



## Hot topic: Gene duplication at the $\alpha$ -lactalbumin locus: Finding the evidence in water buffalo (*Bubalus bubalus* L.)

R. Rullo,\* A. Di Luccia,† L. Chianese,‡ and E. Pieragostini†<sup>1</sup>

\*CNR–ISPAAM, Via Argine, 1085–80147 Ponticelli-Napoli, Italy

†Department PROGESA, University of Bari, 70126 Bari, Italy

‡Department of Food Science, University of Naples Federico II, 80055 Portici (NA), Italy

### ABSTRACT

Studies on milk proteins revealed that a qualitative and quantitative polymorphism may often be found regarding  $\alpha$ -lactalbumin ( $\alpha$ -LA). In mammals, a similar phenomenon was widely documented in the  $\alpha$ -globin system as the result of a gene duplication. The presence of several differently expressed  $\alpha$ -lactalbumin gene (LALBA) products suggests that the mechanism underlying this phenomenon may involve nonallelic genes. To check this hypothesis, an experiment was set up to investigate the LALBA gene arrangement of a water buffalo exhibiting an  $\alpha$ -LA phenotype characterized by a double-band pattern on PAGE isoelectric, focusing analysis of milk protein. In particular, the relative amount of protein inferred from the different intensity of the bands was consistent with a gene duplication. Thus, leukocyte DNA was extracted from a blood sample of the buffalo and amplified with 4 primers (2 RV-IVFW for PCR and 4 FW-IRV for nested PCR). The intergenic segments of the assumed duplicated gene were then amplified with 2 different PCR protocols. First, the segment limited by the third exon in the upstream gene and the second exon in the downstream gene was amplified by simple PCR, which gave aspecific results. Second, this PCR product was subjected to nested PCR, amplifying the segment limited by the fourth exon in the upstream gene and the first exon in the downstream gene, yielding an amplified nucleotide fragment of about 6,200 bp. Blood samples from an additional 15 buffalos were then analyzed in the same manner. The results obtained from the new samples confirmed the presence of an amplified nucleotide fragment of about 6,200 bp in most of them, though they all were characterized by an  $\alpha$ -LA monomorphic phenotype. A couple of 6,200-bp fragments obtained were purified, cloned in pGEM-T easy vector system

(Promega, Madison, WI) and sequenced. The sequence of the large DNA segments, containing the intergenic portion, was aligned with the LALBA gene (accession number AF194373; <http://www.ncbi.nlm.nih.gov/Datatabase/index.html>). They both were found to coincide with the portion containing exon 4 and the untranslated region at the 3' end of the upstream gene and with the portion containing exon 1 and the untranslated region at the 5' end of the downstream gene. These results confirm the hypothesis that a tandemly repeated copy of the LALBA gene is present in water buffalo.

**Key words:**  $\alpha$ -lactalbumin, duplicated gene, nested polymerase chain reaction

According to Ohno's hypothesis, gene duplication plays a major role in evolution (Ohno, 1970). The 2 paralog genes that exist after a gene duplication event usually code for proteins with a different function or structure. A prototype example may be found in the molecular evolution from lysozyme, an antipathogen, to  $\alpha$ -LA, a milk protein, with the duplication occurring perhaps as early as 300 million years ago (Dayhoff, 1976). The acquisition of  $\alpha$ -LA is key for the presence of lactose in milk and also of most milk oligosaccharides (including those of humans) that contain a lactose unit at their reducing end.

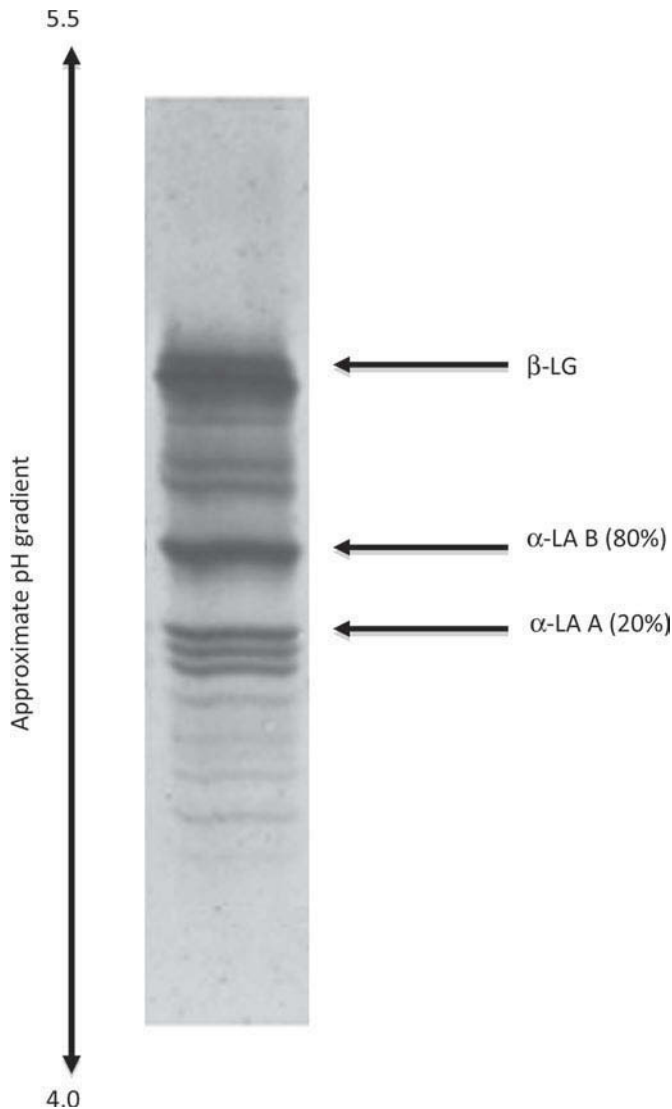
$\alpha$ -Lactalbumin is a major whey protein present in mammalian milk and colostrum. It is a small, acidic protein (14.4 kDa) that has the unusual property of acting as a modifier of galactosyltransferase action to promote lactose synthesis (McKenzie, 1996).  $\alpha$ -Lactalbumin and  $\beta$ 1,4-galactosyltransferase form the enzyme lactose synthetase, which is responsible for lactose synthesis. Lactose regulates osmolarity in milk and draws water into the secretory vesicles of mammary epithelial cells. Because of its association with lactose production,  $\alpha$ -LA is thought to play a role in regulating milk volume (Miglior et al., 2007).

In water buffalo as well as in cattle,  $\alpha$ -LA is characterized by a sequence of 123 AA (Addeo et al., 1976). Two variants, A and B, have been found so far, differing for a nonconservative point mutation (Asn 45→Asp) but

Received August 8, 2009.

Accepted January 22, 2010.

<sup>1</sup>Corresponding author: pierelis@agr.uniba.it



**Figure 1.** Polyacrylamide gel isoelectric of whey protein from individual river buffalo milk containing  $\alpha$ -LA A and  $\alpha$ -LA B, focusing on the pH range 4.0 to 5.5.

occasionally also for their relative proportion (Chianese et al., 2004). Analogously, in Podolian cattle, Pieragostini et al. (2000) observed a different expression in a subject that was supposedly AB heterozygous at the  $\alpha$ -LA locus (*LALBA*), with the A and B variants accounting for 40 and 60% of total protein, respectively.

Bovine and ovine (pseudo)genes homologous to the  $\alpha$ -LA-encoding gene were described by Soulier et al. (1989), who suggested the presence of a family of  $\alpha$ -LA-related sequences in the bovine genome. Moreover, several regulatory elements have been identified in the  $\alpha$ -LA 5' flanking region of various species (Lubon and Henninghausen, 1988; Bleck and Bremel, 1993), providing evidence of *LALBA* gene expression control and

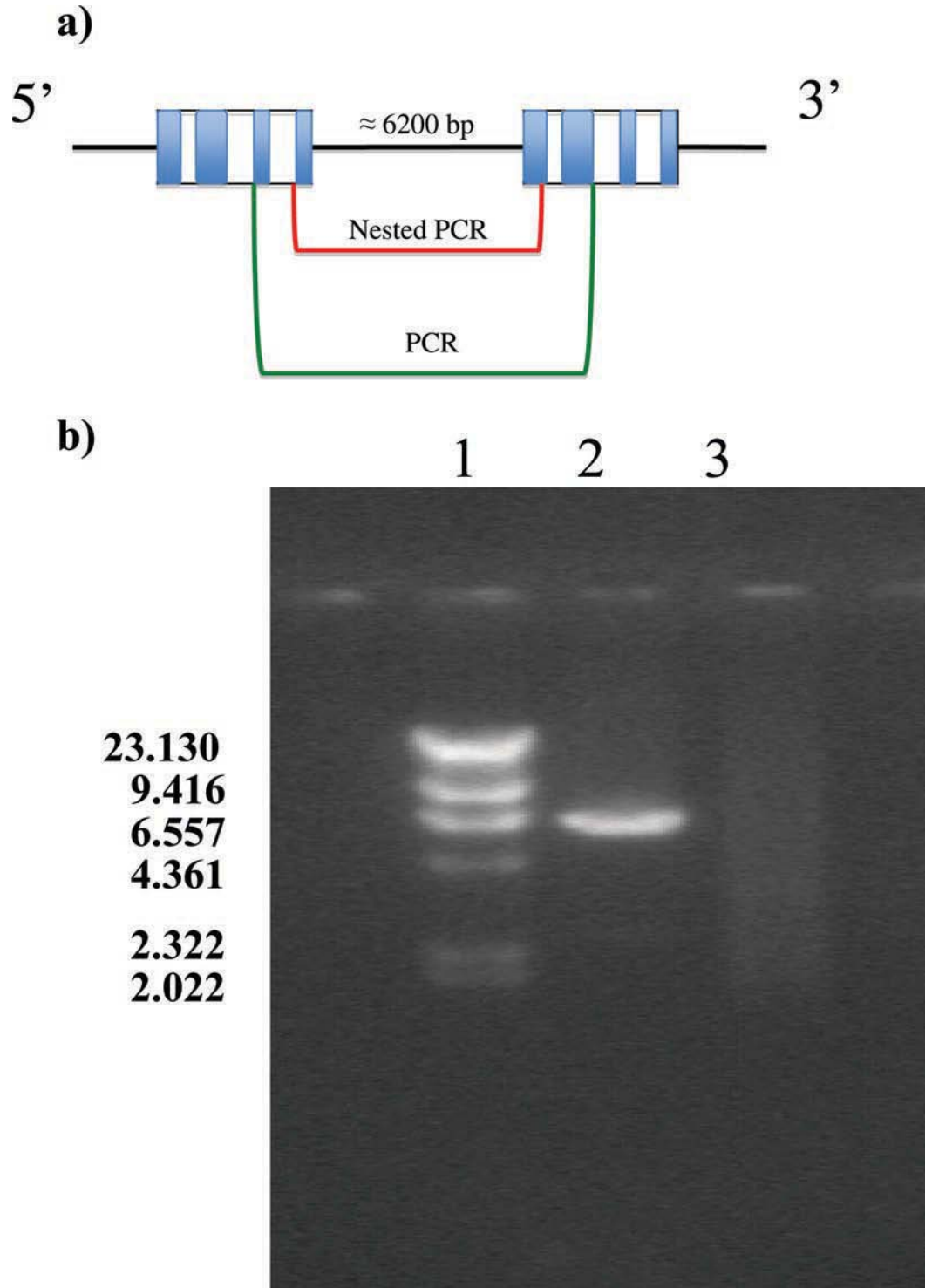
identifying areas that may be associated with quantitative differences in  $\alpha$ -LA production. In mammals, this phenomenon was observed in the  $\alpha$ -globin system as a result of a gene duplication. At the protein level,  $\alpha$ -globin gene expression progressively decreases from the 5' to the 3' end; in particular, as observed in horses, sheep, cattle, and goats, the percent gene efficiencies in the duplicated gene haplotypes are approximately 30 at the 5' end and 20 at the 3' end (Clegg et al., 1984; Vestri et al., 1991; Scaloni et al., 1998; Pieragostini et al., 2005).

In light of all the above considerations and the fact that the  $\alpha$ -LA quantitative phenomenon reported for buffalos is similar to the phenomenon observed in the  $\alpha$ -globin systems, we were encouraged to check the hypothesis of gene duplication of the buffalo *LALBA*. Thus, after a population survey in some buffalo farms of medium size, an adult animal exhibiting  $\alpha$ -LA qualitative and quantitative polymorphism was identified, according to the methodological procedures described by Chianese et al. (2004). Densitometric evaluation of PAGE isoelectric focusing analysis of the milk protein showed that the bands representing  $\alpha$ -LA A and  $\alpha$ -LA B displayed relative intensities of 20 and 80%, respectively (Figure 1). This protein ratio is remarkably similar to that of the A/BI and LH/LL  $\alpha$  gene duplicated arrangements recorded in the horse (Clegg, et al., 1984) and sheep (Pieragostini et al., 2003), respectively. Based on these findings and on the relative amount of protein inferred from the different band intensity, the genotype of the sampled animal was assumed to comprise 3 non-allelic genes directing the synthesis of the same  $\alpha$ -LA B and 1 encoding for  $\alpha$ -LA A.

Genomic DNA was thus isolated from the leukocytes as described by Goossens and Kan (1981). Polymerase chain reaction was performed using approximately 250 ng of genomic DNA and the reaction conditions were optimized using a Gene Amp XL PCR kit (Applied Biosystems, Foster City, CA) with the following final conditions:  $1 \times$  rTth XL PCR buffer, 1.2 mM MgOAc, 0.8  $\mu$ M dNTP (0.2  $\mu$ M each, 2 U of rTth XL DNA polymerase), and 0.3  $\mu$ M of each primer synthesized on the nucleotide sequences of the  $\alpha$ -LA gene from *Bubalus arnee bubalis* (accession number AF194373; <http://www.ncbi.nlm.nih.gov/Database/index.html>).

The intergenic segment of the *LALBA* gene was amplified by PCR and nested PCR, first from the third exon of the upstream gene to the second exon of the downstream gene and then from the fourth exon of the upstream gene to the first exon of the downstream gene (Figure 2a).

Polymerase chain reaction was performed with primer sequences of Exon II Reverse (5'-GGT AAA AAG AAG TTA CTC AC-3') and Exon III Forward (5'-TCC ACC



**Figure 2.** PCR methods and results. a) Schematic representation of PCR strategies to amplify the intergenic portion. b) Agarose gel (0.8%) electrophoresis analysis of PCR (lane 3) and nested PCR (lane 2) compared with  $\lambda$  DNA-*Hind*III DNA size marker in lane 1. Lane 2, exhibiting a single band of 6,200 bp, confirms the presence of an intergenic portion; the smear observed in lane 3 indicates a low yield of specific segment and the presence of many PCR fragments. Color version available in the online PDF.

PCR Fragment	<u>GTCTCCTTCTTCATGATCAGGTTGGCCCATAAAGCACTC</u>	39
LALBA gene	<u>GTCTCCTTCTTCATGATCAGGTTGGCCCATAAAGCACTC</u> *****	2437
PCR Fragment	<i>TGTTCTGAGAAGCTGGATCAGTGGCTCTGTGAGAAGTTGTGAACACCTGCTGTCTTTGCT</i>	99
LALBA gene	<i>TGTTCTGAGAAGCTGGATCAGTGGCTCTGTGAGAAGTTGTGAACACCTGCTGTCTTTGCT</i> *****	2497
PCR Fragment	GCTTCTGTCTCTTTCTGTTCTGGAACCTCTGCCCCATGGCTACCTCGTTTTGCTTC	159
LALBA gene	GCTTCTGTCTCTTTCTGTTCTGGAACCTCTGCCCCATGGCTACCTCGTTTTGCTTC *****	2557
PCR Fragment	TTTGTAGCCCTTGAAGCTAACTCGTCTCTGAGCCCTGGGCCCTGTAGTGGCAATGGACA	219
LALBA gene	TTTGTAGCCCTTGAAGCTAACTCGTCTCTGAGCCCTGGGCCCTGTAGTGGCAATGGACA *****	2617
PCR Fragment	TGTAAGGACTAATCTCCAGGGATGCATGAATGGCGCTCTGGACTTTTGACCCTTGCTCAA	279
LALBA gene	TGTAAGGACTAATCTCCAGGGATGCATGAATGGCGCTCTGGACTTTTGACCCTTGCTCAA *****	2677
PCR Fragment	TGTCCTGATAGCACTTTAATGCAACAGTGCATATCCACTTCTGTCTCTGAATAAAAG	339
LALBA gene	TGTCCTGATAGCACTTTAATGCAACAGTGCATATCCACTTCTGTCTCTGAATAAAAG *****	2737
PCR Fragment	CCTGATTTGAGTGGCTGGCTGTATTTCTCTCGTAGGAGAGGGAGGAAATAGGGTGA	399
LALBA gene	CCTGATTTGAGTGGCTGGCTGTATTTCTCTCGTAGGAGAGGGAGGAAATAGGGTGA *****	2797
PCR Fragment	<p>§</p> <p>GTAGGTAGACTTGGCCATGGGTACAGGCACCTTTATCTCTACTAAGGAGGAGAGAGACTGAACTTATAACAATC AAAGATGGAGATTACTTCTGTATTAATCAATTCAACAGAGTTTATTGATCACCTAGCATAAATCAAAGAATATG GAGGGATCTAAAGTTGACTAAAAGCATCTTACCTAAACTGCTGCTAAGTCGCTTCAGTCGTGTCTGACTCTGTGT GACCCATAGACGGCAGCCACAAGGCTCCCATGTCCTGGGATTCTCCAGGCAAGAACAATGGAGTGGGTGCCATT TCCTTCTCCAGTGCATGAAAGTAAAAGTAAAAGTGAAGTCGCTCAGCCGTGTCTGACTCAGTGACCCCATGGACTGC AGCCACCAGGCTCCTCCGCCATGGGGTTTTCGAGGCAAGGAGTACGTGGAGTGGGGTGCATTGCCTTCTCCACTT ACCTAAACTAGTATACTACAAAACGTTAATATATAGTTATAAATACAAGATTCAATGGACATGGGCTGGGCAACACTG GGAGATGGTGGGGACAGGGAGTCTGGTGGGCTGCAGTCCATGGGGTCAAAAGACTCAGACATGACATAGTGACTG AGCAGCAACAACAACAGCAACATAATACAAGATATAGTGTCAATAGTATTATTAGAGAGTTAGAAAATGTCAATAT CTTCTAATCCCTCTTCTTACCTCTATACCCACTGGGCCTTTTGGCATTACAGACTATATTTGTCTGTCTCAGAGACT AGCAGTCCCAAATGAAGCTGCCATAATCTGTTTAGGGACATATAGGCCCTGTGAGCAGTGAACAGAATATCCCTCT CTTACACTTATGAAAACAACCTCTCGGAAATAACTTGGTATTCATCATATGATTTTAGTCTGAATTTGAACACAATG TGACACATCTCCAGGGTACTGTATGCTATTCCTTCCCTAATATGGGACAAAATGCTTTCCCAAGTCTACAAGG TTTTAGAATAATTGCTATTACAGTCTATGTGGGAGGGCAGGATTGAATGACCATCCACCAACCAACGCACACAT CTCCTTATTATATGCTTCGGTTACAGTCTCAGGAAGGCATCACTTCAGTTCAGTTCAGTCACTCAGTCGTGTCCTCACT CTTTGCAACCCCATGAACCTTAGTACGCAAGGCCTACCTGCCATCAACCACTCCTGGAGTTCACCAAACTCATGTT CATTGAGTCAGTGCATGCCATCCACCATCTCACTCTCTATCGTCCCTTCTCCTCCGGCCCAATCTTTCCAGCAT TAGGGTCTTTCAAATGAGTCACTCTTCATATCAGGTGGCCAAAGTACTGGAGTTTCAGCTTCAACATCAGTCTCTC TAATGAACACCCAGGACTGATCTCCTTTAGAGTGGACTGGTTGGATCTCCTTGCAGTCCATGTGACTCTCAAGAGTCT TCTCCAACACCACAGTTCAAAAGCATCAAGTCTTCAGCGCTCAACTTTCTTACAGTCCAACCTCTCACATCCATACAT GACCACTGGAAAAACCATAGCCTTGAGCAGATGGACCTTTGTTGGCAAAGTAAATGTCTCTGCTTTTGATATGCTATCT AGGTTGGTCATAACTTTCTTCCAAGGAGTAAAGTGTCTTTAATTTTCATGGCTGCAGTCAACATCTGCAGGGATTG GAGCCCCCAAAAATAAAGTCTGACACTGTTTCCACTGTTTCCCATCTATTTCCCATGAAGTATGGGAACAGATGCC ATAATCTTCGTTTCTGAATGTTGAGCTTTAAGCCAACTTTTCCACTCTCCGCTTCACTTTTCAAGAGGCTGTG AGTTCTCTTCACTTTCTGCCATAAGGGTGGTGTCTGATATCTGAGGTTATTGATATTTCTCCAGCAATCTTA ATTCCAGCTTGTGCTTCTTCCAACCCAGCGTTTCTCATGATGTACCCTGCATATAAGTTAAATAAGCAGGGTGACAA ATACAGCCTTGACATACTCCTTTCTCTTCAAACCCAGTCTGTTGTTACTTGTCTAGGGTGATAGAATTTCTATTCC</p>	

**Figure 3.** DNA sequence analysis of the amplified nested PCR segment containing the intergenic portion. The genomic sequence input is the  $\alpha$ -LA (LALBA) gene of *Bubalus bubalus* (accession number AF194373; <http://www.ncbi.nlm.nih.gov/Database/index.html>). The alignments and comparisons performed using the Blast database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) provided the genomic DNA sequence, its inferred protein translation (1-letter code), and the matching parts of the target protein sequence. Numbering of the PCR fragment started from the first nucleotide sequenced up to 6,199; for the LALBA gene, the same numbering reported in the database (AF194373) was used. The primers used for the nested PCR are in underlined letters and exons are in italics. Framed nucleotides refer to various structural patterns with presumed functional roles in the 5' and 3' untranslated regions as CAT and TATA boxes, transcription start (+), recognition signal for polyadenylation, and termination site of the mRNA transcript (§).

TGTTCTCCATGTA CTCTCCAGTGTCTCTGAATGGAATGGAGGAGAAATGCAGTGGAAAGCACTTCTGATATTAGGACT  
 TTCTGCTCATTTCTCTTGGAACACCAAGATCCTCTACACTCATTTCACTCATGGTCTGATGCCAAAACTTGATCTC  
 TGGGCACTGAATTGAATCTTGGAGACAGAGTTTTGGGTGAAGTAGAAGATAATAGCTGTATTGCTTTTCAAAGCAAAG  
 GGAGGACACAATGGGCTCATGCCCTTCAAACCTGTGTGTCCCAAGCTGGGAAGATTTGGTGAGGAATTTTTTGCATG  
 GTTCAAGGGCAGGGTTGTTGATAAGAGTAGGATGTGTTCAAGGCTTCACTCCTTAACTGGCCCTCAGGTAGTCTTC  
 TCTGAAATGAATGCTAACATCTTCTATTTGTTGGAGGTTTTAATTCTGCAAAGAGCTCAAAGATGTTGTTATGTGTAT  
 CCCTTAAGGTGAAACTTGGGGCTCCCTGGTGGCTCAGATGGTAAAGCGTCTGCCTGCAATGCGGGAGACCTGGGTTC  
 AATCCCTGGGTTGGGAGGATCCCTGGAGAAGGAAATGGCAACCCACTCCAGTACTCTTTCTGGAAAAATCCATGGA  
 CAGAGGAGCCTTGTAGGCACAGTCCATGGGGTCTGTAAGAGTTTGACATGACTGAGCGACTTCACTTAAGGTGAAACT  
 AGGACCCTGCCCCAAGGCTGTGCTGTTGTTCTTGGCTGCTGCTTCTGTCTGCACTCCACTTCCCCTGATTGGC  
 AGTTTTTTGAATCTGCCCTTTGGAATTCAGGTAGGGCTTGGAGGCTGGAGTCTATTCCCTACAAAGATAAATGGGG  
 GACTTGAAGAACTTCTGTTTTCAAGAGCCACAGGGTCTGTTAGTTTCACTACTTCATATAGTGATTATCAGTA  
 CTTTATATAGTGATAATCAGTATTTATATAGTGATAATGCCGAGTACATCATGTGAAATGACGGGCTGGATGAAGCA  
 CAAGCTGGGATCAAGATCGCCAGGAGAAATATCAATAACTTCAGATATGCAGATGACACCACCTTATGGCAGAAAA  
 GAAGAGAACTGAAGATTTCTTGGAGAAAGTAAAGAAAGGTTGAAAAGCTGGCTTAAACTCAACATTTCAGATAA  
 CTAAGATCATACATATCATCTGATACCATGACTCCATGGCAAATAGATGGGAAACAATGGAAACAGTGAGAGACTTT  
 ATTTTCTTGGGCTCCAAAATCACTGCAGATGGTACTGCAGCCATGAAATTTAAAGACACTTGCCTTGAAGAAAA  
 GCTATGACTAACCTAGACAGCATATTTAAAGAGCAGAGACATTACCAACAAGGTCACCTAGTCAAAGCTATGGTTTT  
 TCCGGTAGTATGATGGATGTGAGAGACTATAAAGAAAGCTGAGTGTGGAAGAAATGATGCTTTTGAAGCTGGTGT  
 TGGAGAAGACTCTTGGAGTCCCTTGGACTTCAAGGAGTCCAAGCACTCCACTTCTGAAAGGACTCAGTCCCTGGTGT  
 TCATTGGAAGGACTGATGCTGAAGTTGAAGCTCCAATACTTTGGCCACTGGATGCAAAGAACTGACTCATGTGAAAG  
 ACCCTGATACTGGGAAAGATTGAAGGCAGGAGGAGAAGGGATGACAGAGGATGGAAGTGTGGATGGAATCACCAACT  
 CTATGGACATGAGTGTGAGCAAGCTCCAGGAGTTGGTAATGGGCAGGGAAAGCCTGGCATGCTGCAGTCCATGGGTTG  
 CAAAGAGTTGGACACTACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT  
 GGAAGAGTTTGGCCCTGATTCTGAAGAGTTTGAAGATATAAAGTTTGAATAACCTAAACTTTTAGTATCTTTGTTT  
 GGAAGTCTTAAATTATTTACTTAGGATGGGTACCCTGCAATATAAGAAATCAGGTTTTAGAGACTGAGGTAGAGAG  
 AATGAGCCCTGGCATAACAAGACTAACAGCTATTGGTATCATAGCATGAAGCTTGTGAGCAGCAATTTTTAGAACTA  
 GTATCCTAAACTCTACATGCTCCAGGACACTGATCTTAAAGCTCAGGTTGAGAAATCGTGTTTATAGGCTCTAAATGT  
 ATATTGTGGGGCTTCCCTGGTGGCTCAGATGGTAAAGGGTCTGCCTGCAATGTGGGTGATCTGGGTTCAATCCCTGGC  
 TTGGGAAGACTCTCTGGAGAAGGAAATGGCAACCCACTCCAGTCTTGCCTGGAAAATCCATGGCAGAGGAGGCC  
 TTGTAAGCTACAGTCCACGGGATCACAAAGAGTCGGACACGACTGAGCAACTAAGCACAGCAGTACAGTATACACC  
 TGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT  
 GAGACACCAGGGAAGCCCAAGAATACTGGAGTGGGTAGCCTATTCCTTCTCCAGGGGATCTTCTATCCAGAAATG  
 AACTGGAGTCTCCTGCATTTCAAGTGGACTCTTACCAGCTGAAGTACCAGGTGTATACTACTCCAATATTAAGTGC  
 TTAAGTCCAGTTTTCCACCTTTCCCAAAAGGTTGGGTCACTCTTTTTTAACTTCTGTGGCCTACTCTGGGCTT  
 TCTACAAGTTATATATTTATGGACACATTTATTGCAAGCCGTTAGTTTTAGATTTACAATGTGGTATCTGGCTATTT  
 AGTGGTATTGGTGGTTGGGGATGGGGAAGCTGATAACATCTCAGAGGGTAGCTAGATACTGTATACACACTTTTCAA  
 GTTCTCCATTTTTGTGAAATAGAAAGTCTCTGGATCCAAGTTATATGTGATTCTCAGTCTCTGTGGTCATATCTATT

PCR Fragment	CTATTCTGACCACTCAACAAGGAAACCAAGATATCAAG	5509
LALBA Gene	CAACAAGGAAACCAAGATATCAAG	24
	*****	
PCR Fragment	GGACACTTCTTTTGTTCATGCCTGGGTTGAGTGGGCCAGTGTGAGTCTGATCCTGGGA	5569
LALBA Gene	GGACACTTCTTTTGTTCATGCCTGGGTTGAGTGGGCCAGTGTGAGTCTGATCCTGGGA	84
	*****	
PCR Fragment	CCCTGAGGCCCTTTGTTCTGAAAGTAGGTTGGGCACATCTTGCTTCTAGAACCAACA	5749
LALBA Gene	CCCTGAGGCCCTTTGTTCTGAAAGTAGTGTGGGCACATCTTGCTTCTAGAACCAACA	264
	*****	
PCR Fragment	CTACCAGAAACAACATAAATAAAGCCAAATGGGAAACAGGATCATGTTTGTAACTGTT	5809
LALBA Gene	CTACCAGAAACAACATAAATAAAGCCAAATGGGAAACAGGATCATGTTTGTAACTGTT	324
	*****	
PCR Fragment	TGGGCAGGTAACAATACCTAGTATGGACTAGAGATTCTGGGGAGGAAAGGAAAGTGGGG	5869
LALBA Gene	TGGGCAGGTAACAATACCTAGTATGGACTAGAGATTCTGGGGAGGAAAGGAAAGTGGGG	384
	*****	

**Figure 3 (Continued).** DNA sequence analysis of the amplified nested PCR segment containing the intergenic portion. The genomic sequence input is the  $\alpha$ -LA (LALBA) gene of *Bubalus bubalus* (accession number AF194373; <http://www.ncbi.nlm.nih.gov/Database/index.html>). The alignments and comparisons performed using the Blast database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) provided the genomic DNA sequence, its inferred protein translation (1-letter code), and the matching parts of the target protein sequence. Numbering of the PCR fragment started from the first nucleotide sequenced up to 6,199; for the LALBA gene, the same numbering reported in the database (AF194373) was used. The primers used for the nested PCR are in underlined letters and exons are in italics. Framed nucleotides refer to various structural patterns with presumed functional roles in the 5' and 3' untranslated regions as CAT and TATA boxes, transcription start (+), recognition signal for polyadenylation, and termination site of the mRNA transcript (§).

PCR Fragment	TGAAATTACTGAAGGAAGCTAGCTGGCTCAATGTTTCTTTGTTGGTTTTACTGGCCTCTC	5929
LALBA Gene	TGAAATTACTGAAGGAAGCTAGCAGGCTCAATGTTTCTTTGTTGGTTTTACTGGCCTCTC	444
	*****	
PCR Fragment	TCGTCATCCTCTTCTCGGATGTAAGGCTTGATGCCAGGGCCCCTGAGGCTTTTTCCACAA	5989
LALBA Gene	TCGTCATCCTCTTCTCGGATGTAAGGCTTGATGCCAGGGCCCCTGAGGCTTTTTCCACAA	504
	*****	
	+	
PCR Fragment	<b>ATAAAA</b> GGAGGTGAGCAGTGTGGTGACCCCATTTCCAGGATCTTGGGGGTAACCAAAATG	6049
LALBA Gene	<b>ATAAAA</b> GGAGGTGAGCAGTGTGGTGACCCCATTTCCAGGATCTTGGGGGTAACCAAAATG	564
	*****	
	M	
PCR Fragment	ATGTCCTTTGTCTCTCTGCTCCTGGTAGGCATCCTATTCCACGCCACCCAGGCTGAACAA	6109
LALBA Gene	ATGTCCTTTGTCTCTCTGCTCCTGGTAGGCATCCTATTCCATGCCACCCAGGCTGAACAA	624
	*****	
PCR Fragment	TTAACAAAATGTGAGGTGTTCCGGGAGCTGAAAGACTTGAAGGACTATGGAGGTGTCAGT	6169
LALBA Gene	TTAACAAAATGTGAGGTGTTCCGGGAGCTGAAAGACTTGAAGGACTACGGAGGTGTCAGT	684
	*****	
PCR Fragment	<u>TTGCCTGAATGTGAGTTCCTGCTATTTTG</u>	6199
LALBA Gene	<u>L P E Primer Exo I</u>	714
	*****	

**Figure 3 (Continued).** DNA sequence analysis of the amplified nested PCR segment containing the intergenic portion. The genomic sequence input is the  $\alpha$ -LA (LALBA) gene of *Bubalus bubalus* (accession number AF194373; <http://www.ncbi.nlm.nih.gov/Database/index.html>). The alignments and comparisons performed using the Blast database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) provided the genomic DNA sequence, its inferred protein translation (1-letter code), and the matching parts of the target protein sequence. Numbering of the PCR fragment started from the first nucleotide sequenced up to 6,199; for the LALBA gene, the same numbering reported in the database (AF194373) was used. The primers used for the nested PCR are in underlined letters and exons are in italics. Framed nucleotides refer to various structural patterns with presumed functional roles in the 5' and 3' untranslated regions as CAT and TATA boxes, transcription start (+), recognition signal for polyadenylation, and termination site of the mRNA transcript (§).

TGT AAC TCC TGC CA-3') and with the following thermocycler profile: 94°C for 1 min; 10 cycles of 94°C for 30 s, 53°C for 30 s, 68°C for 3 min; 15 cycles of 94°C for 30 s, 53°C for 30 s, 68°C for 3 min + 15 s/cycle; final extension at 72°C for 15 min.

Nested PCR was performed with primer sequences of Exon IV Forward (5'-GTC TCC TTC TTC ATG ATC AG-3') and Exon I Reverse (5'-CAA AAT AGC AGG GAA CTC AC-3') and with the following thermocycler profile: 94°C for 1 min; 10 cycles of 94°C for 30 s, 57°C for 30 s, 68°C for 3 min; 15 cycles of 94°C for 30 s, 57°C for 30 s, 68°C for 3 min + 15 s/cycle; final extension at 72°C for 15 min.

Five microliters of PCR and nested PCR were loaded into a 0.8% agarose gel and visualized by ethidium bromide staining. The primers were drawn on sequences of exons 2 and 3. Figure 2b shows the results of the amplifications with the nested PCR in lane 2 and the PCR in lane 3.

When the PCR amplification was performed using genomic DNA as a template, a smeared DNA product was observed (lane 3); however, if this first amplification product was used as a template in a second PCR-amplification nesting step, the expected 6.2-kbp specific DNA fragment was obtained (lane 2).

The former exhibited a single band (of 6,200 bp), as expected in the presence of an intergenic region,

and the latter displayed a wide smear representing the many PCR fragments present and a low yield of specific segments. Given these findings, blood samples from an additional 15 buffalos were analyzed following the same procedure as in the previous case. The results obtained from these samples confirmed the presence of an amplified nucleotide fragment of about 6,200 bp in most of them, though they were all characterized by an  $\alpha$ -LA B monomorphic phenotype.

The DNA was excised from the gel and purified with the Qiaex II gel extraction kit (Qiagen, Valencia, CA); 50 ng of the eluted DNA fragments were ligated into the pGEM-T Easy Vector (Promega, Madison, WI) at 4°C overnight in 10  $\mu$ L of 1 $\times$  rapid ligation buffer according to the manufacturer's instructions. The top 10 competent cells were transformed with the ligation mixture and grown on a Luria broth agar plate containing 100  $\mu$ g/mL of ampicillin, 9  $\mu$ L of isopropyl  $\beta$ -D-1-thiogalactopyranoside IPTG (200 mg/mL), and 45  $\mu$ L of X-gal (20 mg/mL). Two colonies were picked and grown in 5 mL of Luria broth containing 100  $\mu$ g/mL of ampicillin. Plasmid DNA was purified by QIAprep spin miniprep kit (Qiagen) and sequenced with T7 and Sp6 primers on a Perkin-Elmer sequencer (model ABI-377, Perkin-Elmer, Foster City, CA). Alignments of the nucleotide sequences were performed with CLUSTAL W (<http://www.expasy.ch/tools/#proteome>).

A couple of the 6,200 bp fragments obtained was then purified, cloned in the pGEM-T Easy vector system, and sequenced by an ABI sequencer. The sequences of the amplified nested PCR segments containing the intergenic portion were aligned with the LALBA gene (accession number AF194373; <http://www.ncbi.nlm.nih.gov/Database/index.html>). Both sequences coincided with that of exon 4 and the untranslated region at the 3' of the upstream gene and exon 1 and the untranslated region at the 5' of the downstream gene, indicating a tandemly repeated copy of the LALBA gene (Figure 3a, b, c).

Genomic analysis was performed on the intergenic DNA amplified fragment using the Censor Repeat Mask program (<http://www.girinst.org>). As expected, the analysis detected the presence of some repeat fragments belonging to the ERV/ERV1 and retrotransposon Non LTR/SINE.

Observations from the genomic databases for several eukaryotic species suggest that duplicate genes arise at a very high rate, on average 0.01 per gene per million years (Lynch and Conery, 2000). It has recently been proposed that the fixation process is frequently driven by positive selection for enhanced gene dosage brought about by gene duplication (Kondrashov and Kondrashov, 2006). Disregarding the neutralist–selectionist debate, an increase in the gene product resulting from additional copies of a structural gene has however been reported quite often. For example, both in humans (Ma et al., 2001) and in sheep (Pieragostini et al., 2003) it was found that the greater the number of  $\alpha$ -globin genes, the greater the production of  $\alpha$ -globins. In the case of LALBA, an increase in the gene product may be advantageous to the newborn, as is the case in mice where  $\alpha$ -LA expression was found to be gene dosage-dependent. (Stacey et al., 1995).

Our results lend further support to the hypothesis of Lemay et al. (2009) that milk and mammary genes are more likely to be duplicated in therians than in other genes of the bovine genome and that variation in copy number of the milk protein genes may contribute to the taxonomic diversity of milk composition. In conclusion, this is only the first step in considering LALBA duplication in buffalos because the gene arrangements and their spread in different populations have yet to be investigated.

#### ACKNOWLEDGMENTS

The authors are indebted to Athina Papa for her accuracy in revising the English of the manuscript and Maria Strazzullo for expert bioinformatics assistance.

#### REFERENCES

- Addeo, F., J. Mercier, and B. Ribadeau Dumas. 1976. Elements de structure primaire de l'alpha-lactalbumine et de la beta-lactalbumine de buffle. *FEBS Lett.* 63:255–259.
- Bleck, G. T., and R. D. Bremel. 1993. Sequence and single-base polymorphisms of the bovine alactalbumin 5' flanking region. *Gene* 126:213–218.
- Chianese, L., S. Caira, S. Lilla, F. Pizzolongo, P. Ferranti, G. Pugliano, and F. Addeo. 2004. Primary structure of water buffalo alpha-lactalbumin variant A and B. *J. Dairy Res.* 71:14–19.
- Clegg, J. B., S. E. Y. Goodbourn, and M. Braend. 1984. Genetic organisation of the polymorphic equine alpha globin locus and sequence of the BII alpha 1 gene. *Nucleic Acids Res.* 12:7847–7858.
- Dayhoff, M. O. 1976. The origin and evolution of protein superfamilies. *Gene* 126:213–218.
- Goossens, M., and Y. Y. Kan. 1981. DNA analysis in the diagnosis of hemoglobin disorders. *Methods Enzymol.* 76:805–817.
- Kondrashov, F. A., and A. S. Kondrashov. 2006. Role of selection in fixation of gene duplications. *J. Theor. Biol.* 239:141–151.
- Lemay, D. G., D. J. Lynn, W. F. Martin, M. C. Neville, T. M. Casey, G. Rincon, E. V. Kriventseva, W. C. Barris, A. S. Hinrichs, A. J. Molenaar, K. S. Pollard, N. J. Maqbool, K. Singh, R. Murney, E. M. Zdobnov, R. L. Tellam, J. F. Medrano, J. B. German, and M. Rijnkels. 2009. The bovine lactation genome: Insights into the evolution of mammalian milk. *Genome Biol.* 10:R43.
- Lubon, H., and L. Henninghausen. 1988. Conserved region of the rat  $\alpha$ -lactalbumin promoter is a target site for protein binding in vitro. *Biochem. J.* 256:391–396.
- Lynch, M., and L. C. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Ma, S. K., W. Y. Au, A. Y. Chan, and L. C. Chan. 2001. Clinical phenotype of triplicated alpha-globin genes and heterozygosity for beta0-thalassemia in Chinese subjects. *Int. J. Mol. Med.* 8:171–175.
- McKenzie, H. A. 1996. Alpha-lactalbumins and lysozymes. *EXS* 75:365–409.
- Miglior, F., A. Sewalem, J. Jamrozik, J. Bohmanova, D. M. Lefebvre, and R. K. Moore. 2007. Genetic analysis of milk urea nitrogen and lactose and their relationships with other production traits in Canadian Holstein cattle. *J. Dairy Sci.* 90:2468–2479.
- Ohno, S. 1970. *Evolution by Gene and Genome Duplication*. Springer-Verlag, New York, NY.
- Pieragostini, E., F. Petazzi, and A. Di Luccia. 2003. The relationship between the presence of extra alpha-globin genes and blood cell traits in Altamura sheep. *Genet. Sel. Evol.* 35(Suppl. 1):S121–S133.
- Pieragostini, E., R. Rullo, A. Scaloni, G. Bramante, and A. Di Luccia. 2005. The alpha chains of goat hemoglobins: Old and new variants in native Apulian breeds. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 142:18–27.
- Pieragostini, E., A. Scaloni, R. Rullo, and A. Di Luccia. 2000. Identical marker alleles in Podolic cattle (*Bos taurus*) and Indian zebu (*Bos indicus*). *Comp. Biochem. Physiol. Part B* 127:1–9.
- Scaloni, A., E. Pieragostini, A. Malorni, L. Ferrara, and A. Di Luccia. 1998. Hemoglobin  $\alpha$ -globin chain polymorphism in Podolic cattle: Primary structure determination of two new genetic variants by mass spectrometry and amino acid sequencing. *Biochimie* 80:333–338.
- Soulier, S., J. C. Mercier, J. L. Vilotte, J. Anderson, A. J. Clark, and C. Provat. 1989. The bovine and ovine genomes contain multiple sequences homologous to the alpha-lactalbumin-encoding gene. *Gene* 83:331–338.
- Stacey, A., A. Schnieke, M. Kerr, A. Scott, C. McKee, I. Cottingham, B. Binas, C. Wilde, and A. Colman. 1995. Lactation is disrupted by alpha-lactalbumin deficiency and can be restored by human alpha-lactalbumin gene replacement in mice. *Proc. Natl. Acad. Sci. USA* 92:2835–2839.
- Vestri, R., E. Pieragostini, F. Yang, A. Rando, and P. Masina. 1991. Expression of triplicated and quadruplicated alpha globin genes in sheep. *Br. J. Haematol.* 77:110–116.