



Sequence characterization and comparative analysis of the gastrotropin gene in buffalo (*Bubalus bubalis*)

N.B. Stafuzza, M.M. Borges and M.E.J. Amaral-Trusty

Departamento de Biologia, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista São José do Rio Preto, SP, Brasil

Corresponding author: M.E.J. Amaral-Trusty
E-mail: eamaral@ibilce.unesp.br

Genet. Mol. Res. 13 (4): 10934-10942 (2014)
Received August 8, 2014
Accepted November 5, 2014
Published December 19, 2014
DOI <http://dx.doi.org/10.4238/2014.December.19.15>

ABSTRACT. In this study, we compared the complete sequence of the *FABP6* gene from an animal representing the Murrah breed of the river buffalo (*Bubalus bubalis*) with the gene sequence from different mammals. The buffalo *FABP6* gene is 6105 bp in length and is organized into four exons (67, 176, 90, and 54 bp), three introns (1167, 1737, and 2649 bp), a 5'UTR (93 bp), and a 3'UTR (72 bp). A total of 22 repetitive elements were identified at the intronic level, and four of these (L1MC, L1M5, MIRb, and Charlie4z) were identified as being exclusive to buffalo. Comparative analysis between the *FABP6* gene coding sequence and the amino acid sequence with its homologues from other mammalian species showed a percentage of identity varying from 79 to 98% at the DNA coding level and 70 to 96% at the amino acid level. In addition, the alignment of the gene sequence between the Murrah and the Mediterranean breeds revealed 20 potential single nucleotide polymorphisms, which could be candidates for validation in commercial buffalo populations.

Key words: BAC library; FABP6; Murrah; Next generation sequencing

INTRODUCTION

The fatty acid-binding protein family (FABPs) is a group of intracellular lipid chaperones that act on the metabolic, growth, and inflammatory pathways binding hydrophobic ligands with high affinity, such as saturated and unsaturated long-chain fatty acids. A total of nine functional protein-coding FABP genes have been identified in humans and across different mammals, demonstrating strong evolutionary conservation between species (Smathers and Petersen, 2011). Differences in the expression levels of the FABPs genes contribute significantly to the pathogenesis of many metabolic diseases in humans, making this gene an important target for drug development (Furuhashi and Hotamisligil, 2008).

The *FABP6* gene encodes the ileal lipid binding protein, commonly known as gastrotropin, which has been found in the ileum, stomach, small intestine, ovary, placenta, and adrenal gland, showing a low binding affinity for long-chain fatty acids and a high affinity for bile acids (Iseki et al., 1993; Crossman et al., 1994; Fujita et al., 1995; Agellon et al., 2002; Bonné et al., 2003; Smathers and Petersen, 2011). Bile acids are synthesized in the liver from cholesterol and absorbed in the intestine and the ileum in a process named as enterohepatic circulation. The gastrotropin protein mediates the absorption of the bile acids in the intestine and the ileum, playing a significant role in the homeostasis of bile acids and cholesterol of the organism (Håkansson et al., 2002).

Almeida et al. (2011) identified overexpression of the *FABP6* gene in cattle associated with neurodegeneration by prion caused by spongiform encephalopathy, indicating that this gene could be used in diagnosing, treating, and preventing this disease. The *FABP6* gene has also been identified as a candidate gene for resistance to gastrointestinal nematodes in cattle (Li et al., 2011).

The genome analysis of this important livestock species has advanced significantly in the last decade with the availability of genomic tools, such as the whole genome radiation hybrid panel used for genome mapping (Amaral et al., 2007, 2008) and the genomic BAC library (Stafuzza et al., 2012) used for resequencing of target regions of the genome as well as for the development of high-density physical maps. Recently, a genome sequence from the Mediterranean breed has been publicly released in the National Center for Biotechnology Information database (NCBI-<http://www.ncbi.nlm.nih.gov/assembly/67671>).

Studies involving the *FABP6* gene in buffalo (*Bubalus bubalis*) are absent from the literature. Thus, considering the importance of this gene, we characterized its structure in the Murrah breed through next generation sequencing (NGS) of a Murrah BAC clone and performed a comparative analysis of the DNA coding and amino acid sequences against those from other mammals.

MATERIAL AND METHODS

BAC library screening

The buffalo BAC library was screened (3D screening) using PCR with the cattle-derived primers described by Bionaz and Looor (2008). The 3D screening of the library was carried out in three steps. The first step was performed with 17 superpools, where each superpool contained BAC clones from eight 384-well plates. In the second step, eight single pools, representing BAC clones from each of the eight plates in the positive superpool, were screened.

Finally, the pooled row and column BAC clones from the positive plate were screened. The intersection of the row and column identified the location of the positive BAC clone for the *FABP6* gene.

PCRs included: 10 mM Tris-HCl, 1.5 mM MgCl₂, 50 mM KCl, 10 mM dNTPs, 0.2 mM of each primer, 0.5 U AmpliTaq Gold polymerase (Life Technologies™, USA), and 2 μL BAC DNA in a 25-μL reaction volume. The amplification conditions were as follows: initial denaturation at 94°C for 10 min, followed by 35 cycles at 94°C for 30 s (denaturation), 56°C for 30 s (annealing), 72°C for 30 s (extension), and a final extension at 72°C for 7 min. The PCR products were electrophoresed through 2% agarose gels in 1X TBE buffer containing 0.25 μg/mL ethidium bromide, and photographed under UV light.

Isolation of buffalo BAC DNA

The positive clone for the *FABP6* gene was grown in 50 mL of Luria-Bertani medium (Sigma-Aldrich, USA) containing 12.5 μg/mL chloramphenicol. The BAC DNA was purified using the PhasePrep™ BAC DNA Kit (Sigma-Aldrich, USA) as described by the manufacturer. Briefly, clones were harvested by centrifugation at 5000 g for 10 min. The supernatant was removed and the pellet was resuspended in a 250-μL solution containing RNase and 250 μL lysis components, followed by the addition of 250 μL neutralization solution, and then incubation on ice for 5 min and centrifugation at 16,000 g for 5 min at 4°C. The supernatant was transferred to a 2-mL microcentrifuge tube and 450 μL isopropanol was added. The nucleic acids were collected by centrifugation at 16,000 g for 20 min at 4°C. The pellet was washed with 100 μL 70% ethanol and 500 μL elution solution was added.

The endotoxins and other impurities were removed by addition of 100 μL endotoxin removal solution followed by incubation on ice for 5 min and at 37°C for 5 min. The clear upper phase containing the BAC DNA was transferred to another 2-mL microcentrifuge tube and 540 μL DNA precipitation solution was added, followed by centrifugation at 16,000 g for 20 min at 4°C. The supernatant was removed and the pellet was washed with 150 μL 70% ethanol followed by centrifugation at 16,000 g for 20 min at 4°C. The supernatant was discarded and the BAC DNA was eluted in 100 μL TE buffer.

The DNA concentration was determined using a NanoDrop spectrophotometer (Thermo Scientific, USA) and adjusted to 100 ng/μL. An additional PCR with the purified DNA was performed as described earlier to confirm the amplification of the *FABP6* gene.

BAC DNA sequencing and bioinformatics analysis

The BAC DNA sequencing was obtained by NGS using 454-pyrosequencing-GS FLX Titanium chemistry (Roche) performed at the 454 Sequencing Center, Branford, CT, USA. The sequence data was assembled using the GS De Novo Assembler version 2.6 (Roche).

Repetitive DNA elements in the Murrah BAC clone sequence were identified by RepeatMasker (<http://www.repeatmasker.org/>) and the Augustus software was used to predict genes from the masked nucleotide sequence (<http://bioinf.uni-greifswald.de/augustus/>). The predicted gene sequence was aligned against NCBI reference mRNA sequences from *Bos taurus* using BLAST tools (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and the predicted peptide sequence was aligned against non-redundant protein sequences using BLASTp (protein-protein BLAST) to verify homology among other species and to identify the gene.

The coordinates of the exons, introns, 5'UTR, and 3'UTR were obtained by performing an alignment between the Murrah BAC clone sequence from the Murrah breed against the predicted gene sequence available for the Mediterranean breed on NCBI (<http://www.ncbi.nlm.nih.gov/assembly/67671>). In addition, the Splign program (<http://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi>) was used to recognize introns and splice signals.

VISTA plots were generated using wgVISTA alignment (<http://genome.lbl.gov/cgi-bin/WGVistaInput>) with the bovine chromosome 7 genomic scaffold (GenBank accession No. NW_003104049.1) to show the relative nucleotide conservation between buffalo and bovine.

The Clustal Omega software (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to align coding gene sequences and amino acids sequences from the following species *B. bubalis*, *B. taurus*, *Ovis aries*, *Capra hircus*, *Equus caballus*, *Sus scrofa*, *Rattus norvegicus*, *Mus musculus*, and *Homo sapiens*.

RESULTS

The screening of the buffalo BAC library resulted in the identification of one positive clone for the *FABP6* gene. The 454/Roche GS FLX sequencer generated a total of 2930 reads with an average read length of 669 nucleotides arranged in one contig of 28,709 nucleotides providing a 64.29X coverage, N50 value of 28.73 kb, and 99.89% of Q40+. The Murrah BAC clone DNA sequence was deposited in GenBank under the accession No. KJ635885.

The RepeatMasker tool identified a CG content of 48.79% in the Murrah BAC DNA sequence. Statistical analysis revealed a total of 70 interspersed repeats, such as small interspersed nucleotide elements (SINEs), long interspersed nucleotide elements (LINEs), long terminal repeat elements (LTR), and other DNA elements. Approximately 40% of the DNA sequence comprises interspersed repeats (11,671 bp), represented by 39 SINEs (17.99%), 17 LINEs (11.03%), 6 LTR (7.42%), and 8 DNA elements (4.19%). Other repetitive elements include 6 small RNAs (3.27%), 3 satellites (0.66%), and 5 simple tandem repeats (0.89%). The simple tandem repeats were classified as dinucleotide repeats [(TT)_n, (TG)_n, and (AT)_n] and tetranucleotide repeats [(CATC)_n and (TTTC)_n] ranging from 34 to 72 bp in length. The repetitive elements are summarized in Table 1.

Table 1. Repetitive elements identified in the Murrah BAC DNA sequence regarding the number and type of elements, length (bp), and percentage of the total length of each element observed in the DNA sequence.

Type of element	No.	Length occupied (bp)	Murrah BAC DNA sequence (%)
Interspersed repeats	70	11,671	40.62
SINEs	39	5169	17.99
LINEs	17	3168	11.03
LTR elements	6	2131	7.42
DNA elements	8	1203	4.19
Small RNA	6	939	3.27
Satellites	3	192	0.66
Simple repeats	5	256	0.89

The gene prediction by the Augustus software detected the presence of one gene in the BAC DNA sequence from the Murrah breed. The alignment of the predict gene against the NCBI reference mRNA sequence from *B. taurus* revealed 98% identity and 100% coverage with the bovine fatty acid binding protein 6, ileal (*FABP6*) mRNA (GenBank accession No.

NM_001075675.2). The predicted peptide alignment against non-redundant protein sequences from the *B. taurus* database revealed 95% identity and 100% coverage with the *B. taurus* gastrotropin protein (GenBank accession No. NP_001069143.1).

The alignment between the Murrah BAC sequence and the predicted gene for the Mediterranean breed (GenBank accession No. NW_005785538.1) established the coordinates of the *FABP6* gene from nucleotide 18,574 to nucleotide 24,678, with a total length of 6105 bp. The organization of the gene was as following: 5'UTR (93 bp), exon 1 (67 bp), intron 1 (1167 bp), exon 2 (176 bp), intron 2 (1737 bp), exon 3 (90 bp), intron 3 (2649 bp), exon 4 (54 bp), and 3'UTR (72 bp), with an exon/intron size ratio of 1:14.30 and a CG level of 50.64%. The coding sequence contains a total of 387 nucleotides, which encodes the gastrotropin protein with 128 amino acids.

The comparative analysis using the wgVISTA alignment between the *FABP6* gene from buffalo and bovine (chromosome 7 genomic scaffold-GenBank accession No. NW_003104049.1) showed a genomic conservation of 98.1% on the 5'UTR region, 95.5% on exon 1, 96.5% on intron 1, 98.9% on exon 2, 96.7% on intron 2, 95.6% on exon 3, 95.1% on intron 3, 98.1% on exon 4, and 84.3% on 3'UTR (Figure 1).

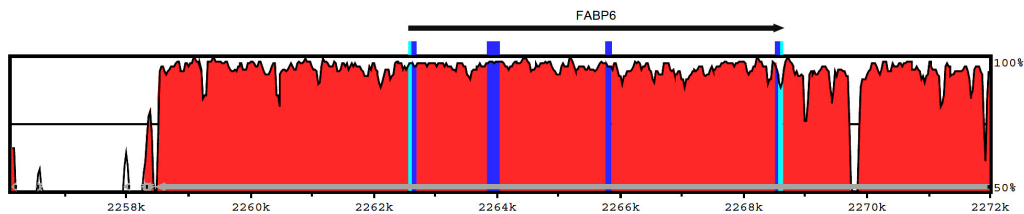


Figure 1. wgVISTA plot obtained from the alignment between the *FABP6* gene from buffalo and bovine (chromosome 7 genomic scaffold - GenBank accession No. NW_003104049.1) showing a genomic conservation (CNS) higher than 95% between both sequences, indicated by red color. The black arrow represents the *FABP6* gene. Exons are indicated by dark blue blocks while the 5'UTR and 3'UTR are indicated by the light blue blocks.

Regarding the presence of repetitive elements in the buffalo *FABP6* gene, the Repeat-Masker tool identified a total of 22 repetitive elements at the intronic level, including 13 SINES, 3 LINES, 2 DNA elements, 2 satellites, and 2 simple tandem repeats, corresponding to 36.11% of the total length of gene (Figure 2). Comparative analysis including other bovids (cattle, sheep, and goat) revealed four buffalo specific repetitive elements, including two LINES (L1MC and L1M5) in intron 2, one SINE (MIRb), and one DNA element (Charlie4z) in intron 3.

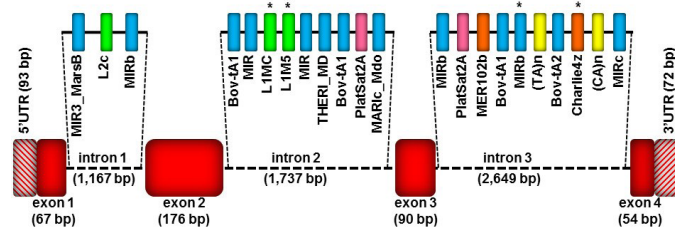


Figure 2. Schematic representation of the buffalo *FABP6* gene showing the repetitive elements found on introns. The repetitive elements are indicated by colored boxes delimited by the size of each intron. SINES are indicated by blue boxes, LINES (green), DNA elements (orange), simple tandem repeats (yellow), and DNA satellites (pink). The asterisks highlight buffalo specific repetitive elements found on intron 2 and 3. The red boxes represent exons 1, 2, 3, and 4 and the hachured boxes represent the 5'UTR and 3'UTR, with their respective sizes in base pairs (bp).

The multiple alignment of the buffalo *FABP6* coding sequences showed 98% identity with cattle (GenBank accession No. NM_001075675.2), 96% with sheep and goat (GenBank accession Nos. XM_004009033.1 and XM_005683245.1, respectively), 84% with pig and horse (GenBank accession Nos. NM_214215.2 and XM_005599253.1, respectively), 81% with human (GenBank accession No. NM_001445.2), and 79% with mouse and rat (GenBank accession Nos. NM_008375.2 and NM_017098.1, respectively), considering 100% coverage (Figure 3).

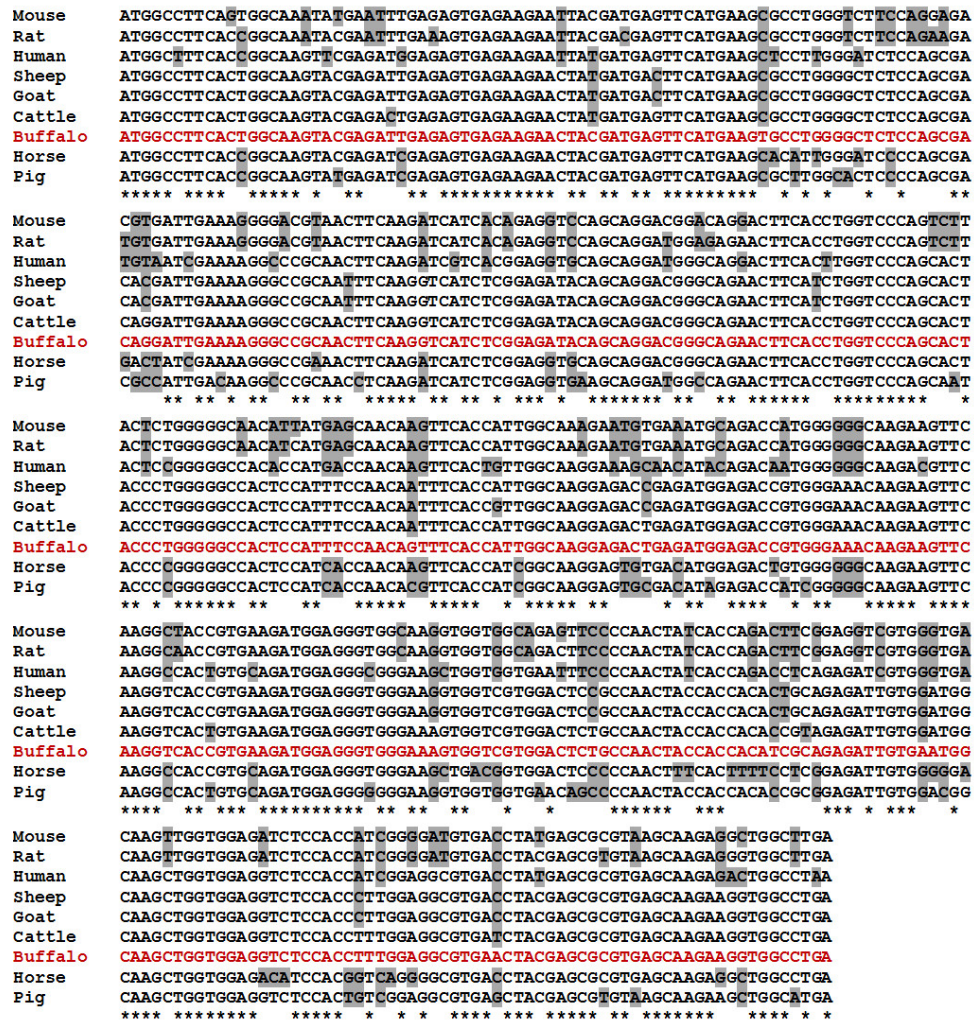


Figure 3. Multispecies alignment using the Clustal Omega software showing the percentage of identity between buffalo *FABP6* coding sequence and mouse (GenBank accession No. NM_008375.2), rat (GenBank accession No. NM_017098.1), human (GenBank accession No. NM_001445.2), sheep (GenBank accession No. XM_004009033.1), goat (GenBank accession No. XM_005683245.1), cattle (GenBank accession No. NM_001075675.2), horse (GenBank accession No. XM_005599253.1), and pig (GenBank accession No. NM_214215.2). Nucleotides substitutions in relation to buffalo *FABP6* are highlighted in gray.

At the protein level, the predicted amino acid sequence of the buffalo gastrotropin revealed 95% identity with cattle (GenBank accession No. XP_004009082.1), 93% with sheep (GenBank accession No. XP_004009082.1), 92% with goat (GenBank accession No. XP_005683302.1), 79% with pig (GenBank accession No. NP_999380.2), 77% with horse (GenBank accession No. XP_005599310.1), 73% with rat (GenBank accession No. NP_058794.1), and 70% with mouse and human (GenBank accession Nos. NP_032401.1 and NP_001436.1, respectively), considering 100% coverage (Figure 4).

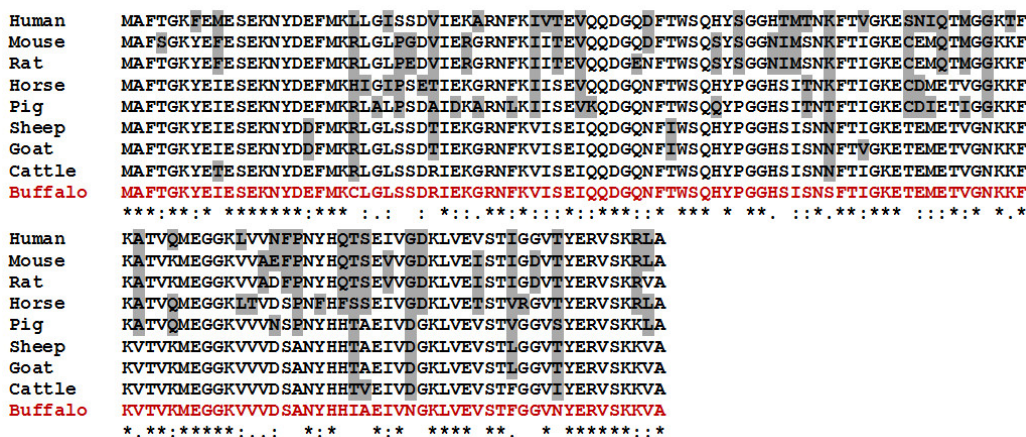


Figure 4. Multispecies alignment using the Clustal Omega software showing the percentage of identity of buffalo gastrotropin amino acid sequence against the protein sequences of human (GenBank accession No. NP_001436.1), mouse (GenBank accession No. NP_032401.1), rat (GenBank accession No. NP_058794.1), horse (GenBank accession No. XP_005599310.1), pig (GenBank accession No. NP_999380.2), sheep (GenBank accession No. XP_004009082.1), goat (GenBank accession No. XP_005683302.1), and cattle (GenBank accession No. NP_001069143.1). Amino acids substitutions in relation to buffalo sequence are highlighted in gray.

The alignment of the *FABP6* gene sequence from the Murrah breed against the predicted gene sequence from the Mediterranean breed (GenBank accession No. NW_005785538.1) revealed 99% identity. Twenty nucleotide substitutions and one deletion were observed in the Murrah breed sequence with the following positions: intron 1 (g.829C>A, g.995G>C, g.1044G>A, and g.1254G>A), intron 2 (g.1897T>A, g.1931C>A, and g.2958delT), intron 3 (g.3472 G>A, g.3705 C>T, g.3709A>G, g.3774C>T, g.4,197A>C, g.4213C>T, g.4237T>C, g.4448T>C, g.4483A>G, g.4934C>A, g.5457A>G, g.5541T>C, and g.5861T>C), and 3'UTR (g.6063T>C).

DISCUSSION

In this study, one BAC clone from the buffalo BAC library was isolated and characterized, showing the complete sequence of the *FABP6* gene in the Murrah breed. The total size of the buffalo *FABP6* gene (6105 bp) showed a similar length to the gene in cattle (6062 bp), sheep (6011 bp), and goat (5783 bp), which was expected considering that they belong to the Bovidae family. Differences in size were more evident when the buffalo gene was compared with the *FABP6* gene from other mammals, such as rat (4670 bp), mouse (5523 bp), horse

(7627 bp), pig (26,800 bp), and human (51,356 bp). Despite differences in the total size of the gene in the mammals outside the Bovidae family, the total size of the coding sequences remained the same (387 bp).

The buffalo *FABP6* gene showed four exclusive repetitive elements located on introns 2 and 3, resulting in differences in the nucleotides when compared to the DNA sequences from cattle, goat, and sheep, contributing to the buffalo gene variability.

The differences observed between coding sequences of buffalo and cattle revealed a total of nine nucleotide substitutions in the buffalo sequence, with two synonymous substitutions and seven nonsynonymous substitutions, resulting in the following amino acids changes: Thr9Ile, Arg21Cys, Asn63Ser, Thr101Ile, Val102Ala, Asp106Asn, and Ile119Asn. These seven amino acid substitutions are responsible for the 95% identity observed between the buffalo and cattle gastrotropin protein.

Comparisons between the Murrah breed gene sequence and the predicted gene sequence for the Mediterranean breed, publicly available at NCBI, showed twenty potential single nucleotide polymorphisms that could be used to uncover the genetic structure of these two buffalo breeds in future population studies.

The comparative analysis of the buffalo *FABP6* gene confirmed the high conservation of its structure when compared with several mammals, providing basic information for future studies of this gene in different buffalo herds involving polymorphisms associated with economically important traits, especially those related to gastrointestinal parasite resistance and metabolic disorders.

ACKNOWLEDGMENTS

Research supported by grants from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil; grant #2011/11889-3) to M.E.J. Amaral-Trusty and a post-doctoral fellowship to NBS (fellowship #2011/02478-0). The Brazilian federal institution CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) funded the master's fellowship to M.M. Borges.

REFERENCES

- Agellon LB, Toth MJ and Thomson AB (2002). Intracellular lipid binding proteins of the small intestine. *Mol. Cell Biochem.* 239: 79-82.
- Almeida LM, Basu U, Williams JL, Moore SS, et al. (2011). Microarray analysis in caudal medulla of cattle orally challenged with bovine spongiform encephalopathy. *Genet. Mol. Res.* 10: 3948-3962.
- Amaral ME, Owens KE, Elliott JS, Fickey C, et al. (2007). Construction of a river buffalo (*Bubalus bubalis*) whole-genome radiation hybrid panel and preliminary RH mapping of chromosomes 3 and 10. *Anim. Genet.* 38: 311-314.
- Amaral ME, Grant JR, Riggs PK, Stafuzza NB, et al. (2008). A first generation whole genome RH map of the river buffalo with comparison to domestic cattle. *BMC Genomics* 9: 631.
- Bionaz M and Looor JJ (2008). ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. *J. Nutr.* 138: 1019-1024.
- Bonné A, Gösele C, den Bieman M, Gillissen G, et al. (2003). Sequencing and chromosomal localization of Fabp6 and an intronless Fabp6 segment in the rat. *Mol. Biol. Rep.* 30: 173-176.
- Crossman MW, Hautft SM and Gordon JI (1994). The mouse ileal lipid-binding protein gene: a model for studying axial patterning during gut morphogenesis. *J. Cell Biol.* 126: 1547-1564.
- Fujita M, Fujii H, Kanda T, Sato E, et al. (1995). Molecular cloning, expression, and characterization of a human intestinal 15-kDa protein. *Eur. J. Biochem.* 233: 406-413.
- Furuhashi M and Hotamisligil GS (2008). Fatty acid-binding proteins: role in metabolic diseases and potential as drug

- targets. *Nat. Rev. Drug Discov.* 7: 489-503.
- Håkansson P, Andersson I, Nyström S, Löfgren L, et al. (2002). Ontogenetic development and spatial distribution of the ileal apical sodium-dependent bile acid transporter and the ileal lipid-binding protein in apoE knockout and C57BL/6 mice. *Scand. J. Gastroenterol.* 37: 1089-1096.
- Iseki S, Amano O, Kanda T, Fujii H, et al. (1993). Expression and localization of intestinal 15 kDa protein in the rat. *Mol. Cell Biochem.* 123: 113-120.
- Li RW, Rinaldi M and Capuco AV (2011). Characterization of the abomasal transcriptome for mechanisms of resistance to gastrointestinal nematodes in cattle. *Vet. Res.* 42: 114.
- Smathers RL and Petersen DR (2011). The human fatty acid-binding protein family: evolutionary divergences and functions. *Hum. Genomics* 5: 170-191.
- Stafuzza NB, Abbey CA, Gill CA, Womack JE, et al. (2012). Construction and preliminary characterization of a river buffalo bacterial artificial chromosome library. *Genet. Mol. Res.* 11: 3013-3019.