated variance components obtained from single-trait analyses.

The solutions of SNP effect estimates ($\hat{u}$) were obtained as a function of the genomic estimated breeding values (GEBVs) using the formula:

$$\hat{u} = \mathbf{DZ}[\mathbf{ZDZ}]^{-1}\hat{a}_g,$$

where D is a diagonal matrix with weights for the SNPs, Z is an incidence matrix of genotypes for each locus, and $\hat{a}_g$ is the vector of GEBVs of genotyped animals. The $\hat{u}$ vector and the matrix D were iteratively recomputed using the following algorithm proposed by Wang et al. (2012):

- 1) In the first iteration $t = 0$ and $D_{(0)} = I$, where t is the iteration number, $D_{(t)}$ is matrix D at iteration t , and I is an identity matrix.
- 2) Calculate the SNP effects at iteration t ($\hat{u}_{(t)}$).
- 3) Recalculate the diagonal elements of D as: $d_{i(t+1)}^* = \hat{u}_{i(t)}^2 / 2p_i(1-p_i)$ for all SNPs, where p_i is the allele frequency of the reference allele of the i^{th} marker and i is the i^{th} SNP.
- 4) Normalize $D_{(t+1)} = (\text{tr}(D_{(0)}) / \text{tr}(D_{(t+1)}^*)) D_{(t+1)}^*$.
- 5) $t = t + 1$.
- 6) Exit, or loop to step 2.

Thus, the matrix G was recalculated for the prediction of SNP effects at each iteration, and $\hat{a}_g$ was obtained only once and does not change during iterations. The algorithm was run over two iterations in total. The BLUPF90 family programs were used for GWAS (Misztal et al., 2008).

QTL Mapping

The proportion of variance explained by SNPs within consecutive nonoverlapping 1-Mb windows was adopted as the criterion to identify potentially important genomic regions. A total of 2,522 windows, containing a mean ($\pm$ standard deviation) of 161.7 ($\pm$ 47.6) SNPs, were considered. The top 10 windows, which explained the highest proportion of additive variance, were defined as potentially important genomic regions. These windows were further investigated by searching for genes and QTLs reported in the QTLdb database (Hu et al., 2013) and located in the same genomic regions, using the UMD3.1 bovine genome assembly (Zimin

Figure 2. Manhattan plot for percentage of variance explained by 1-Mb window for FL1 obtained by wssGBLUP. FL1 = feet and legs evaluated as a binary trait measured at yearling to identify whether (1) or not (0) an animal had feet and leg problems. Genetic variance, % = the proportion of variance explained by SNPs within consecutive nonoverlapping 1-Mb windows.

et al., 2009) as the reference map. The presence of a QTL in adjacent windows (1-Mb to the left and to the right of the top 10 windows) was also evaluated as its effect can be captured by neighboring SNPs due to linkage disequilibrium (LD). The presence of a previously described QTL in the QTLdb database was double checked in the originally listed references to ensure that the same reference assembly was used. Scientific papers published in journals indexed in PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) and ScienceDirect (<http://www.sciencedirect.com/>) were also considered in order to support the biological evidences of those genomic regions located in the top 10 windows. The genes in the top 10 windows were identified using the MapViewer of the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov>).

RESULTS AND DISCUSSION

The estimate of genomic heritability for FL1 (0.16 ± 0.03) was lower than FL2 (0.44 ± 0.10). These results were close to the classical heritability estimated based on pedigree information and the same phenotypic records from a previous study (Vargas et al., 2017). Figures 2 and 3 show the Manhattan plot with the percentages of additive genetic variance explained by 1-Mb windows obtained in the second iteration for FL1 and FL2, respectively. The higher percentage of additive variance explained by a 1-Mb window was equal to 1.37 for FL1 and 1.20 for FL2. The Manhattan plots suggest that FL1 and FL2 are polygenic traits with many genes explaining a relatively small proportion of the total genetic variance of the traits. Some potentially important regions were observed

as described below. However, larger sample sized studies can potentially identify other novel regions.

The top ten 1-Mb windows obtained in the second iteration that explained the highest proportion of genetic variance in FL1 and FL2 can be observed in Tables 2 and 3, respectively. The windows are located on chromosomes 1, 2, 6, 7, 8, 10, and 14 for FL1 and on chromosomes 1, 7, 10, 11, 18, 20, 22, 28, and 29 for FL2 and explained 8.96% and 8.98% of the additive genetic variance for FL1 and FL2, respectively. Describing the results obtained from the genes for FL1 and FL2, it was observed that the window 64 on chromosome 8 (Table 2) and window 40 on chromosome 22 (Table 3) explained the highest proportion of variance estimated for FL1 and FL2, respectively. The SNP explaining the largest variance of each window (leading SNP) captured in some cases almost all the variance explained by the window. In other cases, the leading SNP explained a much smaller percentage of the variance than the whole window, possibly due to high LD between markers within that window. The windows with large effects found for the two traits under investigation were not overlapping, although a negative and moderate genetic correlation between these traits was reported by Vargas et al. (2017) (-0.47 ± 0.02) in Nellore cattle. Therefore, this finding suggests that the genetic correlation between FL1 and FL2 may be due to small-effect genes that contribute to both traits. In addition, the existence of the selection bias in FL2 could contribute to the difference between the genomic regions identified for both traits, FL2 is evaluated only in part of the yearling animals, all of which scored as 0 for FL1.

The present study identified windows associated with FL1 and FL2 that had been previously

Figure 3. Manhattan plot for percentage of variance explained by 1-Mb window for FL2 obtained by wssGBLUP. FL2 = feet and leg scores ranging from 1 (less desirable) to 5 (more desirable), assigned to the top 20% animals according to the selection index applied to this population, which consists of productive and reproductive traits measured at weaning and yearling. Genetic variance, % = the proportion of variance explained by SNPs within consecutive nonoverlapping 1-Mb windows.

associated with feet and leg conformation traits in dairy cattle. In a GWAS using data of contemporary U.S. Holstein cows, [Cole et al. \(2011\)](#) reported top 100 SNP effects for rear legs (side view) that were located within some of the top windows identified in the present study, namely 1_125, 2_25, and 18_59. The authors also detected top 100 SNPs for feet and leg-related traits within the following adjacent windows identified in the present study: 11_11 and 22_41 for feet and leg score, 11_48 for rear legs (rear view), and 20_15 for foot angle and rear legs (rear view). The 1_138 and 14_7 windows associated with FL1 and 10_32 associated with FL2 has also been described by [Wu et al. \(2013\)](#) and [Van Der Spek et al. \(2015\)](#) using data of sole ulcer and rear leg side view in dairy cattle. [Abo-Ismael et al. \(2017\)](#) also detected significant and suggestive SNPs in the 7_92 and 18_59 windows associated with feet and leg conformation and body traits, including bone quality, stature, and dairy strength. The identification of overlapping regions among different studies and independent populations contributes to validate previous candidate genes known to affect feet and leg conformation traits in cattle.

[Tables 4](#) and [5](#) show the annotated genes within the top 10 windows for FL1 and FL2, respectively. Most of these genes are associated with metabolic, cellular component organization or biogenesis, and biological regulation processes with potential biological explanations for the expression of feet and leg conformation traits. For FL1, 85 annotated genes were identified in the top 10 windows, with window 8_64 containing the largest number of genes (22 genes). For FL2, 116 annotated genes were detected in the top 10 windows. The window

with the largest number of genes was 18_59 (47 genes), and there was only one window without annotated genes (1_125). For FL1, the window 2_25 harbors two distalless-related homeodomain transcription factors (*DLX1* and *DLX2* genes). *Dlx* genes play key roles in the development and morphogenesis of the head and limb skeleton. The *DLX2* gene plays important roles in the regulation and migration of ectomesenchymal cells and in osteogenic differentiation ([Depew et al., 2005](#); [Jin, 2005](#)). Its overexpression induces osteogenic differentiation, upregulating bone formation-associated genes ([Sun et al., 2015](#)).

The window 11_47 associated with FL2 has detected representative genes of the group of inflammatory cytokines (*IL-1 β* and *IL-1A*). The *IL-1 β* gene is one of the eleven representatives of the IL-1 family and has been associated with articular cartilage and other elements of joints ([Wojdasiewicz et al., 2014](#)). The *IL-1A* gene has been reported to be related to matrix degradation and loss of mechanical properties of articular cartilage ([Wilson et al., 2007](#)). Both biological functions of these candidate genes (*IL-1 β* and *IL-1A*) are in accordance with biological mechanisms of the traits investigated in the present study.

The *PiT1* (or *SLC20A1*) gene on chromosome 11 associated with FL2, encodes a widely expressed plasma membrane protein that functions as a high-affinity Na⁺-phosphate (Pi) cotransporter and may play a role in bone physiology. The importance of *PiT1* gene in mineralizing processes has been demonstrated in vitro in osteoblasts, chondrocytes, and vascular smooth muscle cells ([Beck et al., 2003](#)). Recent studies have suggested that *PiT1* gene may be implicated in pathological

Table 2. Top 10 windows explaining the highest proportion of genetic variance for feet and legs binary score (FL1) in Nellore cattle

Rank	Chr_Win ^a	Var (%) ^b	nSNPs ^c	posSNP (bp) ^d	varSNP (%) ^e
1 st	8_64	1.37	185	63785867	1.23
2 nd	2_25	1.18	175	24367446	1.14
3 rd	6_51	1.07	184	50956746	1.05
4 th	1_54	0.92	167	53064069	0.92
5 th	7_92	0.88	201	91783596	0.83
6 th	8_83	0.84	197	82855034	0.52
7 th	14_31	0.74	165	30158402	0.31
8 th	1_136	0.67	243	135947945	0.62
9 th	14_6	0.65	259	5347162	0.33
10 th	10_10	0.64	216	9445786	0.36

^aChr = chromosome; Win = 1-Mb window within the chromosome.

^bVar = additive genetic variance explained by the SNPs within the window.

^cnSNPs = number of SNPs within the window.

^dposSNP = position (UMD3.1) of the leading SNP within the window.

^evarSNP: variance explained by the leading SNP.

vascular calcification, leading to diseases such as osteoporosis (Crouthamel et al., 2013). The *CTSL* gene on chromosome 8 is associated with the development of rheumatoid arthritis, a chronic autoimmune joint disease in humans that causes bone destruction and osteoporosis (Kakegawa et al., 1993; Ishibashi et al., 1999). Since the process of skeletal bone formation is highly conserved across species, this gene is a good candidate for conformation traits in Nellore cattle.

The GWAS presented here allowed us to identify regions associated with feet and leg conformation traits already reported in the literature. These findings are of great value to future GWAS meta-analysis and will also serve as a basis for fine mapping studies, aiming to identify causal

mutations for these traits. The identification and validation of QTLs is important for breeding programs, irrespective of the genetic architecture of the trait, because it allows improving the accuracy of genomic predictions (Pérez-Enciso et al., 2015). For instance, by incorporating functional SNPs in the development of new SNP chip panels. In addition, different genes were also reported in other studies conducted for feet and leg conformation traits, suggesting the existence of a genetic difference between these populations. Some factors may contribute to these differences, such as LD structure across populations, coverage of the SNP chip in the analyzed breed, method and population size, since the present study used few animals when compared to most of the studies cited above. However,

Table 3. Top 10 windows explaining the highest proportion of genetic variance for feet and legs categorical score (FL2) in Nellore cattle

Rank	Chr_Win ^a	Var (%) ^b	nSNPs ^c	posSNP (bp) ^d	varSNP (%) ^e
1 st	22_40	1.20	180	39122937	0.94
2 nd	11_47	1.14	180	46647572	1.12
3 rd	1_125	0.95	121	124546798	0.81
4 th	29_25	0.90	188	24618881	0.89
5 th	11_10	0.84	214	9751032	0.18
6 th	7_88	0.84	118	87643539	0.81
7 th	20_16	0.83	105	15412961	0.33
8 th	10_32	0.83	85	31865789	0.63
9 th	18_59	0.74	130	58213152	0.22
10 th	28_21	0.71	165	20385511	0.42

^aChr = chromosome; Win = 1-Mb window within the chromosome.

^bVar = additive genetic variance explained by the SNPs within the window.

^cnSNPs = number of SNPs within the window.

^dposSNP = position (UMD3.1) of the leading SNP within the window.

^evarSNP: variance explained by the leading SNP.

Table 4. Annotated genes within the top 10 windows of feet and legs binary score (FL1)

Chr_IntWin ^a	N ^b	Genes ^c
8_63 to 65	22	<i>CCDC180, TDRD7, LOC785115, LOC101905141, TMOD1, TSTD2, LOC101905362, NCBP1, XPA, FOXE1, LOC100847313, C8H9orf156, HEMGN, LOC101905809, ANP32B, LOC101905751, LOC100847202, NANS, TRIM14, CORO2A, TBC1D2, GABBR2</i>
2_24 to 26	16	<i>TRNAG-UCC, PDK1, ITGA6, MIR2352, TRNAW-CCA, DLX2, DLX1, METAP1D, SLC25A12, HAT1, LOC101904968, DYNC1I2, LOC784052, LOC786733, CYBRD1, LOC101905180</i>
6_50 to 52	2	<i>TRNAS-GGA, LOC100298058</i>
1_53 to 55	8	<i>LOC101904329, CD47, IFT57, MYH15, LOC783477, KIAA1524, DZIP3, TRAT1</i>
7_91 to 93	2	<i>LOC100848524, LOC101904534</i>
8_82 to 84	9	<i>LOC101906474, DAPK1, LOC101906531, CTSL, LOC786322, FBP2, FBP1, LOC101907415, C8H9orf3</i>
14_30 to 32	5	<i>LOC101907975, LOC101907872, MIR124A-2, BHLHE22, CYP7B1</i>
1_135 to 137	8	<i>EPHB1, LOC101903747, LOC101903894, LOC101903829, KY, LOC101905906, CEP63, ANAPC13</i>
14_5 to 7	5	<i>LOC100296770, LOC101907283, COL22A1, LOC101907449, FAM135B</i>
10_9 to 11	8	<i>AP3B1, SCAMP1, LHFPL2, LOC101905812, LOC101905869, ARSB, ARSB, DMGDH</i>

^aChr = chromosome; IntWin = interval of 1-Mb to the left and to the right of the top 10 windows.

^bN = number of genes within the window.

^cGenes = NCBI symbol of annotated genes in *Bos taurus* genome (annotation release 103), using the *Bos taurus* UMD3.1 assembly.

association studies of SNPs for production, carcass and reproductive traits have identified common genomic regions using small population sizes in beef cattle (e.g., Cesar et al., 2014; Martínez et al., 2017; Olivieri et al., 2017; Seabury et al., 2017).

In conclusion, the present GWAS permitted the identification of chromosomal regions associated with feet and leg conformation traits in Nelore cattle. Some of these regions were located within or near previously reported QTL on independent dairy cattle populations, increasing the evidence of the importance of those regions for feet and leg conformation traits. The identification of these

regions contributes to a better understanding of the genetic and physiologic mechanisms regulating both traits, and identifies candidate genes for future investigation of causal mutations.

ACKNOWLEDGMENTS

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil (grant number 2013/25312-5); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil (grant number 303606/2009-6).

Table 5. Annotated genes within the top 10 windows of feet and legs categorical score (FL2)

Chr_IntWin ^a	N ^b	Genes ^c
22_39 to 41	5	<i>CADPS, FEZFL2, C22H3orf14, PTPRG, LOC100337379</i>
11_46 to 48	26	<i>LOC786288, TTL, POLR1B, CHCHD5, LOC101904088, LOC101904038, LOC101904134, SLC20A1, LOC101903687, NT5DC4, CKAP2L, IL1A, LOC101904177, IL1B, IL37, IL36G, IL36A, IL36B, IL36RN, IL1FL10, IL1RN, PSD4, PAX8, LOC100294744, LOC101905095, LOC101905150</i>
1_124 to 126	-	-
29_24 to 26	8	<i>LOC100336918, LOC101904402, SLC6A5, PRMT3, HTATIP2, LOC518027, LOC101906541, LOC101906617</i>
11_9 to 11	15	<i>MRPS9, GPR45, TGFBAP1, LOC100848208, C11H2orf49, FHL2, LOC100848188, LOC101903987, TACR1, POLE4, HK2, LOC101904209, LOC100297235, SEMA4F, M1AP</i>
7_87 to 89	1	<i>LOC100140939</i>
20_15 to 17	7	<i>LOC101906396, RNFL180, LOC101906438, MIR320A-2, LOC101906491, LOC101906576, HTR1A</i>
10_31 to 33	4	<i>DPH6, LOC101903969, TRNAS-AGA, TRNAC-GCA</i>
18_58 to 60	47	<i>LOC617909, MIR99B, MIRLET7E, MIR125A, LOC100337268, LOC101907856, LOC101907942, HAS1, VN2R408P, LOC787554, ZNF613, LOC101902824, LOC101902702, ZNF432, ZNF432, ZNF614, LOC101903149, LOC100300607, ZNF350, LOC787309, BOSTAUV1R406, BOSTAUV1R407, LOC101908164, LOC100337475, BOSTAUV1R410, BOSTAUV1R411, LOC100847477, BOSTAUV1R413, PPP2R1A, LOC101903360, BOSTAUV1R414, BOSTAUV1R-PS409, BOSTAUV1R416, BOSTAUV1R-PS410, LOC101903564, LOC539675, VN1R2, LOC787057, LOC100848895, LOC101902345, LOC101904049, LOC101903979, LOC506495, LOC781189, LOC101904568, LOC101904503, LOC101904435</i>
28_20 to 22	3	<i>LOC101904967, LOC101905021, LOC781358</i>

^aChr = chromosome; IntWin = interval of 1-Mb to the left and to the right of the top 10 windows.

^bN = number of genes within the window.

^cGenes = NCBI symbol of annotated genes in *Bos taurus* genome (annotation release 103), using the *Bos taurus* UMD3.1 assembly.

LITERATURE CITED

- Abo-Ismael, M. K., L. F. Brito, S. P. Miller, M. Sargolzaei, D. A. Grossi, S. S. Moore, G. Plastow, P. Stothard, S. Nayeri, and F. S. Schenkel. 2017. Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle. *Genet. Sel. Evol.* 49:82. doi:10.1186/s12711-017-0356-8
- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752. doi:10.3168/jds.2009-2730.
- Beck, G. R. Jr, E. Moran, and N. Knecht. 2003. Inorganic phosphate regulates multiple genes during osteoblast differentiation, including Nrf2. *Exp. Cell. Res.* 288:288–300. doi:10.1016/S0014-4827(03)00213-1
- Carvalho, R., S. A. Boison, H. H. R. Neves, M. Sargolzaei, F. S. Schenkel, Y. T. Utsunomiya, A. M. P. O'Brien, J. Sölkner, J. C. McEwan, C. P. Van Tassell, et al. 2014. Accuracy of genotype imputation in Nellore cattle. *Genet. Sel. Evol.* 46:69. doi:10.1186/s12711-014-0069-1
- Cesar, A. S. M., L. C. A. Regitano, G. B. Mourão, R. R. Tullio, D. P. D. Lanna, R. T. Nassu, M. A. Mudado, P. S. N. Oliveira, M. L. do Nascimento, A. S. Chaves, et al. 2014. Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. *BMC Genetics*. 15:39. doi:10.1186/1471-2156-15-39
- Chapinal, N., A. Koeck, A. Sewalem, D. F. Kelton, S. Mason, G. Cramer, and F. Miglior. 2013. Genetic parameters for hoof lesions and their relationship with feet and leg traits in Canadian Holstein cows. *J. Dairy Sci.* 96:2596–2604. doi:10.3168/jds.2012-6071
- Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, Jr, B. A. Crooker, C. P. Van Tassell, J. Yang, S. Wang, L. K. Matukumalli, et al. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* 12:408. doi:10.1186/1471-2164-12-408
- Crouthamel, M. H., W. L. Lau, and C. M. Giachelli. 2013. Sodium-dependent phosphate cotransporters and phosphate-induced calcification of vascular smooth muscle cells: redundant roles for PiT-1 and PiT-2. *Arterioscl. Throm. Vas.* 33:2625–2632. doi:10.1161/ATVBAHA.113.302249
- Depew, M. J., C. A. Simpson, M. Morasso, and J. L. Rubenstein. 2005. Reassessing the dlx code: the genetic regulation of branchial arch skeletal pattern and development. *J. Anat.* 207:501–561. doi:10.1111/j.1469-7580.2005.00487.x
- Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. 4th ed. Clarendon Press, Oxford, UK.
- Gianola, D., and J. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15:201–224. doi:10.1186/1297-9686-15-2-201
- Gianola, D., and D. Sorensen. 2002. Likelihood, Bayesian, and MCMC methods in quantitative genetics. In: K., Dietz, M., Gail, K., Krickeberg, J., Samet, A., Tsiatis, eds, 2nd ed. *Statistics for Biology and Health*: Springer, New York, NY; p. 740.
- Häggman, J., J. Juga, M. J. Sillanpää, and R. Thompson. 2013. Genetic parameters for claw health and feet and leg conformation traits in Finnish Ayrshire cows. *J. Anim. Breed. Genet.* 130:89–97. doi:10.1111/j.1439-0388.2012.01007.x
- Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Oper. Res.* 31:1109–1144.
- Henderson, C. R. 1984. Applications of linear models in animal breeding. University of Guelph Press, Guelph, Canada.
- Horimoto, A. R., J. B. Ferraz, J. C. Balieiro, and J. P. Eler. 2007. Phenotypic and genetic correlations for body structure scores (frame) with productive traits and index for CEIP classification in Nellore beef cattle. *Genet. Mol. Res.* 6:188–196.
- Hu, Z. L., C. A. Park, X. L. Wu, and J. M. Reecy. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 41:871–879. doi:10.1093/nar/gks1150
- Ishibashi, O., Y. Mori, T. Kurokawa, and M. Kumegawa. 1999. Breast cancer cells express cathepsins B and L but not cathepsins K or H. *Cancer Biochem. Biophys.* 17:69–78.
- Jin, Y. 2005. Mouse development biology and embryo research method. 1st ed. People's Medical Publishing House, Beijing, China.
- Kakegawa, H., T. Nikawa, K. Tagami, H. Kamioka, K. Sumitani, and T. Kawata. 1993. Participation of cathepsin L on bone resorption. *FEBS Lett.* 321:247–250.
- Martínez, R., D. Bejarano, Y. Gómez, R. Dasoneville, A. Jiménez, G. Even, J. Sölkner, and G. Mészáros. 2017. Genome-wide association study for birth, weaning and yearling weight in Colombian Brahman cattle. *Genet. Mol. Biol.* 40:453–459. doi:10.1590/1678-4685-gmb-2016-0017
- Misztal, I. 2008. BLUPF90 - a flexible mixed model program in Fortran 90. Animal and Dairy Science, University of Georgia, Athens, GA. <http://nce.ads.uga.edu/~ignacy/numpub/blupf90/docs/blupf90.pdf> (accessed February 21, 2016).
- Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92:4648–4655. doi:10.3168/jds.2009-2064
- National Center for Biotechnology Information. 1997. U.S. National Library of Medicine, USA. <http://www.ncbi.nlm.nih.gov/> (accessed February 21, 2016).
- Ødegård, C., M. Svendsen, and B. Heringstad. 2014. Genetic correlations between claw health and feet and leg conformation in Norwegian red cows. *J. Dairy Sci.* 97:4522–4529. doi:10.3168/jds.2013-7837.
- Olivieri, B. F., M. E. Z. Mercadante, J. N. S. G. Cyrillo, R. H. Branco, S. F. M. Bonilha, L. G. Albuquerque, R. M. O. Silva, and F. Baldi. 2017. Genomic regions associated with feed efficiency indicator traits in an experimental Nellore cattle population. *Plos One*. 12:e0171845. doi:10.1371/journal.pone.0164390
- Pérez-Cabal, M. A., and R. Alenda. 2002. Genetic relationships between lifetime profit and type traits in Spanish Holstein cows. *J. Dairy Sci.* 85:3480–3491. doi:10.3168/jds.S0022-0302(02)74437-8
- Pérez-Enciso, M., J. C. Rincón, and A. Legarra. 2015. Sequence- vs. Chip-assisted genomic selection: accurate biological information is advised. *Genet. Sel. Evol.* 47:43. doi:10.1186/s12711-015-0117-5
- R Core Team. 2016. R: a language and environment for statistical computing. R Found. Stat. Comput., Vienna, Austria.
- Roso, V. M., and F. S. Schenkel. 2006. A computer program to assess the degree of connectedness among contemporary groups. In: *Proceeding of the 8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, Minas Gerais, Brazil; pp. 27–26.

- Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478. doi:10.1186/1471-2164-15-478.
- Seabury, C. M., D. L. Oldeschulte, M. Saatchi, J. E. Beever, J. E. Decker, Y. A. Halley, E. K. Bhattarai, M. Molaei, H. C. Freetly, S. L. Hansen, et al. 2017. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics*. 18:386. doi:10.1186/s12864-017-3754-y
- Sorensen, D. A., S. Andersen, D. Gianola, and I. Korsgaard. 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27:229–249. doi:10.1186/1297-9686-27-3-229
- van der Spek, D., J. A. van Arendonk, and H. Bovenhuis. 2015. Genome-wide association study for claw disorders and trimming status in dairy cattle. *J. Dairy Sci.* 98:1286–1295. doi:10.3168/jds.2014-8302
- Sun, H., Z. Liu, B. Li, J. Dai, and X. Wang. 2015. Effects of DLX2 overexpression on the osteogenic differentiation of MC3T3-E1 cells. *Exp. Ther. Med.* 9:2173–2179. doi:10.3892/etm.2015.2378.
- Tsuruta, S., and I. Misztal. 2006. THRGIBBS1F90 for estimation of variance component with threshold and linear models. In: *Proceeding of the 8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, Minas Gerais, Brazil; p. 253.
- Vargas, G., H. H. R. Neves, V. Cardoso, D. P. Munari, and R. Carneiro. 2017. Genetic analysis of feet and leg conformation traits in Nellore cattle. *J. Anim. Sci.* 95:2379–2384. doi:10.2527/jas.2016.1327
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb)*. 94:73–83. doi:10.1017/S0016672312000274.
- Wilson, C. G., A. W. Palmer, F. Zuo, E. Eugui, S. Wilson, R. Mackenzie, J. D. Sandy, and M. E. Levenston. 2007. Selective and non-selective protease inhibitors reduce il-1 induced cartilage degradation and loss of material properties. *Matrix Biol.* 26:259. doi:10.1016/j.matbio.2006.11.001
- Wojdasiewicz, P., Ł. A. Poniatowski, and D. Szukiewicz. 2014. The role of inflammatory and anti-inflammatory cytokines in the pathogenesis of osteoarthritis. *Mediators Inflamm.* 2014:561459. doi:10.1155/2014/561459.
- Wu, X., M. Fang, L. Liu, S. Wang, J. Liu, X. Ding, S. Zhang, Q. Zhang, Y. Zhang, L. Qiao, et al. 2013. Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. *BMC Genomics*. 14:897. doi:10.1186/1471-2164-14-897
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, et al. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42. doi:10.1186/gb-2009-10-4-r42