Systems Genetics Investigations for Feed Intake, Feed Efficiency and Performance in Nellore (*Bos indicus*) Cattle

M.H.A. Santana[1], **H.N. Kadarmideen**[2], **S.D. Pant**[2], **P.A. Alexandre**[1], **G.A. Oliveira Junior**[1], **R.C. Gomes**[3], **Y.T. Utsunomiya**[4], **H.H.F. Neves**[4], **J.F. Garcia**[4], **H. Fukumasu**[1] and **J.B.S. Ferraz**[1].

[1]University of São Paulo, Brazil, [2]University of Copenhagen, Denmark, [3]EMBRAPA, Brazil, [4]State University of São Paulo, Brazil

ABSTRACT: Genome wide association study (GWAS) was performed in 720 Nellore bulls and steers for dry matter intake (DMI), average daily gain (ADG), feed conversion ratio (FCR) and residual feed intake (RFI), using GRAMMAR-Gamma association test. Genes within 50kb flanking regions of SNPs with the highest association with the phenotypes were extracted, and pathway analysis was performed on the KEGG and DAVID databases. GWAS and search for genes/pathways were performed in the statistical environment R by GenABEL and NCBI2R packages, respectively. Several genomic regions that were significantly associated with phenotypes ($p=9.27\times10^{-5}$) were identified by GWAS. Near these regions, many genes were found for the four phenotypes studied (e.g. *ARG2, ATP8A1, KCNJ, PLAG2G7* and *ZNF746*). The analysis identified three main metabolic pathways that influence the four phenotypes concomitantly. These processes are mainly related to ion transport (aldosterone-regulated sodium reabsorption), body composition (T cell receptor signaling pathway) and control of feed intake (proteoglycans).
Keywords: beef cattle; GWAS; system biology

Introduction

Feed intake is very important in livestock production from an economic perspective. Growing and finishing cattle in feedlots is considered a promising solution to reduce deforestation of tropical forests and to reduce methane emissions.

However, expenditure on feed in feedlots represents the largest costs variable (Anderson et al., 2005). Therefore, these animals must be efficient in using food for their growth, or in other words, they must have good feed efficiency. Intake, weight gain and feed efficiency are the main measures for feedlot finishing systems (Nkrumah et al., 2007), together with meat quality.

To better understand these traits, several studies were conducted in beef cattle involving genomic information (Barendse et al., 2007; Moore et al., 2009; Rolf et al., 2012; Lu et al., 2013). However, there are still only a few genome-wide association studies on *Bos indicus*, especially for feedlot traits in Nellore cattle.

Genome-Wide Association Studies (GWAS) with dense panels of genetic markers can pinpoint genomic regions that influence these phenotypes (Rolf et al., 2012). Systems genetics, first proposed in livestock / animal biosciences by Kadarmideen et al. (2006), is an area of genetics within the context of system biology and it combines GWAS results and information genome-wide gene expression and metabolic pathways data. Systems genetics approaches have been shown to be crucial for understanding the biological role of genetic variants and genes that are identified by GWAS. We use this approach in searching for genes and biological processes involved in the regulation of these feed intake/efficiency traits and provide insights into the physiological relationship between feed intake, efficiency and weight gain. These biological processes are not yet well described for Nellore cattle in the literature so far.

The objective of the study is to conduct GWAS and system genetic analyses of feed intake, feed efficiency and weight gain and reveal key genes / genetic variants and physiological processes that are related to the regulation of feed intake/efficiency and performance traits.

Materials and Methods

Phenotypic data. Data of 720 male Nellore (550 ± 115 days of age and 380 ± 51 kg), collected in experiments performed in Brazil from 2007 to 2012, were used in this study. The animals had been evaluated for average daily gain (ADG), dry matter intake (DMI), residual feed intake (RFI) and feed conversion ratio (FCR). These phenotypes were tested for normality (Shapiro-Wilk) and homoscedasticity of residues (Breusch-Pagan). Additionally, outlier observations departing ±3 standard deviations from the average were excluded from further analyses.

Genome wide association studies. The 720 animals were genotyped in two different commercial panels (384 young bulls in Illumina BovineHD with 777,692 SNPs and 336 young bulls and steers in Illumina BovineSNP50 with 54,609 SNPs). To combine the information from two panels of markers an imputation study was performed in software *FImpute v2.2* (Sargolzaei et al., 2012). Only markers imputed with over 95% accuracy were added in the data set. The data were subjected to quality control in which SNPs on sex chromosomes, with a minor allele frequency below 2%, with call rate below 95% and the Fisher exact test for

Hardy-Weinberg equilibrium less than 1×10^{-5} were excluded analysis. After the imputation study and quality control 290,620 SNPs were retained in 672 animals. The association test used was GRAMMAR-Gamma (Svishcheva et al., 2012). The contemporary group and age were included as fixed effects. GWAS for each phenotype was performed in GenABEL v1.7-6 (Aulchenko et al., 2007) package of R statistical environment (R Core Team 2013). The genome-wide significance threshold was determined after applying a modified version of Bonferroni correction to account for multiple testing (Gao et al., 2008). This was computed as $\frac{\propto}{\sqrt{nsnp}}$ where α is the nominal significance threshold of 0.05 and *nsnp* is total number of SNPs used in the analysis (Mantel, 1980), which resulted in $\alpha = 9.27 \times 10^{-5}$.

Finding genes in the vicinity of associated SNPs. The gene set associated with each trait was determined by selecting genes located within 50 kilobases flanking regions of each significantly associated SNP using the function GetRegion in the NCBI2R package (Melville, 2014) available within the R (R Core Team, 2013).

Pathway analyses. Pathway analyses were based on information available within the Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/). Firstly, genes were assigned to biological pathways using GetPathways function available within NCBI2R package (Melville 2014). Subsequently, the number and names of genes in each pathway was determined using a mapPathwaytoname

function (http://biobeat.wordpress.com/category/r/). To determine KEGG pathways significantly associated with different traits, a Fisher's exact test was implemented in R to test if genes contained within the geneset for each trait were overrepresented amongst all genes contained in any given pathway. Since *Bos indicus* genome is poorly annotated, functional annotation of the genesets was also performed using DAVID Gene Functional Classification Tool (Huang et al., 2009) against both bovine (*Bos Taurus*) and human backgrounds.

Results and Discussion

Association analysis. The number of SNPs with different ranges of genome-wide (GW) statistical significance of associations (p-values) are shown in Table 1. Figure 1 shows the Manhattan plots of association tests for ADG, DMI, FCR and RFI with cut-off thresholds represented as the modified Bonferroni.

Table 1. Number of SNPs in each probability (p-value) range in the association test

Trait	≤ 0.001	0.001≥ p ≤0.01	0.01≥ p ≤0.05	
ADG	335	1618	6907	
DMI	520	2225	11,736	
FCR	346	3019	11,934	
RFI		504	2527	12,046

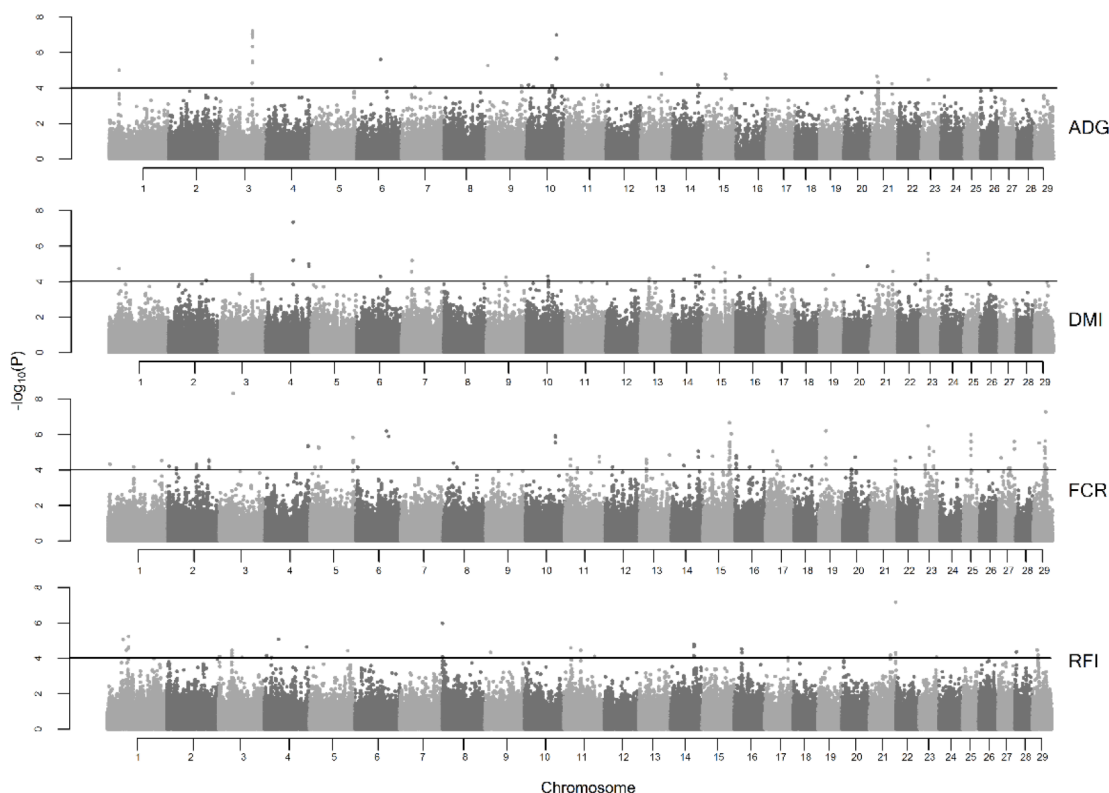**Figure 1: Manhattan plots of −Log (p-value) for average daily gain (ADG), dry matter intake (DMI), feed conversion ratio (FCR) and residual feed intake (RFI)**
The horizontal lines represents the Bonferroni modified significance threshold ($\alpha = 9.27 \times 10^{-5}$).

ADG is affected by GW significant SNPs on chromosome 3, 6 and 10. For DMI, chromosomes 4 and 23 harbours GW significant SNPs. For feed efficiency traits, GW significant SNPs were on chromosomes 10, 15, 19, 23, 25 and 29 for FCR and 4, 8, 14 and 21 for RFI.

System genetics. Based on results from GWAS , a total of 35 SNPs associated with ADG, 43 SNPs associated with DMI, 139 SNPs associated with FCR and another 44 SNPs associated with RFI were used to locate genes within 50 kb flanking regions for gene set enrichment analyses. Several genes were found in genomic regions near these GW significant SNPs. Resulting in genesets containing 32 genes for ADG, 37 genes for DMI, 95 genes for FCR and 50 genes for RFI. Overrepresentation analyses identified one KEGG pathway associated with FCR and another two associated with RFI (Table 2).

Table 2. KEGG pathways found in enrichment analyzes of feed intake, efficiency and performance traits in Nellore cattle

Pathway*	A	B	C	D	p	Trait
1	3	39	92	32568	0.006	**FCR**
2	3	106	47	32501	0.011	**RFI**
3	3	222	47	32385	0.045	**RFI**

*1 = aldosterone-regulated sodium reabsorption, 2 = T cell receptor signaling pathway, 3 = Proteoglycans
A – Frequency in Geneset, B – Frequency in Genome, C – Missing in Geneset, D – Missing in Genome

Aldosterone-regulated sodium reabsorption is mainly related to the renal reabsorption of sodium under the influence of aldosterone that is released from the adrenal glands. Mineral reabsorption is known to be an energetically expensive process and renal mineral reabsorption alone accounts for as much as 10% of the total maintenance energy requirement (Balaban and Mandel 1980, Summers et al., 1988, Kies et al., 2005). Therefore, it is possible that fluctuations in sodium reabsorption indirectly influence feed efficiency by regulating energy expenditure.

T cell receptors are crucial regulators of cell-mediated immunity, besides other cellular signaling pathways that regulate cell fate. While the biological mechanisms enabling T cell receptors to influence feed efficiency are not clear, the T cell receptor signaling pathway has been previously found to be associated with RFI in Angus cattle (Rolf et al., 2012). Members of this pathway are also known to be upregulated in adipose tissue of pigs differing in feed efficiency in response to caloric restrictions (Lkhagvadorj et al., 2010).

Finally, an additional KEGG pathway 'proteoglycans' found to be associated with RFI, and refers to the role that proteoglycans like hyaluronan, syndecans, glypican, perlecan and syndecans, play in influencing tumor progression. Some of these proteoglycans have also been linked with feed intake. For example, syndecan-3 (cell surface proteoglycan that mainly act are receptors) expression on hypothalamic neurons expressing MC3R and MC4R is known to regulate appetite and feed intake (Gantz and Fong, 2003). Therefore, it is possible that other proteoglycans also play specific roles that contribute to the overall regulation of feed intake.

Conclusion

Several genomic regions along the autosomal chromosomes were related to feed intake, feed conversion ratio, residual feed intake and weight gain in the association study. Many genes were found to be related to these traits, like *AKT1, ARG2, ATP8A1, KCNJ, PLAG2G7, SLC12A2* and *ZNF804B*.

However, systematically three major physiological processes related to all measures together were highlighted. These pathways are mainly related to the energy spent on maintenance of the ionic transport, body composition and possible regulatory processes of appetite. These results are consistent with the literature that links these pathways as being fundamental for expression of these phenotypes in animals.

Acknowledgment

Literature Cited

Anderson, R.V., Rasby, R.J., Klopfenstein, T.J. et al. (2005) J Anim Sci 83:694–704.

Aulchenko, Y.S., Ripke, S., Isaacs, A. et al. (2007). Bioinformat. 23(10):1294–1296.

Balaban, R.S., Mandel, L. (1980). J. Physiol. 304(1):331-348.

Barendse, W., Reverter, A., Bunch, R.J. et al. (2007) Genet 176:1893–905.

Gantz, I., Fong, T.M. (2003). Amer. J. Physiol. Endoc. Met. 284(3):468-474.

Gao, X., Starmer, J., Martin E.R. (2008). Gen. Epid. 32:361-369.

Huang, D.W., Sherman, B.T., Lempocki, R.A. (2009). Bioinform. Res. Nat. Protoc. 4(1):44-57.

Kadarmideen, H.N., von Rohr, P., Janss, L.L.G. (2006) Mammalian Genome 17:548-64.

Kies, A.K., Gerrits, W.J.J., Schrama, J.W. et al. (2005). J. Nutr. 135(5):1131-1138.

Lkhagvadorj, S., Qu, L., Cai, W. et al. (2010). Americ. J. Phys. 298(2):494-507.

Lu, D., Miller, S., Sargolzaei, M. et al. (2013). J. Anim. Sci. 91:3612–3633.

Mantel, N. (1980). Biometrics 36: 381-399.

Melville, S. (2014). NCBI2R: R package version 1.4.5.

Moore, S.S., Mujibi, F.D., Sherman, E.L. (2009). J. Anim. Sci. 87(E Suppl.):41–47.

Nkrumah, J.D., Basarab, J.A., Wang, Z. et al. (2007). J. Anim. Sci. 85:2711–2720.

Rolf, M., Taylor, J.F., Schnabel, R.D. et al. (2012). Animal Genet 43(4):367-374.

R Core Team (2013). R Found Stat Comp, Vienna, Austria.

Sargolzaei, M., Chesnais, J., Schenkel, F. et al. (2012). Open Ind. Session, 30.

Summers, M., McBride, B.W., Milligan, L.P. (1988). Dobson, Cornell Univ. Press: 273:21-37.

Svishcheva, G.R., Axenovich, T.I., Belonogova, N.M. et al. (2012). Nat. Genet. 44:1166-1170.