



Sequence analysis of the *S1PR1* gene in river buffalo

N.B. Stafuzza^{1,2}, B.C.M. Naressi¹, M.M. Borges¹ and M.E.J. Amaral-Trusty¹

¹Departamento de Biologia, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista Júlio de Mesquita Filho, São José do Rio Preto, SP, Brasil

²Departamento de Ciências Exatas, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista Júlio de Mesquita Filho, Jaboticabal, SP, Brasil

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

Genet. Mol. Res. 15 (1): gmr.15016363

Received February 20, 2015

Accepted October 26, 2015

Published January 29, 2016

DOI <http://dx.doi.org/10.4238/gmr.15016363>

ABSTRACT. Recent developments in methodologies for genomic analyses have enabled a significant advance in understanding of the river buffalo genome. The *S1PR1* gene has been mapped to buffalo chromosome 6 and bovine chromosome 3; this gene is of interest as it is a candidate for marbling in meat, an important economic trait. Here, we performed next generation sequencing in a buffalo BAC DNA clone and obtained a 54.5-kb sequence encompassing the entire buffalo *S1PR1* gene as well as the 27 kb upstream region and the 22 kb downstream region. The gene had a total length of 4716 bp, including three exons and two introns; exons 1 and 2 were classified as non-protein-coding. In comparison with homologues from other species, the structural organization of buffalo *S1PR1* was closest to that of the goat and in both species exon 2 of the gene was non-protein-coding. One hundred and nine repetitive elements were found within the buffalo gene and its boundary regions, with 50 SINE repeats being the most abundant. Alignment of *S1PR1* sequences from the Murrah and Mediterranean breeds revealed two nucleotide substitutions (g.1176C>G

and g.2740T>C), which represent potential SNPs that could be used in further studies of buffalo genetic structure.

Key words: BAC library; *Bubalus bubalis*; Murrah; Pyrosequencing; Shingosine-1-phosphate receptor 1

INTRODUCTION

Sphingosine-1-phosphate receptors (SIPRs) regulate essential cellular processes such as proliferation, migration, cytoskeletal organization, adherens junction assembly, and morphogenesis (Sanchez and Hla, 2004). Although originally known as endothelial differentiation G-protein coupled receptors, the *SIPR* gene family was renamed once sphingosine-1-phosphate protein (S1P) was established as their ligand (Chun et al., 2010).

Five S1PR proteins have been identified in mammals (S1PR1 to S1PR5). S1PR1 is expressed by the majority of innate and adaptive immune cells such as macrophages, eosinophils, dendritic cells, mast cells, natural killer cells, T cells, B cells, and natural killer T cells (Dorsam et al., 2003; Jolly et al., 2004; Roviezzo et al., 2004; Czeloth et al., 2007; Allende et al., 2007; Hughes et al., 2008; Rivera et al., 2008). The other receptors from this family (S1PR2, S1PR3, S1PR4, and S1PR5) are expressed in only a few cell types of the immune system with an overlapping but distinct pattern of expression according to the tissue (Sanchez and Hla, 2004; Rivera et al., 2008).

The river buffalo (*Bubalus bubalis*) plays an important role in the worldwide economy through the production of high-quality milk and meat. Genomic analysis of this species has advanced significantly in the last decade with the development of methodologies such as radiation hybrid panels for whole genome mapping (Amaral et al., 2007; Stafuzza et al., 2009) and genomic BAC libraries (Stafuzza et al., 2012) for region-specific resequencing and characterization of target genes (Stafuzza et al., 2014a,b,c).

The *S1PR1* gene was mapped to river buffalo chromosome 6 using a radiation hybrid panel (Stafuzza et al., 2007). In cattle, quantitative trait locus analysis revealed that the gene is located on chromosome 3 within a region for marbling of meat and is considered a candidate gene for this trait (Yamada et al., 2006). Single nucleotide polymorphisms (SNPs) within the gene similarly show an association with increased levels of meat marbling in Japanese Black beef cattle (Yamada et al., 2009a,b).

Since *S1PR1* might also be an important gene for meat quality in buffalo, we investigated its molecular structure in the Murrah breed, obtained the sequence of the full-length gene and its boundary regions, and searched for repetitive elements that might contribute to buffalo specific genomic signatures. Additionally, the DNA sequence was compared in Murrah and Mediterranean breeds to search for potential SNPs.

MATERIAL AND METHODS

Bacterial artificial chromosome (BAC) library screening

The river buffalo BAC library was three-dimensionally screened by PCR as described by Stafuzza et al. (2012) using the following primers: forward 5'-TGGCCCTCTCAGACCTGTTG-3' and reverse 5'-TGGCGAGGAGACTGAACACG-3' (Amaral et al., 2008).

PCR was performed using a 25- μ L reaction mixture containing 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. The amplification conditions were: initial denaturation at 94°C for 10 min, followed by 35 cycles at 94°C for 30 s (denaturation), 50°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 separated on 2% agarose gels in 1X TBE buffer containing 0.25 mg/mL ethidium bromide and photographed under UV light.

Isolation of BAC DNA and pyrosequencing

The positive clone for *S1PR1* was grown in 50 mL Luria-Bertani medium (Sigma-Aldrich, USA) containing 12.5 mg/mL chloramphenicol. The BAC DNA was purified using the PhasePrepBAC DNA Kit (Sigma-Aldrich) following the manufacturer instructions.

DNA concentration was determined using a NanoDrop spectrophotometer (Thermo Scientific, USA) and adjusted to 100 ng/µL. Amplification of *S1PR1* was confirmed by PCR using the purified DNA.

DNA sequencing and bioinformatic analysis

BAC DNA sequencing was performed by next generation sequencing (NGS) using 454-pyrosequencing - GS FLX Titanium chemistry (Roche) 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/>). 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 by BLASTp (protein-protein BLAST) to verify homology among species and identify the predicted genes.

The coordinates of exons, introns, 5'UTR, and 3'UTR of the gene were obtained by alignment of the BAC DNA sequence from the buffalo 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 identify introns and splice signals.

The Clustal Omega software (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to align coding gene sequences and amino acids sequences from buffalo (*B. bubalis*), cow (*B. taurus*), yak (*Bos mutus*), sheep (*Ovis aries*), goat (*Capra hircus*), horse (*Equus caballus*), human (*Homo sapiens*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), and chicken (*Gallus gallus*).

RESULTS

A clone from the buffalo BAC library was sequenced using the NGS platform at 454/Roche GS FLX. A total of 4014 reads were generated with an average read length of 711 nucleotides arranged in one contig of 54,510 nucleotides providing a 50.54X coverage and 99.86% of Q40+. The BAC DNA sequence was deposited in GenBank under accession No. KJ635883.

The Augustus software predicted one gene in the DNA sequence showing 99% identity and 100% coverage when aligned with bovine sphingosine-1-phosphate receptor 1 (*S1PR1*) mRNA

(GenBank accession No. NM_001013585.4) at the NCBI database. Alignment of the predicted peptide against non-redundant protein sequences from the *B. taurus* database revealed 100% identity and 100% coverage with bovine S1PR1 protein (GenBank accession No. NP_001013603.2).

Alignment of our Murrah breed buffalo DNA sequence against the Mediterranean breed buffalo sequence available at NCBI (GenBank accession No. NW_005784259.1, unplaced genomic scaffold scf7180021616379, UMD_CASPUR_WB_2.0) suggested a total length of 4716 bp for the *S1PR1* gene. This sequence included three exons (exon 1, 366 bp; exon 2, 1115 bp; and exon 3, 1308 bp) and two introns (intron 1, 292 bp and intron 2, 230 bp) and a coding exon/intron size ratio of 1:2.2; the CG level of the sequence was 48.9%. The exon/intron organizations of buffalo *S1PR1* and homologs in other bovids (yak, bovine, sheep, and goat), rodents (mouse and rat), horse, human, and chicken are shown in Figure 1. According to the NCBI database, exon 1 and exon 2 are classified as non-protein-coding exons, and similarly for the first 159 bp of exon 3. Alignment of *S1PR1* in Murrah and Mediterranean breeds revealed 99% identity, with only two nucleotide substitutions in exon 2 (g.1176C>G) and 3 (g.2740T>C); these variants represent potential SNPs that might be of use for studying the genetic structure of buffalo breeds.

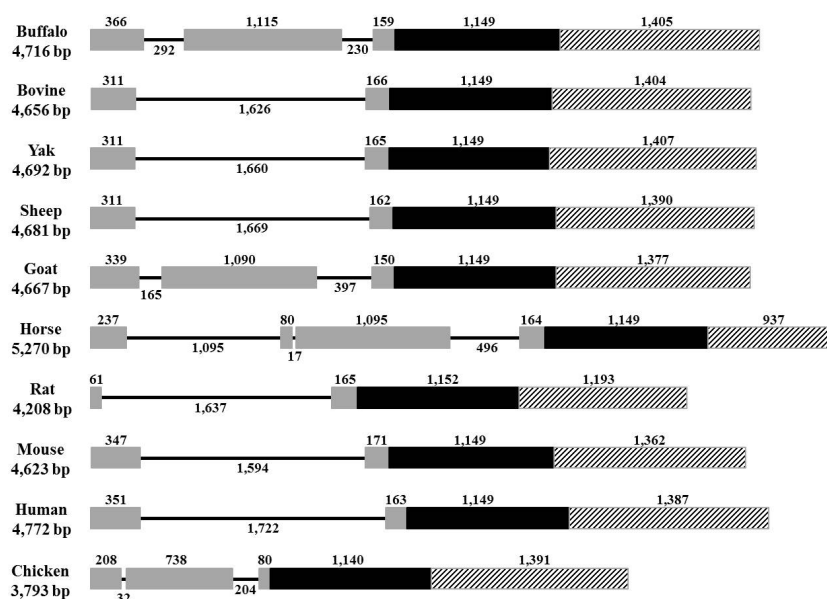


Figure 1. Schematic illustration of the exon/intron organization of *S1PR1* in river buffalo (*Bubalus bubalis*), cow (*Bos taurus*), yak (*Bos mutus*), sheep (*Ovis aries*), goat (*Capra hircus*), horse (*Equus caballus*), human (*Homo sapiens*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), and chicken (*Gallus gallus*). Coding exons are indicated by black boxes while non-protein-coding exons are indicated by gray boxes. Hatched boxes represent the 3'UTR regions. Introns are indicated by solid black lines. Numbers indicate the size of exons, introns and 3'UTRs.

DISCUSSION

Although buffalo *S1PR1* showed a similar size to other bovids, its organization showed greater similarity to that of the goat gene: both species contained a non-protein-coding exon 2. Among the other analyzed vertebrates, the presence of a non-protein-coding exon 2 was only observed in the chicken. *S1PR1* showed most difference in gene organization in the horse with the

presence of three non-protein-coding exons instead of two as in buffalo, goat, and chicken (Figure 1).

Comparative sequence analysis between buffalo and bovine genes revealed 56 indels (insertions/deletions), with deletions concentrated in non-protein-coding exon 2 and insertions in the 3'UTR region. A total of 65 nucleotide substitutions was identified and these occurred mainly in introns. Eighteen substitutions were found in the 5'UTR region (g.3312T>C, g.3360T>G, g.3398G>A, g.3513G>A, g.3520T>C, g.3623C>T, g.3665C>G, g.3807A>G, g.3980C>A, g.3986A>T, g.3989C>A, g.3092C>T, g.4207A>G, g.4210T>G, g.4274T>A, g.4396T>C, g.4462G>A, and g.4586A>T) while only six were observed in the 3'UTR region (g.40G>A, g.68A>C, g.79T>C, g.150G>T, g.1979A>T, and g.2033G>A). Nucleotide substitutions in these regions of *S1PR1* were previously suggested to be associated with the level of marbling in meat of Japanese Black beef cattle (Yamada et al., 2009a; 2009b). Eight synonymous substitutions were also found in the coding region (g.2280C>T, g.2295T>C, g.2508C>G, g.2682T>C, g.2856T>G, g.2913A>G, g.3015T>C, and g.3102T>C). Comparison of the buffalo *S1PR1* coding region with other bovids (goat and sheep) showed 18 nucleotide substitutions (14 synonymous and 4 nonsynonymous); these substitutions resulted in the following amino acids changes in the buffalo protein: Ile173Val, Ser216Ala, Ile219Val, and Ser338Gly (Figure 2).

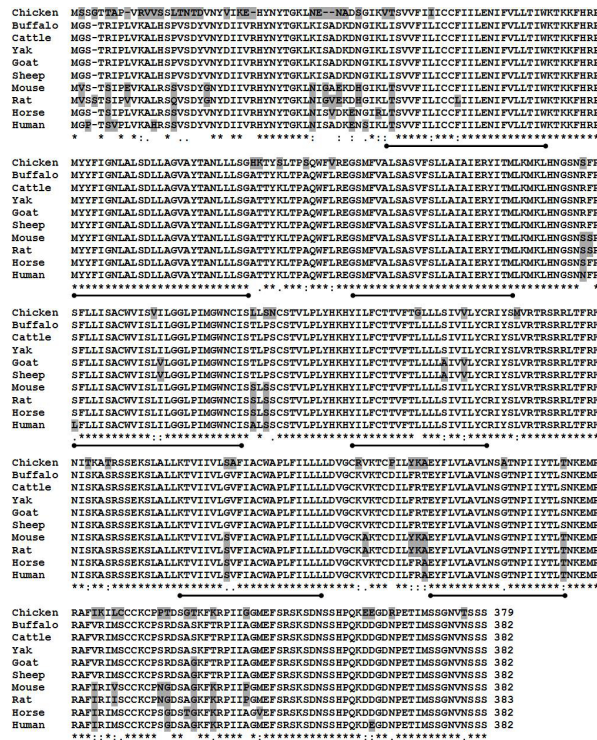


Figure 2. Multispecies amino acid alignment using the Clustal Omega software showing the identity between the buffalo S1PR1 sequence against its counterparts in chicken (GenBank accession No. XP_422305.3), cattle (GenBank accession No. NP_001013603.2), yak (GenBank accession No. XP_005906341.1), goat (GenBank accession No. XP_005678088.1), sheep (GenBank accession No. XP_004002290.1), mouse (GenBank accession No. NP_031927.2), rat (GenBank accession No. NP_058997.1), horse (GenBank accession No. XP_005610319.1), and human (GenBank accession No. NP_001391.2). Amino acids substitutions are highlighted in gray. The black solid lines indicate the seven transmembrane domains of the S1PR1 protein.

The *S1PR1* coding sequence was 1149 nucleotides and encoded a protein of 382 amino acids. This protein is identical in size to that of other mammals, such as bovids (cattle, yak, sheep, and goat), human, and mouse. Chickens have a smaller protein with only 379 amino acids, while the rat has a protein with an additional amino acid (serine at position four) compared to other mammals. Alignment of the amino acid sequences of these various species (Figure 2) indicated high conservation of the transmembrane domain of this protein, as previously described in humans (Rosen et al., 2013). The amino acid sequence of the buffalo *S1PR1* protein showed 100% identity with cattle and yak, 99% with goat and sheep, 95% with horse, 94% with human, 92% with mouse and rat, and 83% with chicken (Figure 2). All alignments showed 99% coverage, except the alignment with amino acid sequences from other bovids that showed 100% coverage.

The RepeatMasker tool was used to analyze the *S1PR1* gene sequence as well as a 27 kb upstream and a 22 kb downstream region of the gene to cover a total of 54.5 kb. The region had a GC content of 41.05% and included 109 repetitive elements encompassing 30.7 kb (56.31% of the region). The rate of repetitive elements in this region was elevated compared to the level of 46.5% in the whole bovine genome (Adelson et al., 2009).

Analysis of the repetitive elements (109) showed 92 interspersed repeats (50 SINEs, 28 LINEs, 7 LTRs, 6 DNA transposons and 1 unclassified), 5 small RNAs, 2 satellites, 3 low complexity DNAs, and 8 simple tandem repeats. A summary of these elements is given in Table 1. The large proportion of interspersed repeats in buffalo DNA sequence was expected as this category also represents 40 to 50% of the total length of the human, mouse and cattle genomes (International Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002; Jurka et al., 2007; Adelson et al., 2009).

Table 1. Repetitive elements in the buffalo BAC DNA sequence. The type of repeat, the total number of repeats, the total length of the repeats, and the relative total size of each element are given.

Type of element	Number	Total length (bp)	Proportion (%) DNA sequence
SINE ¹	50	6,210	11.39
LINE ²	28	18,348	33.65
LTR ³	7	1,098	2.01
DNA Transposon	6	2,731	5.01
Satellite	2	1,094	2.01
Small RNA	5	842	1.54
Simple repeat	8	265	0.49
Low complexity DNA	3	112	0.21

¹Short interspersed sequences. ²Long interspersed sequences. ³Long terminal repeats transposons.

The two simple repeats in buffalo *S1PR1* in the non-protein-coding exon 2 (GT_n) and in intron 2 (TTTC_n) represented 1.46% of the total length of gene. Both repeats were also observed in the *S1PR1* gene of other bovids except for the goat that only had a repeat TTTC_n.

The molecular characterization and the comparative analysis of the buffalo *S1PR1* gene corroborate the expected high conservation of its structure when compared with several vertebrates, and provide basic information for future studies of this gene in different buffalo populations such as investigation of the association of polymorphisms with economically important traits, especially those related with meat quality.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by FAPESP (São Paulo Research Foundation, grant #2011/11889-3) to M.E.J. Amaral-Trusty. CNPq (National Counsel of Technological and Scientific Development) funded the post-doctoral fellowship to N.B. Stafuzza and CAPES (Coordination for the Improvement of Higher Education Personnel) funded the master fellowship to B.C.M. Naressi and M.M. Borges.

REFERENCES

- Adelson DL, Raison JM and Edgar RC (2009). Characterization and distribution of retrotransposons and simple repeats in the bovine genome. *Proc. Natl. Acad. Sci. U. S. A.* 106: 12855-12860. <http://dx.doi.org/10.1073/pnas.0901282106>
- Allende ML, Zhou D, Kalkofen DN, Benhamed S, et al. (2007). S1P1 receptor expression regulates emergence of NKT cells in peripheral tissues. *FASEB J.* 22: 307-315. <http://dx.doi.org/10.1096/fj.07-9087.com>
- 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. <http://dx.doi.org/10.1111/j.1365-2052.2007.01587.x>
- 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. <http://dx.doi.org/10.1186/1471-2164-9-631>
- Chun J, Hla T, Lynch KR, Spiegel S, et al. (2010). International Union of Basic and Clinical Pharmacology. LXXVIII. Lysophospholipid receptor nomenclature. *Pharmacol. Rev.* 62: 579-587. <http://dx.doi.org/10.1124/pr.110.003111>
- Czeloth N, Schippers A, Wagner N, Müller W, et al. (2007). Sphingosine-1 phosphate signaling regulates positioning of dendritic cells within the spleen. *J. Immunol.* 179: 5855-5863. <http://dx.doi.org/10.4049/jimmunol.179.9.5855>
- Dorsam G, Graeler MH, Seroogy C, Kong Y, et al. (2003). Transduction of multiple effects of sphingosine 1-phosphate (S1P) on T cell functions by the S1P1 G protein-coupled receptor. *J. Immunol.* 171: 3500-3507. <http://dx.doi.org/10.4049/jimmunol.171.7.3500>
- Hughes JE, Srinivasan S, Lynch KR, Proia RL, et al. (2008). Sphingosine-1-phosphate induces an antiinflammatory phenotype in macrophages. *Circ. Res.* 102: 950-958. <http://dx.doi.org/10.1161/CIRCRESAHA.107.170779>
- International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860-921. <http://dx.doi.org/10.1038/35057062>
- Jolly PS, Bektas M, Olivera A, Gonzalez-Espinosa C, et al. (2004). Transactivation of sphingosine-1-phosphate receptors by FcepsilonRI triggering is required for normal mast cell degranulation and chemotaxis. *J. Exp. Med.* 199: 959-970. <http://dx.doi.org/10.1084/jem.20030680>
- Jurka J, Kapitonov VV, Kohany O and Jurka MV (2007). Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* 8: 241-259. <http://dx.doi.org/10.1146/annurev.genom.8.080706.092416>
- Rivera J, Proia RL and Olivera A (2008). The alliance of sphingosine-1-phosphate and its receptors in immunity. *Nat. Rev. Immunol.* 8: 753-763. <http://dx.doi.org/10.1038/nri2400>
- Rosen H, Stevens RC, Hanson M, Roberts E, et al. (2013). Sphingosine-1-phosphate and its receptors: structure, signaling, and influence. *Annu. Rev. Biochem.* 82: 637-662. <http://dx.doi.org/10.1146/annurev-biochem-062411-130916>
- Roviezzo F, Del Galdo F, Abbate G, Bucci M, et al. (2004). Human eosinophil chemotaxis and selective in vivo recruitment by sphingosine 1-phosphate. *Proc. Natl. Acad. Sci. U. S. A.* 101: 11170-11175. <http://dx.doi.org/10.1073/pnas.0401439101>
- Sanchez T and Hla T (2004). Structural and functional characteristics of S1P receptors. *J. Cell. Biochem.* 92: 913-922. <http://dx.doi.org/10.1002/jcb.20127>
- Stafuzza NB, Ianella P, Mizziara MN, Agarwala R, et al. (2007). Preliminary radiation hybrid map for river buffalo chromosome 6 and comparison to bovine chromosome 3. *Anim. Genet.* 38: 406-409. <http://dx.doi.org/10.1111/j.1365-2052.2007.01612.x>
- Stafuzza NB, Abbassi H, Grant JR, Rodrigues-Filho EA, et al. (2009). Comparative RH maps of the river buffalo and bovine Y chromosomes. *Cytogenet. Genome Res.* 126: 132-138. <http://dx.doi.org/10.1159/000245912>
- 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. <http://dx.doi.org/10.4238/2012.May.22.6>
- Stafuzza NB, Borges MM and Amaral-Trusty MEJ (2014a). Comparative analysis of the river buffalo somatostatin gene. *Genet. Mol. Res.* 13: 10017-10024. <http://dx.doi.org/10.4238/2014.November.28.6>
- Stafuzza NB, Borges MM and Amaral-Trusty MEJ (2014b). Sequence characterization and comparative analysis of the gastrotropin gene in buffalo (*Bubalus bubalis*). *Genet. Mol. Res.* 13: 10934-10942. <http://dx.doi.org/10.4238/2014.December.19.15>

- Stafuzza NB, Borges MM and Amaral-Trusty MEJ (2014c). Genetic analysis of superoxide dismutase 1 gene in murrh river buffalo. *J. Anim. Plant Sci.* 24: 1869-1875.
- Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562. <http://dx.doi.org/10.1038/nature01262>
- Yamada T, Taniguchi Y, Nishimura S, Yoshioka S, et al. (2006). Radiation hybrid mapping of genes showing intramuscular fat deposition-associated expression changes in bovine musculus longissimus muscle. *Anim. Genet.* 37: 184-185. <http://dx.doi.org/10.1111/j.1365-2052.2006.01426.x>
- Yamada T, Sasaki S, Sukegawa S, Miyake T, et al. (2009a). Novel SNP in 5' flanking region of EDG1 associated with marbling in Japanese Black beef cattle. *Anim. Sci. J.* 80: 486-489. <http://dx.doi.org/10.1111/j.1740-0929.2009.00665.x>
- Yamada T, Itoh M, Nishimura S, Taniguchi Y, et al. (2009b). Association of single nucleotide polymorphisms in the endothelial differentiation sphingolipid G-protein-coupled receptor 1 gene with marbling in Japanese Black beef cattle. *Anim. Genet.* 40: 209-216. <http://dx.doi.org/10.1111/j.1365-2052.2008.01822.x>