Polymorphisms in the ghrelin gene and their associations with milk yield and quality in water buffaloes

Animal Science Department, São Paulo State University (Unesp), Jaboticabal-SP, 14884-900 Brazil

**ABSTRACT**

Ghrelin is a gastrointestinal hormone that acts in releasing growth hormone and influences the body general metabolism. It has been proposed as a candidate gene for traits such as growth, carcass quality, and milk production of livestock because it influences feed intake. In this context, the aim of this study was to verify the existence of polymorphisms in the ghrelin gene and their associations with milk, fat and protein yield, and percentage in water buffaloes (*Bubalus bubalis*). A group of 240 animals was studied. Five primer pairs were used and 11 single nucleotide polymorphisms (SNP) were found in the ghrelin gene by sequencing. The animals were genotyped for 8 SNP by PCR-RFLP. The SNP g.960G>A and g.778C>T were associated with fat yield and the SNP g.905T>C was associated with fat yield and percentage and protein percentage. These SNP are located in intronic regions of DNA and may be in noncoding RNA sites or affect transcriptional efficiency. The ghrelin gene in buffaloes influences milk fat and protein synthesis. The polymorphisms observed can be used as molecular markers to assist selection.

**Key words:** *Bubalus bubalis*, fat, protein, single nucleotide polymorphism

**INTRODUCTION**

Milk production and quality traits in buffaloes have heritability values that vary from moderate to high (Aspilcueta-Borquis et al., 2010a,b,c,d; Tonhati et al., 2011), indicating that it is possible to obtain genetic gains by the selection of genetically superior animals (Seno et al., 2012). To evaluate candidate sires for selection, the progeny test is used, but milk production traits are expressed only in females. This fact increases the cost and the generation interval.

Marker-assisted selection enhances quantitative selection by providing greater accuracy for additive genetic values, besides reducing the generation interval. The ghrelin gene is a candidate to be studied because this hormone induces hunger and is, therefore, related to feed intake. This trait is hard to measure, mainly when animals are raised in a pasture system.

Ghrelin is a gastrointestinal hormone first identified in the stomach of rats (Kojima et al., 1999). It is a peptide composed of 28 amino acids with an octanoic modification in its hydroxyl group on serine-3, which is essential for ghrelin to perform its function of releasing growth hormone (Bednarek et al., 2000). It was first isolated from the oxyntic mucosa of the stomach and is predominantly produced by cells of the gastrointestinal tract. According to Kojima et al. (1999), ghrelin is also produced in small quantities in the central nervous system, kidneys, placenta, and heart. Growth hormone is one of the most important galactopoietic hormones and has an influence on the distribution of nutrients for milk production in lactating cows (Sejrsen et al., 1999).

Roche et al. (2006) conducted a study to determine whether genetic selection for milk production would result in an increase in the plasma concentration of ghrelin in Holstein-Friesian cows. It was observed that cows with a high genetic merit for milk production have higher plasma concentrations of ghrelin and growth hormone. It is associated with an increase in DM consumption, showing that consumption and milk production are closely related.

Sun et al. (2011) detected 11 polymorphisms in the ghrelin gene in noncoding regions in 5 Chinese cattle breeds through the PCR-single-strand conformation polymorphism (SSCP) technique. The authors reported that in the ghrelin gene, no significant associations were revealed between any variant sites and BW, ADG, body sizes for different growth periods (6, 12, 18, and 24 mo old), or milk yield at 305 d, milk protein rate, and milk fat percentage. Kowalewska-Luczak et al. (2011) analyzed the ghrelin gene in 169 Holstein-Friesian cows and found a G/A polymorphism on intron 3 of the gene. The authors analyzed the association of this polymor-
phism with milk-production traits, but no significant
differences between the genotypes were described.

The human ghrelin gene is composed of 5 exons and
4 introns. It has 2 different transcriptional initiation
sites, resulting in 2 different mRNA (Sato et al., 2012).
The bovine ghrelin gene is located on chromosome 22
(BTA22) and is also composed of 5 exons and 4 introns,
but with only 1 mRNA (Colinet et al., 2009; Sato et
al., 2012).

The aim of this work was to verify the existence of
polymorphisms in the ghrelin gene of the water buff-
falo (*Bubalus bubalis*) and its associations with milk,
fat, and protein yields and fat and protein percentages
in the milk. It is known that the ghrelin gene in buff-
faloes is characterized (GenBank: EF583468.1; http://
www.ncbi.nlm.nih.gov/genbank/) and that is expressed
(GenBank: EU604028.1), but no evidence exists of vari-
ants in the gene or association with production traits.

### MATERIALS AND METHODS

#### Animals

The study was conducted with 240 water buffaloes
from a herd belonging to a farm located in Dourado,
São Paulo, Brazil. This farm integrates the dairy buff-
falo milk-recording program of São Paulo State Univer-
sity (Unesp), Jaboticabal-SP, Brazil. The animals were
randomly chosen from the herd, but all of them had to
be in a contemporary group (previous analyses were
done to only genotype animals within a contemporary
group).

#### DNA Extraction and PCR

The DNA was extracted from hair follicle samples
by the phenol-chloroform-isoamyl alcohol method.
Five primer pairs were designed (Table 1) based on
the sequence related to *Bubalus bubalis* of the ghrelin
gene deposited in GenBank database, under accession
code EF583468.1. This sequence has 3,995 bp and 4
exons and 3 introns. The pairs were designed by using
the Primer3 program (available at http://frodo.wi.mit.
edu/prim3/).

The amplification reactions had a final volume of 25
μL, containing 2 μL of DNA (100 ng), 1.5 μL of each
primer (15 pM), 12.5 μL of GoTaq Colorless Master
Mix 1× Taq DNA polymerase and 7.5 μL of nuclease-
free water. The amplification cycles were performed in
an Eppendorf Mastercycler Gradient 5331 thermocycler
(Eppendorf AG, Hamburg, Germany) under the follow-
ing program: denaturing at 94°C for 5 min, followed by
39 denaturing cycles at 94°C for 45 s, specific annealing
for each primer pairs for 45 s, extension at 72°C for 45
s, and final extension at 72°C for 5 min. After amplifi-
cation, a 3-μL aliquot of each sample was diluted with
2 μL of loading buffer (4 μL/mL of GelRed; Uniscience
do Brasil Ind., São Paulo, Brazil) and submitted to
1.5% agarose gel electrophoresis.

### Table 1. Primer pairs and region and size of the amplified product and annealing temperature

<table>
<thead>
<tr>
<th>Primer pair number</th>
<th>Primer sequence</th>
<th>Region of the amplified fragment</th>
<th>Size of the amplified fragment (bp)</th>
<th>Primer annealing temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F: GCCATGCCCCGCCCCGCTGGAC R: TCTGCACCTTCCGCTGATCTTC</td>
<td>Exon 2, intron 2 and part of exon 3</td>
<td>408</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>F: AGCACCTCCTCTTCCCTGAGGA R: TTGCCTGTTGATGCGACGAAACTA</td>
<td>Exon 3 and part of intron 3</td>
<td>281</td>
<td>63.4</td>
</tr>
<tr>
<td>3</td>
<td>F: TGCTTGCACGGTGTGTTCTCTTAC R: AGAATCTGACCCGCTCATCAGT</td>
<td>Part of intron 3</td>
<td>402</td>
<td>63.6</td>
</tr>
<tr>
<td>4</td>
<td>F: TGACGTCTTCGGAGACTTGAAAGGT R: ATGGATGTGATCTGGCAAAGGACT</td>
<td>Part of intron 3 and exon 4</td>
<td>444</td>
<td>60.2</td>
</tr>
<tr>
<td>5</td>
<td>F: TGTTGAGCCCTGCCCATGAGTAACAA R: TCATCAGCCAGGGTTCTGAATGC</td>
<td>Intron 4 and exon 5</td>
<td>448</td>
<td>59.5</td>
</tr>
</tbody>
</table>

1F = forward; R = reverse.

#### SSCP and Sequencing

For the SSCP, 4 μL of the PCR product was used
along with 4 μL of SSCP buffer (95% formamide, 20
mM EDTA, 0.05% bromophenol blue, and 0.05% xy-
lene cyanol), for a final volume of 8 μL. A total of 100
samples for each primer pairs were used in the analyses.
The samples were denatured at 95°C for 5 min, followed by
39 denaturing cycles at 94°C for 45 s, specific annealing
for each primer pairs for 45 s, extension at 72°C for 45
s, and final extension at 72°C for 5 min. After amplifi-
cation, a 3-μL aliquot of each sample was diluted with
2 μL of loading buffer (4 μL/mL of GelRed; Uniscience
do Brasil Ind., São Paulo, Brazil) and submitted to
1.5% agarose gel electrophoresis.
submitted to purification according to the protocol recommended for the Wizard SV Gel and PCR Clean-Up System kit (Promega Corp., Madison, WI). The PCR product was sequenced for both primers (forward and reverse) by the dideoxynucleotide chain termination reaction, using an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems Inc., Foster City, CA) and an ABI 3730 XL automatic sequencer (Applied Biosystems Inc.). The sequences of the ghrelin fragments analyzed were aligned by the CodonCode multiple sequence alignment program (http://www.codoncode.com/aligner/download.htm), with the sequences published in the GenBank database taken as reference to identify the polymorphisms.

**RFLP**

The restriction enzymes were chosen by using the NEBcutter 2.0 program (available at http://tools.neb.com/NEBcutter2/). For the RFLP reactions, we used 2.5 μL of the PCR product, 5 U of the specific restriction enzyme for each polymorphism, and one-tenth of the specific buffer for each enzyme in a final volume of 7.5 μL. Then, the samples were diluted with 2 μL of loading buffer (4 μL/mL of GelRed) and submitted to gel electrophoresis (3.0%) to evaluate the migration patterns of the samples. The restriction enzymes used in the PCR-RFLP analysis are shown in Table 2.

**Statistical Analysis**

The linkage disequilibrium ($r^2$) was estimated using the Plink computational program (available at http://pngu.mgh.harvard.edu/~purcell/plink/) to find out which SNP were inherited together. Considering 2 loci with 2 alleles for each locus (A1/A2 and B1/B2), the following formula applies:

$$r^2 = D^2/[\text{freq}(A1) \times \text{freq}(A2) \times \text{freq}(B1) \times \text{freq}(B2)]$$

(Hill and Robertson, 1966), where $D = \text{freq}(A1_B1) \times \text{freq}(A2_B2) - \text{freq}(A1_B2) \times \text{freq}(A2_B1)$ (Hill, 1981).

The buffaloes of the study were all from first lactations; they were daughters of 29 sires, with a minimum of 5 daughters per sire and a maximum of 20 daughters per sire. The milk recording was done monthly. The studied traits milk, fat, and protein yield were accumulated and truncated at 305 d. Milk fat and protein percentage values were obtained by averaging the monthly test-day records per lactation.

Analyses of variance through the mixed process of the SAS package (SAS Institute, 2000) were done. The contemporary group was considered to be the combination of year and calving season. The averages of milk, fat and protein yield, and fat and protein percentages were compared using the Tukey test ($P < 0.05$). To check for associations between the SNP and the above traits, the following statistical model was used:

$$Y_{ijkl} = \mu + GC_i + S_j + M_k + \beta_1(I_{ijkl} - \bar{I}) + \beta_2(I_{ijkl} - \bar{I}) + e_{ijkl}$$

where $Y_{ijkl}$ is milk yield and fat and protein yield/percentages; $\mu$ is average of the trait in the population; $GC_i$ is fixed effect of contemporary group; $S_j$ is random effect of the sire, with mean 0 and variance $\sigma^2_S$; $M_k$ is fixed effect of the SNP; $\beta_1$ is linear regression coefficient of trait $Y_{ijkl}$ in relation to buffalo age at testing; $\beta_2$ is quadratic regression coefficient of trait $Y_{ijkl}$ in relation to buffalo age at testing; $I_{ijkl}$ is buffalo age at testing; $\bar{I}$ is mean age of the buffalo at testing; and $e_{ijkl}$ is residual random effect associated with trait $Y_{ijkl}$, with mean 0 and variance $\sigma^2_e$.

**RESULTS AND DISCUSSION**

By the SSCP technique, the primer pairs 1, 2, 3, and 5 presented different migration patterns. A sequencing of the amplified fragments which had different migration patterns (by SSCP) was done and 11 polymorphisms were detected (Table 3). The sequences and polymorphisms found were submitted to the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) by the accession number JQ647511.1.

Despite 11 SNP being detected, only 8 of them were genotyped for all the population by PCR-RFLP. It was impossible to genotype the animals for 3 of the SNP identified (g.256G>C, g.986C>T, and g.1763C>T) using PCR-RFLP. The polymorphism g.380T>C is found in the gene’s coding region. Therefore, substitution of the ATG codon by the ACG codon occurred (i.e., there was substitution of synthesis of the amino acid methionine by

<table>
<thead>
<tr>
<th>SNP</th>
<th>Primer pair</th>
<th>Restriction enzyme used for genotyping</th>
</tr>
</thead>
<tbody>
<tr>
<td>g.122A&gt;G</td>
<td>1</td>
<td>BsoDI</td>
</tr>
<tr>
<td>g.380T&gt;C</td>
<td>2</td>
<td>BcrI</td>
</tr>
<tr>
<td>g.778C&gt;T</td>
<td>3</td>
<td>PciI</td>
</tr>
<tr>
<td>g.808T&gt;C</td>
<td></td>
<td>NcoI</td>
</tr>
<tr>
<td>g.905T&gt;C</td>
<td></td>
<td>Tsp451I</td>
</tr>
<tr>
<td>g.906G&gt;A</td>
<td></td>
<td>MpiI</td>
</tr>
<tr>
<td>g.1065T&gt;C</td>
<td>5</td>
<td>HindI</td>
</tr>
<tr>
<td>g.1666A&gt;G</td>
<td></td>
<td>Tsp451I</td>
</tr>
</tbody>
</table>
the amino acid threonine). The others were in intron regions. These SNP have not been previously described in the literature.

Table 4 presents the estimated pairwise $r^2$ values between the 8 SNP genotyped in the population. When $r^2$ is greater than 0.33, the linkage disequilibrium is considered to be strong (Ardlie et al., 2002). These values here varied from 0.010 to 0.806. The SNP g.960G>A and g.778C>T had $r^2$ values lower than 0.33 with all the other SNP and between each other, meaning that they were in linkage equilibrium and were inherited separately. The SNP g.808T>C, g.905T>C, g.1665T>C, and g.1666A>G presented values higher than 0.33 among one another, meaning that these SNP were in linkage disequilibrium and were inherited together. Finally, SNP g.122A>G and g.380T>C presented high $r^2$ values with some SNP of the previous group, but not with all of them. Despite this, because the $r^2$ values were near 0.33, we considered them all to be in the same group, to obtain a more plausible biological explanation. According to these results, 3 groups of SNP exist for calculation of association analyses: group 1, containing SNP g.960G>A; group 2, composed of SNP g.122A>G, g.380T>C, g.808T>C, g.905T>C, g.1665T>C, g.1666A>G; and group 3, containing g.778C>T.

Table 4. Estimated pairwise linkage disequilibrium ($r^2$) values for the 8 SNP found on the ghrelin gene

<table>
<thead>
<tr>
<th>SNP</th>
<th>g.122A&gt;G</th>
<th>g.380G&gt;C</th>
<th>g.960G&gt;A</th>
<th>g.808T&gt;C</th>
<th>g.778C&gt;T</th>
<th>g.905T&gt;C</th>
<th>g.1665T&gt;C</th>
</tr>
</thead>
<tbody>
<tr>
<td>g.122A&gt;G</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>g.380G&gt;C</td>
<td>0.296</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>g.960G&gt;A</td>
<td>0.010</td>
<td>0.013</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>g.808T&gt;C</td>
<td>0.447</td>
<td>0.592</td>
<td>0.041</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>g.778C&gt;T</td>
<td>0.083</td>
<td>0.156</td>
<td>0.034</td>
<td>0.097</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>g.905T&gt;C</td>
<td>0.503</td>
<td>0.635</td>
<td>0.050</td>
<td>0.806</td>
<td>0.073</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>g.1665T&gt;C</td>
<td>0.275</td>
<td>0.323</td>
<td>0.043</td>
<td>0.544</td>
<td>0.096</td>
<td>0.463</td>
<td>—</td>
</tr>
<tr>
<td>g.1666A&gt;G</td>
<td>0.511</td>
<td>0.613</td>
<td>0.031</td>
<td>0.671</td>
<td>0.060</td>
<td>0.757</td>
<td>0.468</td>
</tr>
</tbody>
</table>

In the association analyses, for group 2, SNP 1576T/C was chosen for the calculations because of its better distribution of genotypic frequencies, which were $f(TT) = 0.23$, $f(TC) = 0.51$, and $f(CC) = 0.26$. Table 5 below shows the correlations of these SNP with milk, fat, and protein yield, and fat and protein percentage.

As can be seen in Table 5, g.960G>A had an association with fat yield ($P < 0.05$). The AA genotype presented the highest average production for this trait.

The SNP g.905T>C showed an association with fat yield and fat and protein percentage ($P < 0.05$). The TT genotype presented the highest fat yield. For the fat and protein percentage traits, the genotypes that presented the highest averages were TT/TC.

The SNP g.778C>T showed an association with fat yield. The genotypes CC and CT did not have significant differences between each other and had the highest averages for the fat yield trait.

The AA genotype of g.960G>A, the TT and TC genotypes of SNP g.905T>C, and the CC genotype of SNP g.778C>T were associated with a higher average fat yield, and (or) fat percentage and (or) protein percentage, but without a smaller average milk yield. This indicates that selection can be done for these genotypes to increase the desired traits without altering milk yield.

An ANOVA was also performed on the genotypes for g.380T>C, the only one found in an exonic region. However, the results were not significant at $P < 0.05$.

Reports of polymorphisms on the ghrelin gene are scarce in the literature, as are works on the association between ghrelin and milk production traits. These are the first SNP and the first associations reported for buffaloes. In cattle, Zhang et al. (2012) studying the promoter of ghrelin, found some SNP that influence ischium width and indicated them as possible markers for growth. Sun et al. (2011) found 11 SNP in the ghrelin gene, but no association was found with growth and milk production traits.

According to the results of the current study, the SNP that show association with the traits of interest...
Table 5. Mean milk, fat and protein yields, and fat and protein percentages and nominal P-values for group 1, formed by SNP g.960G>A; for group 2, formed by g.905T>C; and for group 3, formed by 778C>T1

<table>
<thead>
<tr>
<th>Item</th>
<th>MY (kg)</th>
<th>FY (kg)</th>
<th>PY (kg)</th>
<th>%F</th>
<th>%P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA (g.960G&gt;A)</td>
<td>1,592.49</td>
<td>111.96</td>
<td>67.79</td>
<td>7.14</td>
<td>4.32</td>
</tr>
<tr>
<td>AG (g.960G&gt;A)</td>
<td>1,527.22</td>
<td>105.56</td>
<td>65.99</td>
<td>7.10</td>
<td>4.32</td>
</tr>
<tr>
<td>GG (g.960G&gt;A)</td>
<td>1,525.43</td>
<td>101.59</td>
<td>65.76</td>
<td>6.93</td>
<td>4.34</td>
</tr>
<tr>
<td>Genotype 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT (g.905T&gt;C)</td>
<td>1,545.71</td>
<td>107.84</td>
<td>66.81</td>
<td>7.18b</td>
<td>4.34b</td>
</tr>
<tr>
<td>TC (g.905T&gt;C)</td>
<td>1,524.85</td>
<td>102.94</td>
<td>65.93</td>
<td>6.99ab</td>
<td>4.35b</td>
</tr>
<tr>
<td>CC (g.905T&gt;C)</td>
<td>1,545.15</td>
<td>102.18</td>
<td>65.79</td>
<td>6.87a</td>
<td>4.29a</td>
</tr>
<tr>
<td>Genotype 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC (g.778C&gt;T)</td>
<td>1,537.65</td>
<td>104.36</td>
<td>66.37</td>
<td>7.02</td>
<td>4.34</td>
</tr>
<tr>
<td>CT (g.778C&gt;T)</td>
<td>1,522.47</td>
<td>102.01</td>
<td>65.62</td>
<td>6.97</td>
<td>4.33</td>
</tr>
<tr>
<td>TT (g.778C&gt;T)</td>
<td>1,447.42</td>
<td>88.06</td>
<td>60.79</td>
<td>6.43</td>
<td>4.23</td>
</tr>
</tbody>
</table>

*Means within a column followed the same superscript letter are not significantly different (Tukey test, P > 0.05).
1MY = milk yield; FY = fat yield; PT = protein yield; %F = fat percentage; %P = protein percentage.
2MY (P = 0.17); FY (P = 0.0068); PY (P = 0.50); %F (P = 0.11); %P (P = 0.83).
3MY (P = 0.34); FY (P = 0.021); PY (P = 0.52); %F (P = 0.018); %P (P = 0.040).
4MY (P = 0.21); FY (P = 0.016); PY (P = 0.083); %F (P = 0.15); %P (P = 0.39).

are located in noncoding regions of DNA (i.e., they are located on the introns). Greenwood and Kelsoe (2003) and Le Hir et al. (2003) have determined that introns can affect transcriptional efficiency and Nakaya et al. (2007) suggested that many of the introns produce non-coding RNA of proteins that help regulate the majority of protein-coding genes. These results indicate that ghrelin introns in buffaloes may be important in the process described above and future studies related to them to better understand it may be done.

Leptin is the hormone with contrary action to ghrelin (i.e., the hormone for satiety). Tanpure et al. (2012), studying this gene in buffaloes, found an SNP that presented a significant association with the fat percentage trait at 150 d. This SNP was in the intronic region. This result is similar to those found here, where the SNP of intronic regions presented correlations with milk production traits. This suggests that the intronic variants in hormones related to consumption and feed intake seem to be important.

Based on these results, an association exists between these SNP identified and fat and protein yield/percentage. Therefore, future studies are necessary to verify the influence of other gene regions on milk quality, as well as studies on other genes.

CONCLUSIONS

The results obtained in the present study permit the inference that 11 SNP were detected on the ghrelin gene of water buffaloes. Three SNP (g.960G>A, g.778C>T, and g.905T>C) were associated with fat production and (or) fat and protein percentages (P < 0.05), indicating an association with milk quality. These SNP are candidates to be used as molecular markers in buffaloes, especially because an assisted selection can increase fat and protein percentage without reducing milk yield. Further studies are important to characterize other regions of interest on the same gene as well as to understand the genetic processes behind these SNP.

REFERENCES


