https://doi.org/10.17221/14/2018-CJAS

The haemoglobin subunits alpha and beta: Old and new genetic variants in the Italian Mediterranean buffalo

Rosario Rullo[1], Aldo Di Luccia[2], Elena Ciani[3], Elisa Pieragostini[4]*

[1]*Institute for the Animal Production Systems in the Mediterranean Environment, Napoli, Italy*
[2]*Department of Agricultural, Food and Environmental Sciences, University of Foggia, Foggia, Italy*
[3]*Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari "Aldo Moro", Bari, Italy*
[4]*Department of Emergency and Organ Transplantation, University of Bari "Aldo Moro", Bari, Italy*
**Corresponding autor: elena.ciani@uniba.it*

Citation: Rullo R., Di Luccia A., Ciani E., Pieragostini E. (2019): The haemoglobin subunits alpha and beta: Old and new genetic variants in the Italian Mediterranean buffalo. Czech J. Anim. Sci., 64, 279–290.

Abstract: Haemoglobin (HB), the most widely distributed respiratory pigment in the animal kingdom, is among the best characterized oxygen-binding proteins, both at functional and molecular level. However, very little information is available about the genomic features of HB in river buffalo (*Bubalus bubalis*), even though there are reports in literature confirming the presence of interesting polymorphisms at the protein level in Mediterranean buffalo. We hence address the characterization of exonic as well as intronic nucleotide polymorphism in the haemoglobin subunit alpha and beta in a set of nine Italian Mediterranean buffaloes exhibiting different HB phenotypes. The nine buffaloes were selected from a random set of 398 samples, previously analysed for their HB protein polymorphism, in order to account for both globin variants and the evolution of intron variability within the most common domesticated species of the family *Bovidae*. All four sequenced clones of the subunit alpha were 1311 bp, whereas the length of the five different sequenced clones of the subunit beta ranged from 1841 to 1960 bp, due to an insertion of 119 nucleotides. Six polymorphic sites were detected in the four amplicons of alpha subunit. Among them, two variations concern exclusively haplotype *A*, while four sequence variations were found to be specific to haplotype *B*. Several variations, both in exonic and intronic regions, were detected in the *B. bubalis* subunit beta. In conclusion, the nucleotide sequence variants observed in this work substantiate the known haemoglobin protein polymorphisms, and an updated protein nomenclature is provided here. In addition, we observed a high sequence similarity in the overall pattern of variation in the haemoglobin subunits, possibly the results of a concerted evolution, with relatively more extensive gene homogenization in river buffalo than in other ruminant species.

Keywords: *Bubalus bubalis*; haemoglobin subunit variation; biodiversity; transposable elements; gene duplication

The Italian Mediterranean buffalo (*Bubalus bubalis*), which belongs to the *Bubalus* genus of the *Bovidae* family, is generally classified as a river sub-type of water buffalo, similar to the buffalo breeds of Hungary, Romania and the Balkan countries (Bigi and Zanon 2008). It was introduced into Italy probably in Roman times or with the Arab conquests, spreading across Southern Italy, where today it forms an economically significant proportion of livestock. Widely used in the past as a draught animal, the Italian Mediterranean water buffalo is now mainly kept for milk production to

https://doi.org/10.17221/14/2018-CJAS

make mozzarella cheese. A herd-book was opened in 1980, and the breed was officially recognised in 2000. Since then, as selection pressure has focused mainly on milk production, it may be assumed that its genetic variability has been decreasing. As is widely known, a decrease in genetic diversity can, in the long run, limit the ability of populations to adapt to future environmental changes. In recent years, growing attention has focused on functional genetics, an area that is recurrently included amongst animal breeding research priorities. Thanks to its accessibility and obvious biological importance, haemoglobin has been one of the most studied polymorphisms in vertebrate species since the infancy of both population and evolutionary genetics. The functional effect of the haemoglobin phenotype on haematological patterns has been demonstrated in humans as well as in mammalian species (Pieragostini et al. 2010). Four statements are crucial to our research mindset. First, being the most widely-distributed respiratory pigment in the animal kingdom, haemoglobin is the best characterized oxygen-binding protein, both at functional and molecular level. Secondly, genetic diversity is generally quantified by alleles and genotype frequencies at one or more loci. Thirdly, though the structure and organization of haemoglobin subunit clusters has been reported in many species, very little information is available about the genomic features of river buffalo. Finally, interesting protein polymorphisms in the haemoglobin genetic system have been found in Mediterranean buffalo (Di Luccia et al. 1991a; Iorio et al. 2004).

Thus, our research moved toward the genetic characterization of exonic as well as intronic polymorphism in the subunits alpha and beta, in order to account for both globin variants and the evolution of intron variability within the most common domesticated species of the family *Bovidae*.

MATERIAL AND METHODS

Nomenclature. The nucleotide sequences of all the buffalo's haemoglobin alpha and beta subunits have been deposited in the EMBL/GenBank DNA database and the accession numbers are listed in Table 1. Parts A and B of Table 1 show the gene names as well as the protein denominations reported according to the National Center for Biotechnology Information guidelines (Murphy et al. 2006). As shown in Figure 1, the subunits in haplotype *A* and those in haplotype *B* encode four different globin chains in total, namely αS and αN the former, and αI and αK the latter. Within subunits encoded by haplotype *A*, αS differs from αN at two amino acid positions (Leu130Phe and Ser132Asn). These replacements are not present in the alpha globin subunits encoded by haplotype *B*; in addition, both αS and αN differ from αI and αK at the amino acid residue 65 (Ala65Asn) (Ferranti et al. 2001). As for the subunits encoded by genes of haplotype *B*, αI and αK exhibit a single amino acid substitution each, namely Ile11Val and Gln12Lys, respectively. Finally, as for the beta globin cluster, only two adult beta globin chains

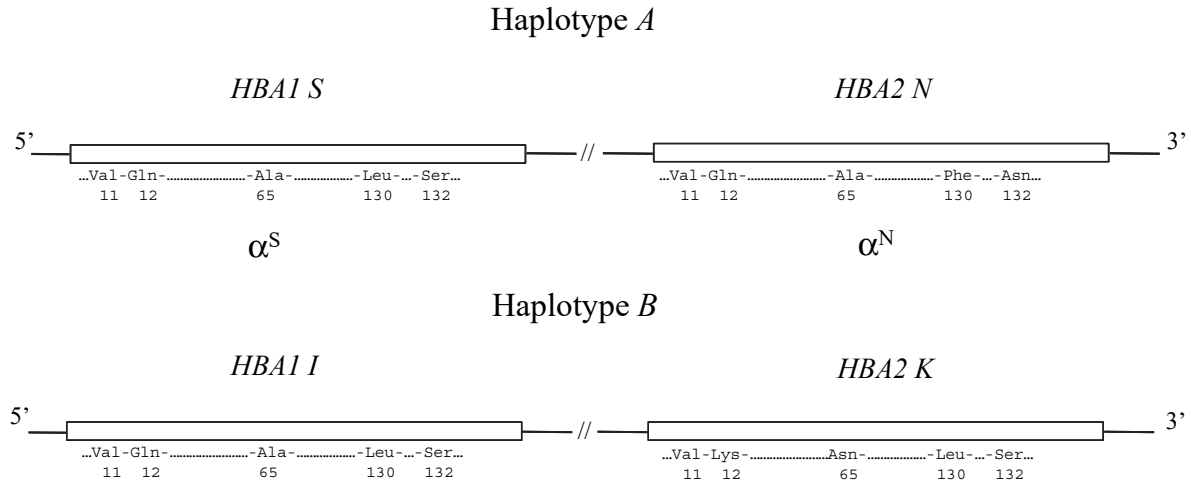

Figure 1. Schematic representation of the alpha subunits (haplotypes *A* and *B*) from *B. bubalis*. Current name of gene, and the corresponding protein, are indicated on the top, and bottom, respectively. Only variant positions of the amino acid sequence are indicated

https://doi.org/10.17221/14/2018-CJAS

Table 1. Part A shows the name and Accession number of alpha and beta subunits. Part B compares new names of haemoglobin (HB) with the previous ones and indicates the corresponding tetramers. Both gene and protein denominations are reported according to the National Center for Biotechnology Information guidelines (Murphy et al. 2006). Part C summarizes the tetramers observable depending on the haemoglobin electrophoretic phenotypes, which, in turn, depend on the genetic polymorphism at subunits alpha and beta. (The gene and protein variants have been named using as superscript letter the letter code indicating the substitution in the mature protein consequent to the mutation in the codons (Table 2). The superscript letter S assigned to the HB phenotype indicates the presence of the slow electrophoretic mobility of HB's carrying the β^T variant; the phenotypes are obviously named according to the homozygosity at *HBB* subunits and to the fact that *HBA* diplotypes exhibit equal or different subunits)

Part A (genes)				Part B (proteins)				Part C		
Subunit alpha	Accession No.	subunit beta	Accession No.	HB previous name[1]	HB new name	HB tetramer	globin chain responsible for the HB name	alpha diplotype	HB phenotype	HB tetramer per phenotype
HBA1 S	AJ242731	*HBB A*	AM886147	HB1	HB S	$\alpha^S{}_2\beta^A{}_2$	α^S	*AA*	AA	$\alpha^S{}_2\beta^A{}_2$, $\alpha^N{}_2\beta^A{}_2$
				HB3	HB N	$\alpha^N{}_2\beta^A{}_2$	α^N		AA^S	$\alpha^S{}_2\beta^A{}_2$, $\alpha^N{}_2\beta^A{}_2$, $\alpha^S{}_2\beta^T{}_2$, $\alpha^N{}_2\beta^T{}_2$
HBA2 N	AJ242733	*HBB T*	AM886148	HB2	HB I	$\alpha^I{}_2\beta^A{}_2$	α^I		A^SA^S	$\alpha^S{}_2\beta^T{}_2$, $\alpha^N{}_2\beta^T{}_2$
				HB4	HB K	$\alpha^K{}_2\beta^A{}_2$	α^K		AB	$\alpha^S{}_2\beta^A{}_2$, $\alpha^N{}_2\beta^A{}_2$, $\alpha^I{}_2\beta^A{}_2$, $\alpha^K{}_2\beta^A{}_2$
HBA1 I	AJ242732	*HBB G*	AM886151	HB1s	HB ST	$\alpha^S{}_2\beta^T{}_2$	α^S and β^T	*AB*	A^SB^S	$\alpha^S{}_2\beta^T{}_2$, $\alpha^N{}_2\beta^T{}_2$, $\alpha^I{}_2\beta^T{}_2$, $\alpha^K{}_2\beta^T{}_2$
				HB3s	HB NT	$\alpha^N{}_2\beta^T{}_2$	α^N and β^T		AB A^SB^S	$\alpha^S{}_2\beta^A{}_2$, $\alpha^N{}_2\beta^A{}_2$, $\alpha^I{}_2\beta^A{}_2$, $\alpha^K{}_2\beta^A{}_2$, $\alpha^S{}_2\beta^T{}_2$, $\alpha^N{}_2\beta^T{}_2$, $\alpha^I{}_2\beta^T{}_2$, $\alpha^K{}_2\beta^T{}_2$
HBA2 K	AJ242734	*HBB E*	AM886150	HB2s	HB IT	$\alpha^I{}_2\beta^T{}_2$	α^I and β^T	*BB*	BB	$\alpha^I{}_2\beta^A{}_2$, $\alpha^K{}_2\beta^A{}_2$
				HB4s	HB KT	$\alpha^K{}_2\beta^T{}_2$	α^K and β^T		BB^S	$\alpha^I{}_2\beta^A{}_2$, $\alpha^K{}_2\beta^A{}_2$, $\alpha^I{}_2\beta^T{}_2$, $\alpha^K{}_2\beta^T{}_2$
		HBB P	AM886149	HBG	HB G	$\alpha^S{}_2\gamma_2$	γ		B^SB^S	$\alpha^I{}_2\beta^T{}_2$, $\alpha^K{}_2\beta^T{}_2$
				HBE	HB E	$\alpha^S{}_2\varepsilon_2$	ε			

[1]Iorio et al. (2004)

Czech Journal of Animal Science, 64, 2019 (7): 279–290

https://doi.org/10.17221/14/2018-CJAS

Table 2. Variant codons and predicted amino-acid residues in the alpha subunits of the haemoglobin genes in *Bovinae* and *Caprinae* subfamilies as reported in the GenBank database

Farmed species	α Subunits	Accession No.	Codon										
			8	9	11	12	24	65	72	90	116	130	132
Bubalus bubalis	*HBA1 S*	AJ242731	AAG Lys	AGC Ser	GTC Val	CAG Gln	GAT Asp	GCC Ala	GGC Gly	CAC His	AAT Asn	TTG Leu	AGC Ser
	HBA2 N	AJ242733	AAG Lys	AGC Ser	GTC Val	CAG Gln	GAT Asp	GCC Ala	GGC Gly	CAC His	AAT Asn	TTC Phe	AAC Asn
	HBA1 I	AJ242732	AAG Lys	AGC Ser	ATC Ile	CAG Gln	GAT Asp	AAC Asn	GGC Gly	CAC His	AAT Asn	TTG Leu	AGC Ser
	HBA2 K	AJ242734	AAA Lys	AGC Ser	GTC Val	AAG Lys	GAT Asp	AAC Asn	GGC Gly	CAC His	AAT Asn	TTG Leu	AGC Ser
Bos taurus	*HBA1 Y*	AJ242799	AAG Lys	GGC Gly	GTC Val	AAG Lys	GAG Glu	GCC Ala	GAA Glu	TAC Tyr	AGT Ser	TTG Leu	AAC Asn
	HBA1 N	AJ242797	AAG Lys	GGC Gly	GTC Val	AAG Lys	GAG Glu	GCC Ala	GAA Glu	CAC His	AGT Ser	TTG Leu	AAC Asn
	HBA2 S	AJ242798	AAG Lys	GGC Gly	GTC Val	AAG Lys	GAG Glu	GCC Ala	GAA Glu	CAC His	AGT Ser	TTG Leu	AGC Ser
Capra hircus	*HBA1*	J00043	AAG Lys	TCC Ser	GTC Val	AAG Lys	GCT Ala	GCC Ala	GGC Gly	CAC His	AAT Asn	TTG Leu	AAC Asn
	HBA2	J00044	AAG Lys	TCC Ser	GTC Val	AAG Lys	GCT Ala	GCC Ala	GGC Gly	CAC His	AGT Ser	TTG Leu	AAC Asn
Ovis aries	*HBA1*	X70213	AAG Lys	TCC Ser	GTC Val	AAG Lys	GCT Ala	GCT Ala	GGC Gly	CAC His	AAT Asn	TTG Leu	TTG Leu
	HBA2	X70215	AAG Lys	TCC Ser	GTC Val	AAG Lys	GCT Ala	GCT Ala	GGC Gly	CAC His	AAT Asn	TTG Leu	TTG Leu

https://doi.org/10.17221/14/2018-CJAS

Table 3. List of the nine individuals characterized by different haemoglobin (HB) electrophoretic phenotypes selected in this study for the DNA sequencing

Sample	Phenotype	Globin symbols for HB tetramers	*HBA* alleles				*HBB* alleles	
1	AA	$\alpha^{S}_{2}\beta^{A}_{2}$, $\alpha^{N}_{2}\beta^{A}_{2}$	*HBA1 S*	*HBA2 N*	–	–	*HBB A*	–
2	AA^{S}	$\alpha^{S}_{2}\beta^{A}_{2}$, $\alpha^{N}_{2}\beta^{A}_{2}$, $\alpha^{S}_{2}\beta^{T}_{2}$, $\alpha^{N}_{2}\beta^{T}_{2}$	*HBA1 S*	*HBA2 N*	–	–	*HBB A*	*HBB T*
3	$A^{S}A^{S}$	$\alpha^{S}_{2}\beta^{T}_{2}$, $\alpha^{N}_{2}\beta^{T}_{2}$	*HBA1 S*	*HBA2 N*	–	–	–	*HBB T*
4	BB	$\alpha^{I}_{2}\beta^{A}_{2}$, $\alpha^{K}_{2}\beta^{A}_{2}$	–	–	*HBA1 I*	*HBA2 K*	*HBB A*	–
5	BB^{S}	$\alpha^{I}_{2}\beta^{A}_{2}$, $\alpha^{K}_{2}\beta^{A}_{2}$, $\alpha^{I}_{2}\beta^{T}_{2}$, $\alpha^{K}_{2}\beta^{T}_{2}$	–	–	*HBA1 I*	*HBA2 K*	*HBB A*	*HBB T*
6	$B^{S}B^{S}$	$\alpha^{I}_{2}\beta^{T}_{2}$, $\alpha^{K}_{2}\beta^{T}_{2}$	–	–	*HBA1 I*	*HBA2 K*	–	*HBB T*
7	AB	$\alpha^{S}_{2}\beta^{A}_{2}$, $\alpha^{N}_{2}\beta^{A}_{2}$, $\alpha^{I}_{2}\beta^{A}_{2}$, $\alpha^{K}_{2}\beta^{A}_{2}$	*HBA1 S*	*HBA2 N*	*HBA1 I*	*HBA2 K*	*HBB A*	–
8	$A^{S}B^{S}$	$\alpha^{S}_{2}\beta^{T}_{2}$, $\alpha^{N}_{2}\beta^{T}_{2}$, $\alpha^{I}_{2}\beta^{T}_{2}$, $\alpha^{K}_{2}\beta^{T}_{2}$	*HBA1 S*	*HBA2 N*	*HBA1 I*	*HBA2 K*	*HBB A*	*HBB T*
9	AB $A^{S}B^{S}$	$\alpha^{S}_{2}\beta^{A}_{2}$, $\alpha^{N}_{2}\beta^{A}_{2}$, $\alpha^{I}_{2}\beta^{A}_{2}$, $\alpha^{K}_{2}\beta^{A}_{2}$, $\alpha^{S}_{2}\beta^{T}_{2}$, $\alpha^{N}_{2}\beta^{T}_{2}$, $\alpha^{I}_{2}\beta^{T}_{2}$, $\alpha^{K}_{2}\beta^{T}_{2}$	*HBA1 S*	*HBA2 N*	*HBA1 I*	*HBA2 K*	*HBB A*	*HBB T*

were found, differing in a neutral-to-neutral amino acid replacement, Ala134Thr (Di Luccia et al. 1991a; Iorio et al. 2004). Part C of Table 1 summarizes the tetramers observable depending on the haemoglobin electrophoretic phenotypes, which, in turn, depend on the genetic polymorphism at subunits alpha and beta. Particularly, the gene and protein variants have been renamed according to the letter code indicating the substitution in the mature protein consequent to the mutation in the codons, as shown in Table 2.

Blood samples. Previously, a total of 398 adult animals had been randomly sampled in two breeding areas of the Campania region and analysed for HB polymorphism (Iorio et al. 2004). Thereafter, nine individuals characterized by different HB phenotypes (Table 3) were selected in this study for DNA sequencing.

DNA analyses. Polymerase chain reaction (PCR) of haemoglobin subunits alpha and haemoglobin subunits beta was carried out using about 20–50 ng of genomic DNA extracted from leukocytes according to Goossens and Kan (1981). The genes were amplified with 1 U of AmpliTaq Gold DNA Polymerase (Applied Biosystems, USA) using 0.4 μM heterologous primers designed on the basis of the goat haemoglobin alpha subunit and the sheep beta one, Accession No. J00044 and X14727, respectively. After an initial activation step (95°C for 10 min), the reaction mixture was subjected to a three-step cycle repeated 30 times, each step consisting of 30 s denaturation at 95°C, 30 s annealing at 55°C, and 1 min extension at 72°C.

The primers used were: α-forward (5'-T_{-349}CCCTTCCGTTCTCAAGACA$_{-330}$-3'), α-reverse (5'-A_{+962}GGGTGGCCATTTCAGCACT$_{+943}$-3'), β-forward (5'-A_{-157}AGTCAGTGCCAGGAAGACC$_{-138}$-3'), β-reverse (5'-C_{+1803}ATCTGAGTAAGAGACAGTG$_{+1784}$-3').

The amplicons were purified on agarose gel by QIAquick Gel Extraction Kit (Qiagen, USA), and ligated to the pGEM-T Easy Vector (Promega, USA). After transformation into the *E. coli* TG1 strain (Promega), positive colonies containing the plasmid DNA were purified by QIAprep Spin Miniprep Kit (Qiagen) and sequenced with T7 and SP6 primers. Restriction and modifying enzymes were from GE Healthcare Bio-Sciences (USA) or Promega; oligonucleotide synthesis and nucleotide sequencing was carried out at Primm (Italy); ampicillin, isopropyl-β-D-thiogalactopyranoside (IPTG) and X-Gal were from Sigma-Aldrich (USA).

Phylogenetic classification. For an evolutionary study on the haemoglobin subunit alpha of *B. bubalis*, the corresponding genes of other farmed species were included in the analysis. The nucleotide sequence alignment and the percentage of protein identity were obtained with CLUSTALW2 or EMBOSS Needle software, respectively; interspersed elements were analysed by a RepeatMasker Web Server. The phylogenetic tree was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood

https://doi.org/10.17221/14/2018-CJAS

(MCL) approach, and then selecting the topology with superior log likelihood value. The robustness of the inferred tree was evaluated by applying 1000 bootstrap replications. Analyses were conducted in MEGA X (Kumar et al. 2018).

RESULTS

Evidence from HB sequence analysis. Aligning the nucleotide sequences of the various amplicons obtained from the buffalo's DNA samples (Table 3), and the sequence of both the goat and sheep alpha and beta subunits, gave the following results: (i) the size of each of the four different sequenced clones of the subunit alpha was 1311 bp; (ii) the length of the five different sequenced clones of the subunit beta ranged from 1841 to 1960 bp, due to an insertion of 119 nucleotides (data not shown); (iii) the nucleotide sequences of the amplicons of both alpha and beta subunits show the presence of canonical hallmarks characteristic of the haemoglobin genetic structures, namely three exons and two introns with consensus intron/exon splice junctions; (iv) the 5' UTR included all sequences defined as important for transcription and translation events such as the CCAAT box, the ATA box and the presumed mRNA Cap site, whereas, in the 3' UTR, the hexanucleotide AATAAA polyadenylation signal and the poly (A) addition site were present; (v) the ATG start codon and the TAA stop codon were present.

In addition, Table 2 shows codons differing among the alpha subunits in *B. bubalis* and/or when contrasted to the corresponding codons in *Bovinae* and *Caprinae* subfamilies, as reported in the GenBank according to the related sequence Accession number. As for *B. bubalis,* six codons were detected to vary among the four amplicons of the alpha subunit. Notably, at codon position 8 (with reference to the Accession No. AJ242731), the subunit *HBA2 K* presents the AAA triplet while the three other subunits present the AAG triplet (c.24G>A), both of them coding for lysine. At codon position 11, the subunit *HBA1 I* presents the ATC triplet while the three other subunits present the GTC triplet (c.31G>A), the former coding for isoleucine while the latter codes for valine. At codon position 12, the subunit *HBA2 K* presents the AAG triplet (lysine) while the three other subunits present the CAG triplet (glutamine) (c.34A>C). At codon position 65, two subunits (*HBA2 K* and *HBA1 I*) present the AAC triplet (asparagine) while the two other subunits present the GCC triplet (alanine) (c.193G>A and c.194A>C). At codon position 130, the subunit *HBA2 N* presents the TTC triplet (phenylalanine) while the three other subunits present the TTG triplet (leucine) (c.390G>C). At codon position 132, the subunit *HBA2 N* presents the AAC triplet (asparagine) while the three other subunits present the AGC triplet (serine) (c.395A>G). The variant codons at positions 8, 11 and 12 are located in the first exon, the one at position 65 in the second exon, and the two at positions 130 and 132 are

Table 4. Variant codons in the coding region of the beta subunits of *B. bubalis* and deduced amino-acid replacements. An asterisk (*) indicates the conserved nucleotide with respect to the *HBB A* allele. Synonymous codons are indicated by the abbreviation "syn". In the *HBBP* gene, a stop codon at position 58 (with reference to the Accession No. AM886149) terminates the amino-acid sequence

Subunit beta	Codon						
	58	81	96	101	115	128	134
HBB A	AAG Lys	AAG Lys	CAT His	AAC Asn	CGC Arg	GAC Asp	GCT Ala
HBB T	***	**A syn	***	***	***	***	A** Thr
HBG	***	***	**C syn	***	***	***	***
HBE	***	***	***	T** Tyr	*A* His	ACT Ser	A** Thr
HBBP	T** End	*** –	*** –	*** –	*A* –	ACT –	A** –

https://doi.org/10.17221/14/2018-CJAS

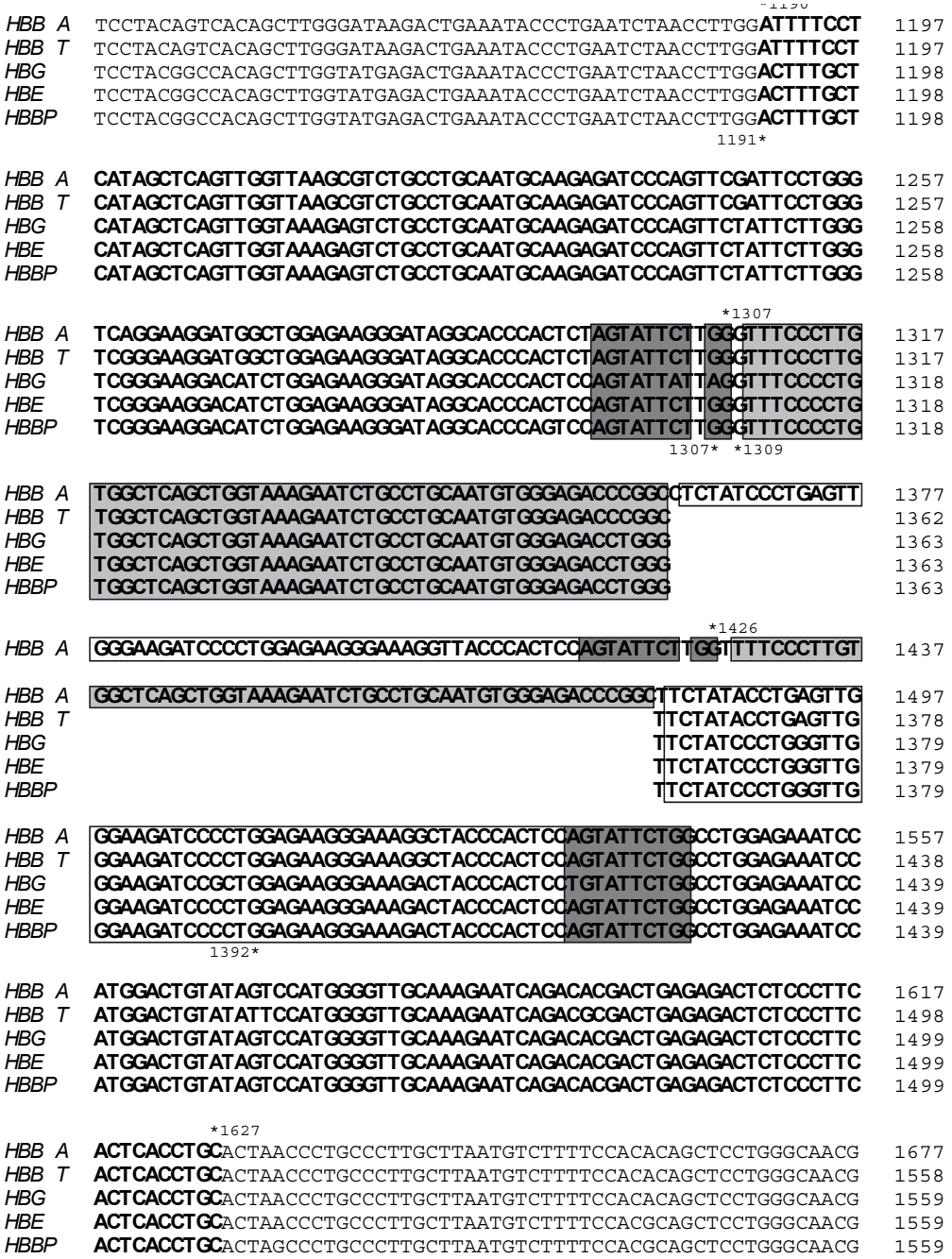

Figure 2. Multiple sequence alignment of the five beta subunits of *B. bubalis* in the region surrounding the 119 bp insertion of the *HBB A* allele (Accession No. AM886147). The SINE elements are bolded; an asterisk identifies the start or end point for each element; top numbers define the location of SINE in the *HBB A* allele, with respect to the other allele (*HBB T*) or the other subunits (bottom numbers). In the *HBB A* allele, three SINEs were located at positions g.1190–1307, g.1307–1426 and g.1426–1627 (top numbers), the beginning of the second and third SINE is coincident with the end of the first and second SINE, respectively. In the *HBB T* allele (Accession No. AM886148) and remaining subunits, two SINEs were located at positions g.1191–1307 and g.1309–1392 (bottom numbers). Partial segments of the g.1363–1481 insertion of the *HBB A* allele are repeated in adjacent upstream or downstream regions of the other beta subunits, and in particular: (i) the first one (unshadowed box extending from g.1364 to 1415) is downstream repeated at position g.1483–1534 of all five beta subunits; (ii) the second one (dark grey box extending from g.1416 to 1426) is a copy of the g.1297–1307 upstream region and is essentially repeated in the g.1535–1544 downstream region; (iii) the third one (light grey box extending from g.1428 to 1481) is a copy of the g.1309–1362 upstream region

https://doi.org/10.17221/14/2018-CJAS

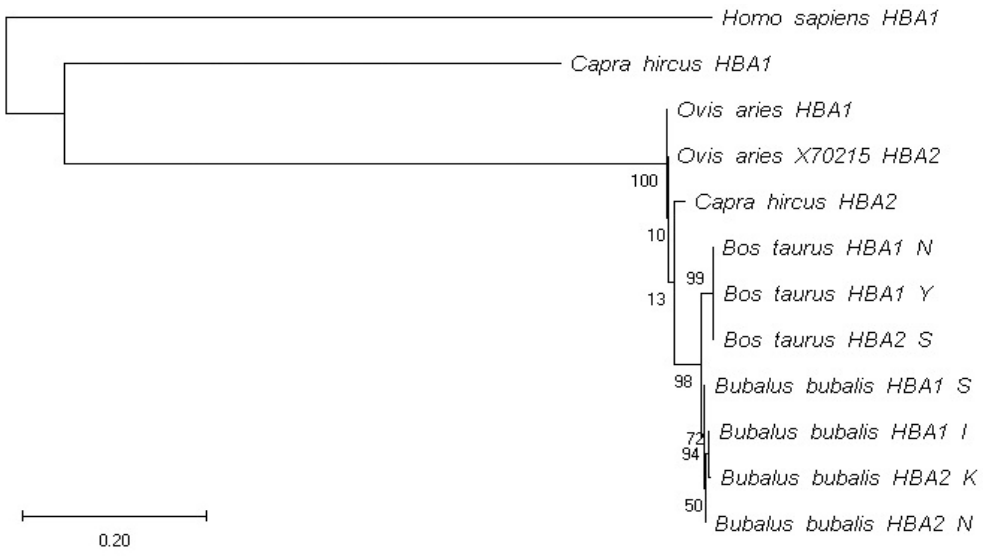

Figure 3. Maximum Likelihood phylogenetic tree of the alpha subunit nucleotide sequences. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Numbers on nodes refer to the bootstrap confidence values obtained after 1000 replicates. The accession numbers of the considered livestock sequences are listed in Table 2. The human sequence (Accession No. V00491) was included as out-group

located in the third exon of the corresponding genes. Table 4 presents codons differing among the beta subunits of *B. bubalis* as well as the deduced amino acid replacements. Divergent nucleotides with respect to the *HBB A* allele sequence are shown, while asterisks point to the conserved nucleotides. In the *HBBP* gene, a stop codon at position 58 (with reference to the Accession No. AM886149) terminates the amino acid sequence.

Comparison of the multiple alignments among the amplified beta subunits of *B. bubalis* nucleotide sequences evidenced the presence of several variations located in the exonic and intronic regions. As shown in Table 1, *HBB A* and *HBB T* alleles correspond to the codified beta globin chains β^A and β^T, respectively. In their coding sequence (with reference to the Accession No. AM886147), they differ only for two codons, i.e. at codon 81 (second exon), where AAG and AAA triplets, both coding for lysine, are respectively present, and at codon 134 (third exon), where GCT (alanine) and ACT (threonine) are present. However, the most relevant difference between the *HBB A* and *HBB T* alleles concerns their second intronic regions where, in *HBB A,* eight random variations and the aforementioned insertion of 119 nucleotides after position g.1363 (Accession No. AM886147) can be found (Figure 2). Notably, the insertion is essentially constituted by three repeated segments, namely: (i) the first (unshadowed box extending g.1364–1415), which is downstream repeated at position g.1483–1534 in all five subunits in the beta cluster; (ii) the second (dark grey box extending g.1416–1426), which is a copy of the g.1297–1307 upstream region and is essentially repeated in the g.1535–1544 downstream region; (iii) the third (light grey box extending g.1428–1481), which is a copy of the g.1309–1362 upstream region.

As for the three remaining beta subunits, comparison of the multiple alignments evidenced that: (i) the *HBG* gene exhibits a variation, located at position c.288 (Accession No. AM886151) of the second exon, leading to a synonym triplet (CAC instead of CAT), coding for His96; (ii) in the *HBE* gene, a nucleotide substitution, located at position c.301 (Accession No. AM886150) of the second exon, changes the AAC codon for Asn101 into the TAC codon for Tyr101; (iii) the *HBBP* gene shows a sequence variant in the exonic region at position c.172 (Accession No. AM886149) of the second exon replacing the AAG codon for Lys58 with the TAG stop codon.

Furthermore, when compared to the other genes, *HBE* and *HBBP* share five sequence variations in the third exon, located at c.344, c.382, c.383, c.384 and c.400 (with reference to both Accession No. AM886150 and AM886149), all producing amino-acid substitutions. The above cited amino-acid replacements are listed in Table 4.

Evidence from the phylogenetic analysis. By observing the nucleotide differences found in

https://doi.org/10.17221/14/2018-CJAS

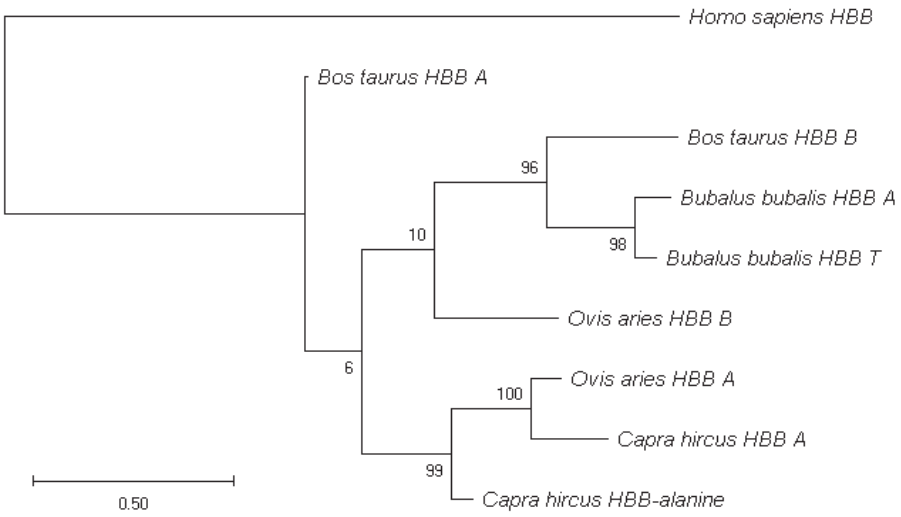

Figure 4. Maximum Likelihood phylogenetic tree of the beta subunit nucleotide sequences. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Numbers on nodes refer to the bootstrap confidence values obtained after 1000 replicates. Accession numbers of the considered sequences are, for *B. bubalis*, AM886147 (*HBB A*) and AM886148 (*HBB T*); for *B. taurus*, AB512625 (*HBB A*) and AB512624 (*HBB B*); for *O. aries*, DQ352470 (*HBB A*) and DQ352471 (*HBB B*); for *C. hircus*, DQ350619 (*HBB A*) and M15387 (*HBB-alanine*). Despite showing a nucleotide sequence different from DQ350619, the sequence M15387 also encodes for *HBB A*. The human sequence (Accession No. KU350152.1) was included as out-group

eleven codons among the alpha subunits of the four ruminant species considered (Table 2), we notice that most variations involve only genes of *B. bubalis*, such as those related to codons 8, 11, 12, 65, 130 and 132. Conversely, comparing the sequences in the non-buffalo species, codons 9, 24 and 132 account for the most relevant differences among them.

Figure 3 shows the phylogenetic tree built on the basis of a comparison between the corresponding alpha subunits in the considered species. Clearly, an early branch allowed the separation of the *Bovinae* from the *Caprinae* clade; a successive branching separated *B. bubalis* from *B. taurus*, followed by a separation between *C. hircus* and *O. aries*. Figure 4 shows the phylogenetic tree built on the basis of a comparison between the corresponding beta subunits in the considered species. Again in this case, the branching pattern was indicative of a recent common ancestor between *B. bubalis* and *B. taurus* beta subunits.

DISCUSSION

Haemoglobin subunits alpha. The nucleotide sequences of the four alpha subunits confirm the existence of two haplotypes named *A* and *B* whose sequence variations are restricted to the exonic regions. Based on the evidence provided by a comparison within and among the considered species, the mutational events related to the sequence variations in the various alleles seem to have occurred more or less recently in evolutionary time, thus highlighting a different rate of gene homogenization (Elder and Turner 1995; Liao 1999; Carson and Scherer 2009; Ganley 2013). In other words, the presence of a reduced or high number of variant alleles might be explained as recent or ancient mutations, respectively. For instance, the nucleotide variation corresponding to Ile11Val may represent a recent sequence variation, whereas the one corresponding to Gln12Lys seems to be more ancient. Similarly, Leu130Phe may correspond to a recent mutation while Asn132Ser may correspond to an ancient one. As for the other farmed species, the point mutations responsible for the His90Tyr and Asn132Ser replacements could correspond to recent sequence variation in *B. taurus*, similarly to Asn116Ser in *C. hircus*. On the other hand, the codon differences among the various species at position 9, 24, 72 and 116 seem to be the oldest ones, as they are almost completely homogenized and fixed in each species. In this regard, no discriminant codons apparently exist for *Bovinae,* whereas *Caprinae* exhibit discriminant codons in the above-mentioned positions.

https://doi.org/10.17221/14/2018-CJAS

Concerning codon 65, a number of mammalian species exhibit a putatively ancestral codon GAC for Asp (Kleinschmidt and Sgouros 1987). This codon, due to an early c.194A>C variation, transformed into GCC, coding for the Ala present in the less common haplotype *A* of river buffalo, in *B. taurus*, *C. hircus* and *O. aries*, as well as in other *Artiodactyla* species. Another c.193G>A variation changed the GAC for Asp into AAC for Asn, present in the *B* haplotype of river buffalo, as well as in northern elk; this substitution involves a CpG dinucleotide, considered to be a hot spot mutation site (Perutz 1990). Furthermore, based on these findings and a definitive knowledge of the deduced primary structures, some discrepancies emerge with respect to a previous report (Ferranti et al. 2001), regarding an erroneous assignment at amino acid position 9, 44 and 116. The gene sequence analysis showed that the correct codons are AGC, TTC and AAT corresponding to Ser9, Phe44 and Asn116, instead of the erroneously assigned Gly, Pro and Ser.

The separation between *Bovinae* and *Caprinae* appears relevant on the phylogenetic tree, showing that these families are divided into two coherent groups. In particular, the *Caprinae* and *Bovinae* clades took a different evolutionary route from a common ancestor. The earlier evolutionary differentiation involved the *Bovinae* clade, followed by a branch separating *B. bubalis* from *B. taurus*.

Haemoglobin subunits beta. The five beta subunits amplified from *B. bubalis* L. were classified as adult (*HBB A* and *HBB T*), fetal (*HBG*), embryonic (*HBE*) and pseudogene (*HBBP*) on the basis of a comparison of their deduced amino-acid sequences with those from related species. The first observation on beta subunits from river buffalo regarded the presence of a nucleotide insertion spanning 119 bp found in the second intronic region of only the *HBB A* allele. This region could be considered as an inverted duplicated segment, probably deriving from a cross-chromosome rearrangement; it is likely that transposable elements were responsible for this modification, thus confirming their pivotal role in shaping the genome (Finnegan 1989). Indeed, an analysis performed with the RepeatMasker Web Server allowed the identification of transposable elements belonging to the family of short interspersed nuclear elements (SINE) in all five beta subunits. In the *HBB A* allele (Accession No. AM886147), three SINEs were located at position g.1190–1307, 1307–1426 and g.1426–1627, whereas in the *HBB T* allele (Accession No. AM886148) and remaining beta subunits, two SINEs were located at g.1191–1307 and g.1309–1392 (Figure 2).

The evolutionary classification of the five beta subunits of river buffalo was better assessed through pair-wise alignments of their deduced amino acid sequences performed among the various β-chains, thanks also to the comparison with other species (Table 5). Indeed, a progressive decrease in amino-acid identity is evident moving from adult to embryonic beta globin in *O. aries*, *C. hircus* or *B. taurus*. A similar decrease observed for beta globins in *B. bubalis* allowed a reliable assignment of the corresponding beta subunits of this species, although the five beta globins of river buffalo showed higher degrees of identity, with values ranging from 99.3 to 91.7%. In the other species, the amino-acid identity decreased significantly up to 66.0% for *O. aries*, 65.3% for *C. hircus* or 61.9% for *B. taurus*. Meanwhile, in *B. bubalis,* the amino-acid identity suggests that the product of the *HBG* gene is a fetal beta chain, (95.2% identity with β^A), the product of the *HBE* gene is an embryonic β-chain (92.4% identity with β^A) whereas, due to the presence of a stop codon in the second exon, *HBBP* is a pseudogene. However, the termination apart, as regards the deduced β-chain from *HBBP*, an overall 99.3% identity with the embryonic beta chain encoded by *HBE* was found, suggesting that *HBBP* and *HBE* may result from the duplication of a common ancestral gene.

Contrary to what was observed for the alpha subunits, the Maximum Likelihood tree topology for the beta subunits did not show a clear separation between *Bovinae* and *Caprinae.* However, only branches connecting beta subunit sequences within the two sub-families were strongly supported by bootstrap values. In addition, beta subunits have recently been highlighted as not particularly suited for phylogenetic reconstruction, due to the low rate of substitutions in the coding regions (Jiang et al. 2015). Finally, the higher affinity for oxygen of the beta-like globin gene cluster (Nasuda-Kouyama et al. 1983) as well as its complexity due to the presence of both developmentally and environmentally regulated genes (Pieragostini et al. 2010; Philipsen and Hardison 2018) account for the difference in the adaptive value of the beta subunits compared to the alpha subunits.

https://doi.org/10.17221/14/2018-CJAS

Table 5. Comparison of the amino-acid identities obtained from a pair-wise alignment among the deduced beta chains in different farmed species

	Accession No.	Beta chains	Amino acid identity (%)				
			β^A	β^T	γ	ε	ψ
Bubalus bubalis	AM886147	β^A	100				
	AM886148	β^T	99.3	100			
	AM886151	γ	95.2	94.5	100		
	AM886150	ε	92.4	93.1	97.2	100	
	AM886149	γ	91.7	92.4	96.6	99.3	100
			β^A	β^B	γ	ε^4	ε^2
Bos taurus	P02070	β^A	100				
	AB512624	β^B	90.3	100			
	P02081	γ	80.7	84.1	100		
	P06643	ε^4	65.3	66.7	67.3	100	
	P06642	ε^2	61.9	63.3	63.3	94.6	100
			β^A	$\beta^{D\ malta}$	β^C	γ	ε^2
Capra hircus	P02077	β^A	100				
	unknown[1]	$\beta^{D\ malta}$	99.3	100			
	P02078	β^C	81.4	85.5	100		
	P02082	γ	81.4	80.7	78.6	100	
	P02105	ε^2	65.3	65.3	63.3	66.7	100
			β^A	β^B	β^C	γ	
Ovis aries	Q1KYZ7	β^A	100				
	Q1KYZ6	β^B	95.2	100			
	P68056	β^C	89.0	85.5	100		
	P02083	γ	81.4	81.4	77.9	100	

[1]Bannister et al. 1979

CONCLUSION

Our results confirm that the overall pattern of genetic variation among the haemoglobin subunits is characterized by a high sequence similarity, and corroborate the hypothesis that the phenomenon is the result of a concerted evolution. The hallmark of 'concerted evolution' is that, in a group of species sharing a repetitive sequence family, the sequences are more homogeneous within each species than they are when compared between species. This observation implies that the common ancestral genome already had the repetitive sequence family, and that there was homogenization of the diverging sequences as they descended into the various present-day species. In the light of these theories, the relative gene homogenization of river buffalo seems to be more extensive than in other ruminant species.

In addition, our findings provide further proof of haemoglobin's generally accepted role as a model protein for studies aimed at monitoring genetic diversity among farmed animal species subjected to intensive artificial selection towards production traits. In this regard, the genetic characterization of the globin chains in *B. bubalis* in the present study substantiate the haemoglobin protein polymorphisms in river buffalo previously reported by Di Luccia et al. (1991b) and Iorio et al. (2004). Moreover, according to the latter authors, at the time of their survey, in the buffalo population in Campania, alpha haplotypes *A* and *B* (inferred from electrophoretic patterns) were found with a frequency of 0.044 and 0.956 respectively, while the beta alleles were respectively present with a frequency of 0.622 and 0.378. Based on these reference data and on the fact that neither alpha nor beta subunits have been the target of artifi-

https://doi.org/10.17221/14/2018-CJAS

cial selection, a new survey might be useful to investigate possible loss of variability. Such an investigation might be relevant mostly because, following the establishment of the herd book and the increasingly marked demands for mozzarella cheese, in the last two decades the buffalo population has undergone intensive selection for milk production.

Acknowledgement. The authors wish to thank the language consultant, Anthony Green, for proofing the English manuscript.

REFERENCES

Bannister J.V., Bannister W.H., Wilson J.B., Lam H., Miller A., Huisman T.H. (1979): The structure of goat hemoglobins V. A fourth β chain variant (β-D-Malta; 69 Asp is replaced by Gly) with decreased oxygen affinity and occurring at a high frequency in Malta. Hemoglobin, 3, 57–75.

Bigi D., Zanon A. (2008): Atlas of Native Italian Breeds: Bovines, Equines, Sheep, Goats and Pigs Reared in Italy. Edagricole, Milan, Italy. (in Italian)

Carson A.R., Scherer S.W. (2009): Identifying concerted evolution and gene conversion in mammalian gene pairs lasting over 100 million years. BMC Evolutionary Biology, 9, 156.

Di Luccia A., Iannibelli L., Addato E., Masala B., Manca L., Ferrara L. (1991a): Evidence for the presence of two different β-globin chains in the hemoglobin of the river buffalo (Bubalus bubalis L.). Comparative Biochemistry and Physiology B, 99, 887–892.

Di Luccia A., Iannibelli L., Ferranti P., Manca L., Masala B., Ferrara L. (1991b): Electrophoretic and chromatographic evidence for allelic polymorphisms in the river buffalo α-globin gene complex. Biochemical Genetics, 29, 421–430.

Elder Jr. J.F., Turner B.J. (1995): Concerted evolution of repetitive DNA sequences in eukaryotes. The Quarterly Review of Biology, 70, 297–320.

Ferranti P., Facchiano A., Zappacosta F., Vincenti D., Rullo R., Masala B., Di Luccia A. (2001): Primary structure of α-globin chains from river buffalo (Bubalus bubalis L.) hemoglobins. Journal of Protein Chemistry, 20, 171–179.

Finnegan D.J. (1989): Eukaryotic transposable elements and genome evolution. Trends in Genetics, 5, 103–107.

Ganley A.R.D. (2013): Concerted evolution. In: Maloy S., Hughes K. (eds): Brenner's Encyclopedia of Genetics, Vol 2. Academic Press, San Diego, USA, 126–130.

Goossens M., Kan Y.Y. (1981): DNA analysis in the diagnosis of hemoglobin disorders. Methods in Enzymology, 76, 805–817.

Iorio M., Vincenti D., Annunziata M., Rullo R., Bonamassa R., Di Luccia A., Pieragostini E. (2004): Biochemical and molecular investigations on qualitative and quantitative Hb polymorphism in the river buffalo (Bubalus bubalis L.) population reared in Southern Italy. Genetics and Molecular Biology, 27, 167–173.

Jiang Y., Wang X., Kijas J.W., Dalrymple B.P. (2015): Beta-globin gene evolution in the ruminants: Evidence for an ancient origin of sheep haplotype B. Animal Genetics, 46, 506–514.

Kleinschmidt T., Sgouros J.G. (1987): Hemoglobin sequences. Biological Chemistry Hoppe-Seyler, 368, 579–615.

Kumar S., Stecher G., Li M., Knyaz C., Tamura K. (2018): MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution, 35, 1547–1549.

Liao D. (1999): Concerted evolution: Molecular mechanism and biological implications. American Journal of Human Genetics, 64, 24–30.

Murphy M., Brown G., Wallin C., Tatusova T., Pruitt K., Murphy T., Maglott D. (2006): Gene Help: Integrated Access to Genes of Genomes in the Reference Sequence. Collection. Available from https://www.ncbi.nlm.nih.gov/books/NBK3841/ (accessed Jan 10, 2018).

Perutz M.F. (1990): Frequency of abnormal human hemoglobins caused by C–T transitions in CpG dinucleotides. Journal of Molecular Biology, 213, 203–206.

Philipsen S., Hardison R.C. (2018): Evolution of hemoglobin loci and their regulatory elements. Blood Cells, Molecules and Diseases, 70, 2–12.

Pieragostini E., Alloggio I., Petazzi F. (2010): Insights into hemoglobin polymorphism and related functional effects on hematological pattern in Mediterranean cattle, goat and sheep. Diversity, 2, 679–700.

Received: 2018–01–28
Accepted: 2018–12–13