Genet. Sel. Evol. 31 (1999) 239–253

Original article

Genetic polymorphism of milk proteins in African *Bos taurus* and *Bos indicus* populations. Characterization of variants α_{s1}-Cn H and κ-Cn J

Marie-Françoise Mahé[a], Guy Miranda[b], Rémy Queval[c], Abou Bado[c], Paul Souvenir Zafindrajaona[d], François Grosclaude[a*]

[a] Laboratoire de génétique biochimique et de cytogénétique, Institut national de la recherche agronomique, 78352 Jouy-en-Josas cedex, France
[b] Laboratoire de biochimie et structure des protéines, Institut national de la recherche agronomique, 78352 Jouy-en-Josas cedex, France
[c] Centre international de recherche-développement sur l'élevage en zone subhumide, 01 B.P. 454 Bobo-Dioulasso 01, Burkina-Faso
[d] Laboratoire de recherches vétérinaires et zootechniques de Farcha, B.P. 433, N'Djamena, Chad

(Received 30 October 1999; accepted 16 February 1999)

Abstract – The polymorphism of caseins, α-lactalbumin and β-lactoglobulin was investigated in African *Bos taurus* (N'Dama, Baoulé, Kuri) and *Bos indicus* (Shuwa Arab, Sudanese Fulani) populations. The respective frequencies of alleles α_{s1}-Cn^B and α_{s1}-Cn^C in the N'Dama (0.89 and 0.11) and Baoulé (0.92 and 0.08) breeds were almost opposite to those found in the Shuwa Arab zebu (0.22 and 0.78), a true zebu, which confirms a phenomenon already documented in the literature. Because the α_{s1}-Cn^B, β-Cn^{A1}, κ-Cn^B haplotype was strongly predominant in N'Dama and Baoulé (0.56 and 0.60), as compared to the α_{s1}-Cn^C, β-Cn^{A2}, κ-Cn^A haplotype in the Shuwa Arab zebu (0.63), an opposite trend in frequencies was also observed between taurines and zebus at the β-Cn and κ-Cn loci. These results confirm that the polymorphism of caseins provides an efficient marker system to discriminate *Bos taurus* from *Bos indicus* origins. The Kuri was at an intermediate position, since, in this population, the α_{s1}-Cn^B allele predominated as in taurines, while the

* Correspondence and reprints
E-mail: jngroscl@biotec.jouy.inra.fr

α_{s1}-Cn^C, β-Cn^{A2}, κ-Cn^A haplotype was the most frequent, as in zebus. This may be interpreted as revealing intercrossings with zebus in the previous history of this cattle type. Conversely, but to a lesser degree, the polymorphism of the Sudanese Fulani zebu indicates a taurine influence, in accordance with what is accepted about the origins of this cattle type. No polymorphism of α_{s2}-casein could be identified, while α-lactalbumin was polymorphic in all populations. Two additional variants, probably specific to African cattle, were observed. Variant H of α_{s1}-casein, found in Kuri, is characterized by the deletion of the eight amino acid residues (51–58) coded by exon 8, a probable consequence of exon skipping. Allele α_{s1}-Cn^H is derived from allele α_{s1}-Cn^B. Variant J of κ-casein, found in Baoulé, is derived from variant B by the substitution of Ser 155 (B) → Arg (J). The existence of at least another allele of α_{s1}-casein was suggested. © Inra/Elsevier, Paris

genetic polymorphism / milk proteins / Africa / *Bos taurus* / *Bos indicus*

Résumé – Polymorphisme génétique des protéines du lait dans des populations de taurins et de zébus africains. Caractérisation des variants α_{s1}-Cn H et κ-Cn J. Le polymorphisme des caséines, de l'α-lactalbumine et de la β-lactoglobuline a été analysé dans des populations bovines africaines de type taurin (N'Dama, Baoulé, Kouri) et zébu (Choa, Peuhl). Les fréquences respectives des allèles α_{s1}-Cn^B et α_{s1}-Cn^C chez les taurins N'Dama (0,89 et 0,11) et Baoulé (0,92 et 0,08) tendent à être inverses de celles trouvées chez le zébu Choa, un zébu vrai (0,22 et 0,78), ce qui confirme un phénomène déjà signalé dans la littérature. L'haplotype α_{s1}-Cn^B, β-Cn^{A1}, κ-Cn^B prédominant nettement chez ces taurins (0,56 et 0,60), par contraste à l'haplotype α_{s1}-Cn^C, β-Cn^{A2}, κ-Cn^A chez le zébu Choa (0,63), l'inversion des fréquences entre taurins et zébus s'observe également aux loci β-Cn et κ-Cn. Ces résultats confirment que le groupe des gènes des caséines fait partie des marqueurs de choix pour discriminer entre des origines de type *Bos taurus* et *Bos indicus*. Le Kouri occupe une position intermédiaire puisque l'allèle α_{s1}-Cn^B prédomine comme chez les taurins, alors que l'haplotype le plus fréquent est α_{s1}-Cn^C, β-Cn^{A2}, κ-Cn^A comme chez les zébus. Ces particularités peuvent être interprétées comme révélant des pratiques de métissage plus ou moins anciennes avec des zébus. Inversement, mais à un bien moindre degré, le polymorphisme du zébu Peuhl révèle une influence taurine, en accord avec ce qui est admis sur les antécédents de ce type de bovin. Aucun polymorphisme de la caséine α_{s2} n'a pu être identifié, alors que l'α-lactalbumine est polymorphe dans toutes les populations. Deux variants supplémentaires, probablement spécifiques des populations africaines, ont été identifiés. Le variant H de la caséine α_{s1}, trouvé chez le Kouri, se caractérise par la délétion de la séquence de huit résidus d'acides aminés (51–58) codée par l'exon 8, conséquence vraisemblable d'une anomalie d'épissage de l'ARN messager, l'allèle α_{s1}-Cn^H dérivant de l'allèle α_{s1}-Cn^B. Le variant J de la caséine κ, trouvé chez le Baoulé, dérive du variant B par la substitution Ser 155 (B) → Arg (J). L'existence d'au moins un autre allèle de la caséine α_{s1} est suggérée. © Inra/Elsevier, Paris

polymorphisme génétique / protéines du lait / Afrique / *Bos taurus* / *Bos indicus*

1. INTRODUCTION

More than 40 years after the pioneer work of Aschaffenburg and Drewry on β-lactoglobulin [2], a vast amount of information has been collected on the genetic polymorphism of the six main bovine lactoproteins: α_{s1}-, α_{s2}-, β-

and κ-caseins, controlled by four tightly clustered loci (α_{s1}-Cn, α_{s2}-Cn, β-Cn, κ-Cn), α-lactalbumin and β-lactoglobulin, controlled by independent loci (α-La, β-Lg) [24, 28]. Investigations were primarily carried out in dairy breeds of European origin and were stimulated by the search for correlations between those polymorphisms and milk production traits, which have proved to be successful [10, 28]. In addition, the work was also extended to beef breeds, since milk protein polymorphisms are valuable markers for population studies [10, 11, 28].

Data available on African *Bos taurus* and *Bos indicus* populations, as well as on zebus as a whole, are comparatively scarce and, when they do exist, they are far less complete. As an example, the only publication providing haplotype frequencies of the casein cluster of genes is that by Grosclaude et al. on Madagascar zebus [12]. As early as 1968, Aschaffenburg and coworkers [1, 4] drew attention to the interesting features of the lactoprotein polymorphisms in *Bos indicus*, namely the predominance, at the α_{s1}-casein locus, of the C allele, contrasting with the usual higher frequency of the B allele in *Bos taurus*, and the occurrence of a polymorphism of α-lactalbumin, contrasting with the monomorphism of this protein in the various breeds of *Bos taurus* which had been investigated at that time; α-lactalbumin was, however, later found to be polymorphic in southern European breeds and this made the differentiation between taurines and zebus less clear [21, 24, 28].

The lack of data on the genetic polymorphism of milk proteins in African cattle is unfortunate because the diversity of these populations is exceptionally high, since they were derived from successive *Bos taurus* and *Bos indicus* introductions which tended to substitute for, or to mix in a complex way. According to Epstein [7] the first domestic cattle in Africa were humpless longhorn animals introduced through Egypt from South-West Asia in the second half of the 5th millenium B.C. This type is now restricted to two West-African populations, the N'Dama, whose breeding centre is the Fouta Djallon plateau in Guinea, and the Kuri, located in the Lake Chad basin (*figure 1*). A second *Bos taurus* type, the humpless shorthorn cattle, originating from the same domestication area in South-West Asia, was introduced into Africa in the 2nd millenium B.C. In West Africa, humpless shorthorns, known as Baoulé, Somba, Muturu and Lagune, are now mainly found in the coastal regions from Gambia to Cameroon. Present African zebus are derived from shorthorned thoracic humped animals which spread rapidly westwards from the Horn of Africa after the Arab invasion (about 700 A.D.). In West Africa, this type now extends along a narrow belt south of the Sahara desert (from west to east: Maure, Tuareg, Azawak and Shuwa zebus). Finally, cattle of mixed origin are widely distributed in eastern and southern Africa. In West Africa, they are represented by the long or giant horned Fulani zebus, which extend between the taurine area in the south and the zebu belt in the north. According to Epstein [7] Fulani cattle were derived from crossbreedings between longhorn humpless cattle and thoracic humped zebus.

This paper presents the results of the analysis of milk protein polymorphisms in the two longhorn humpless populations, N'Dama and Kuri, in the humpless shorthorn Baoulé, in the Shuwa Arab true zebu and in the Sudanese Fulani cattle. The four above-mentioned cattle groups are thus represented.

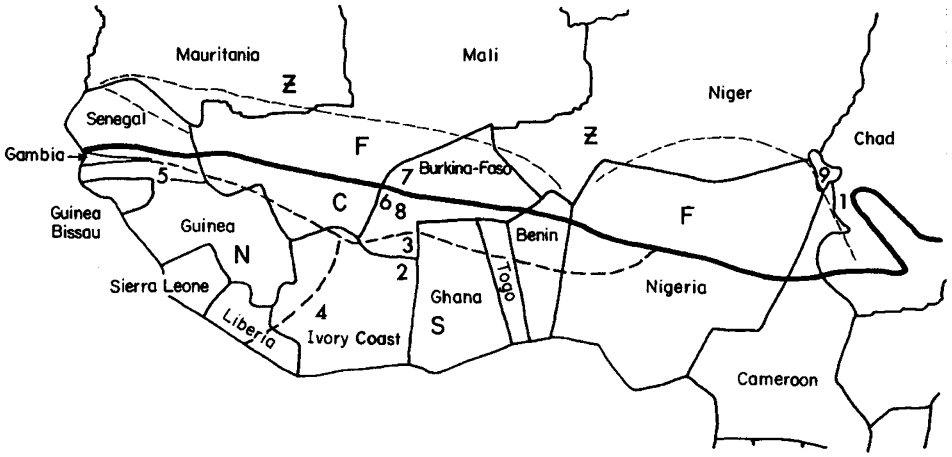

Figure 1. Location of the main West-African cattle populations and origins of milk samples collected for this study. The map is based on those of Rege et al. [34], Lhoste [16] and Epstein [7]. Heavy bold line: approximate northern limit of the Tsetse-infested zone. Dashed lines: approximate limits of the extension areas of the genetic types. N: N'Dama; S: West-African Shorthorn; Z: thoracic humped zebus; F: Fulani cattle; C: crossbreds. Origin of milk samples: Shuwa Arab: 1; Baoulé: 2, 3; N'Dama: 4, 5; Sudanese Fulani: 6–8; Kuri: 9.

2. MATERIALS AND METHODS

2.1. Equipment

The reverse phase-HPLC equipment was from Spectra Physics, San José, CA, USA; the absorbance detector (lambda Max 481) and automatic injector (712 WISP) were from Waters, Milford, MA, USA; the Nucleosil C18 N 225 column (250 × 4.6 mm, 10 nm, 5 μm) was from Shandon HPLC, Runcorn-Cheshire, UK; the Vydac C4 214TP54 column (150 × 4.6 mm; 30 nm; 5 μm) was from Touzart et Matignon, Vitry-sur-Seine, France; the FPLC system and Mono Q (HR10/10) column were from Pharmacia, Uppsala, Sweden; the amino acid analyser LC3000 was from Eppendorf-Biotronik, Maintal, Germany; the Procise 494-610A protein sequencer, 377 A automated DNA sequencer and 480 thermal cycler were from Perkin Elmer-Applied Biosystems, San José, CA, USA; the matrix-assisted laser desorption ionization linear time of flight mass spectrometer (MALDI-MS) G2025A, equipped with a Pentium PC using a sofware supplied by the manufacturer, was from Hewlett Packard, Palo Alto, CA, USA; the QIA quick PCR purification kit was from Qiagen, Courtabœuf, France.

2.2. Nomenclature

The known variants of α_{s1}-casein being A, B, C, D, E [25], F [8, 30] and G [32, 33], the additional one found in the present study was named H. In the same way, the additional variant of κ-casein was named J, next to A, B, C, D, E [25], F [14], G [9], H and I [31].

2.3. Milk samples

Individual milk samples from Shuwa Arab cows were collected in 1973 in private herds of the N'Djamena area, Chad (location 1 in *figure 1*). Samples from Baoulé cows were provided, in 1990, by the experimental farms of Minankro (IDESSA), near Bouaké, Ivory Coast ($N = 52$) and Banankeledaga (CIRDES), near Bobo-Dioulasso, Burkina Faso ($N = 46$), the animals of both herds originating from the Lobi district (locations 2 and 3). Samples from N'Dama cows were also provided in 1990 by the ranch of Marahoué (IDESSA), Ivory Coast ($N = 37$) and by the experimental farm of Kolda (ISRA), Senegal ($N = 48$) (locations 4 and 5). Samples from Sudanese Fulani cows were collected, between 1990 and 1996, from private herds from 11 villages in Burkina Faso, nine of which are located around Bobo-Dioulasso (location 8), the remaining two being more distant (locations 6 and 7). Samples from Kuri cows were collected in 1994, in private herds from the Bol district, in the Lake Chad basin (location 9). After milking, the samples were frozen until air-dispatching to the laboratory in Jouy-en-Josas. Only a few samples were not suitable for analysis.

The genotype of the Kuri cow, whose milk was used to produce α_{s1}-casein H, was homozygous for the α_{s1}-Cn^{H}, β-Cn^{A2}, κ-Cn^{A} haplotype. The genotypes of the two Baoulé cows, whose milk was used to produce κ-casein J, were α_{s1}-$Cn^{B/B}$, β-$Cn^{A1/A2}$, κ-$Cn^{B/J}$ and α_{s1}-$Cn^{B/B}$, β-$Cn^{A2/A2}$, κ-$Cn^{A/J}$, respectively, because no homozygous cow was available.

2.4. Methods

2.4.1. Electrophoresis of milk samples

Milk samples from Shuwa Arab cattle were analysed by starch gel and polyacrylamide gel electrophoresis as described by Grosclaude et al. [12]. Samples from the other populations were analysed by isoelectric focusing according to Mahé and Grosclaude [19].

2.4.2. Preparation of κ-casein

Whole casein, acid-precipitated at pH 4.6 from skim-milk, was chromatographed on a mono Q column as described by Guillou et al. [13]. The order of retention times of the non-glycosylated κ-casein fractions ($\kappa 0 - Cn$) of the three genetic variants was J < B < A. $\kappa 0 - Cn$ fractions were exhaustively dialysed against distilled water and freeze-dried.

2.4.3. Preparation of α_{s1}-casein

Whole casein, solubilized (10 g/L) in 20 mM Bis-Tris buffer pH 7.0, 4 M urea and 0.05 % DTT, was chromatographed on the C4 column (40 °C, 1 mL/min) using a linear gradient from 65 % solvent A (0.115 % TFA) and 35 % solvent B (CH_3CN/H_2O: 80/20; 0.10 % TFA) to 35 % solvent A and 65 % solvent B. The collected fractions were dried under vacuo in a speedvac (Savant Instruments).

2.4.4. Gel electrophoresis of native and renneted whole casein

Starch-gel electrophoresis of whole casein at an alkaline pH was carried out according to Aschaffenburg and Michalak [3]. Renneted samples were obtained by mixing 10 μL of a 1/50 diluted rennet solution (containing 520 mg chymosin per litre) with whole casein (24 mg/mL). Once coagulated (after 20 min at 32 °C), the samples were loaded onto the gel.

2.4.5. Preparation of deglycosylated CMP (CMP0) of the variants κ-Cn B and κ-Cn J

CMP0s B and J were prepared by a two-step precipitation of the supernatant of a chymosin hydrolysate of whole casein (κ-Cn AJ and κ-Cn BJ) with 5 and 12 % trichloracetic acid successively, according to Yvon et al. [39]. The CMP0 fraction was chromatographed at 40 °C on the C18 nucleosil column at a 1 mL/min rate, using a linear gradient from 100 % solvent A (0.115 % TFA) to 100 % solvent B (CH_3CN/H_2O/TFA 60/40/0.10 %), collected and dried with a speedvac evaporator concentrator. Retention times of the CMP0s of variants A, B and J were in the order of A < J < B.

2.4.6. Enzymatic and chemical hydrolysis

Chymosin hydrolysate (E/S: 10^{-5}) of the whole casein was performed at 37 °C for 20 min in 25 mM citrate buffer, pH 6.5. The reaction was stopped by increasing the pH to 9.0 with NaOH. α_{s1}-Casein H was hydrolysed by TPCK-treated trypsin (E/S: 0.01, W/W) at 37 °C for 18 h in 200 mM Tris-HCl buffer, pH 8.2 and the reaction was stopped by decreasing the pH to 2.0 with TFA. Endoproteinase Asp-N hydrolysis (E/S: 0.01, W/W) was performed in 50 mM sodium phosphate buffer, pH 8.0, at 37 °C overnight. CnBr cleavage (CNBr/Met: 100) was performed in 70 % formic acid at room temperature for 20 h in the dark. κ0-Casein J was hydrolysed with carboxypeptidase A (E/S: 0.015) at 40 °C for 16 h in 200 mM N-ethylmorpholine acetate buffer, pH 8.5. CMP0 was digested by *Staphylococcus aureus* protease V8 (E/S: 0.033, W/W) at 37 °C overnight in 50 mM ammonium acetate buffer, pH 4.0.

2.4.7. RP-HPLC chromatography of enzymatic hydrolysates

Tryptic and endo Asp-N hydrolysates of α_{s1}-casein H were fractionated on the C18 column (40 °C, 1 mL/min) using a linear gradient (50 min) from 100 % solvent A (0.0115 % TFA) to 60 % solvent B (CH_3CN/H_2O/TFA: 80/20/0.10 % TFA). The CNBr hydolysate was chromatographed on the C4 column (40 °C, 1 mL/min) using a linear gradient (60 min) from 80 % solvent A (0.115 % TFA) to 80 % solvent B (CH_3CN/H_2O: 90/10/0.10 % TFA). An enzymatic hydrolysate of CMP0 was run on the C18 column (40 °C, 1 mL/min) using a linear gradient from 100 % solvent A (0.115 % TFA) to 80 % solvent B (CH_3CN/H_2O: 60/40, 0.10 % TFA).

2.4.8. Molecular mass determination

The molecular masses of α_{s1}-Cn and CNBr peptides of α_{s1}-Cn were measured by MALDI-MS. First, 1 μL of the sample was mixed with 1 μL of the matrix

(sinapinic acid for α_{s1}-Cn; 4-hydroxy-α-cyano-cinnamic acid for CNBr peptides of α_{s1}-Cn). Then, 1.2 µL of the solution was deposited on the gold-10 position multiple sample probe. The droplet was allowed to dry in a vacuum, resulting in a uniform layer of fine granular matrix crystals. Proteins and peptides were desorbed and ionized (positive polarity) by a pulsed N2 laser (337 nm) with an energy of around 6 µJ. The pressure in the tube of flight (1 m in length) was about 10^{-7} Torr and the acceleration voltage of ions was 28 kV. The final mass spectrum was averaged out for about 200 simple shot spectra.

2.4.9. Polymerase chain reaction amplification and analysis of PCR products

In vitro DNA amplification was performed with the thermostable DNA polymerase of *Thermus aquaticus* in a thermal cycler [35]. A typical 50 µL reaction mixture consisted of 5 µL of 10 × PCR buffer (500 mM KCl, 100 mM Tris-HCl, 15 mM $MgCl_2$, 0.1 % (W/V) gelatin, pH 8.3), 2.5 µL of 5 mM dNTPs mix, 0.5 µL (50 pmol) of each amplimer, 0.5 µL (cDNA synthesis reaction mixture) to 1.5 µL (1.5 µg of genomic DNA) of template DNA and 0.3 µL (2.5 units) Taq DNA polymerase (Promega). To minimize evaporation loss, the mixtures were overlaid with two drops of light mineral oil. After an initial denaturing step (94 °C for 10 min), the reaction mixture was subjected, unless otherwise indicated, to the following three-step cycle which was repeated 35 times: denaturation for 1.5 min at 94 °C, annealing for 2 min at 58 °C and extension for 2 min at 72 °C. Five microlitres of each reaction mixture were analysed by electrophoresis in the presence of ethidium bromide (0.5 µg/mL), in a 2 % agarose slab gel (Appligene) in TBE buffer (1 M Tris, 0.9 M boric acid, 0.01 M EDTA).

2.4.10. DNA sequence analysis

Amplification of the genomic sequence including exons 7 and 8 was carried out using the oligonucleotide probes bov15 (5′TTATTCTTCATACCTGACTA AG 3′) and bov14 (5′ CTTAAAGCATAGAGCATATTC 3′), complementary to sequences located upstream of exon 7 and downstream of exon 8, respectively. PCR products were first purified on QIA quick spin columns and then directly sequenced according to the dideoxynucleotide chain termination procedure [36], with primer bov15 for exon 7 and primer bov 14 for exon 8, using the ABI Prism Big Dye terminator cycle sequencing ready reaction kit with Amplitaq DNA polymerase FS (Perkin Elmer). Sequencing products were analysed on polyacrylamide gel using the DNA sequencer.

3. RESULTS

3.1. Allelic and haplotypic frequencies

Among the six main lactoproteins, only α_{s2}-casein was not found to be polymorphic with the techniques used. *Table I* gives the allelic frequencies at the loci of the five other proteins, and *table II* the frequencies of haplotypes of the casein loci cluster, calculated by the method of Ceppellini et al. [6]. This

Table I. Allelic frequencies of milk protein polymorphisms in samples of three taurine and two zebu African populations. The data of Grosclaude et al. [12] on Madagascar zebus were added as a reference for *Bos indicus*.

Alleles		Populations (number of cows)					
		Taurines				Zebus	(1)
		Baoulé	N'Dama	Kuri	Sudanese Fulani	Shuwa Arab	*Madagascar*
		(96)	(75)	(169)	(1022)	(95)	(*586–778*)
	B	0.92	0.89	0.52	0.39	0.22	*0.17*
	C	0.08	0.11	0.43	0.61	0.78	*0.83*
α_{s1}-Cn	**H**	–	–	0.04	–	–	–
	x	–	–	0.01	< 0.01		–
	A^1	0.63	0.60	0.29	0.22	0.08	*0.10*
	A^2	0.36	0.37	0.62	0.71	0.86	*0.90*
β-Cn	B	0.01	0.03	0.09	0.05	0.06	–
	D	–	–	–	0.02	–	–
	A	0.34	0.27	0.74	0.69	0.83	*0.75*
	B	0.64	0.73	0.25	0.31	0.17	*0.25*
κ-Cn	C	–	–	0.01	–	–	–
	J	0.02	–	–	–	–	–
	A	0.04	0.02	0.14	0.29	0.21	*0.17*
α-La	B	0.96	0.98	0.86	0.71	0.79	*0.83*
	A	0.09	0.10	0.19	0.09	0.14	0.27
β-Lg	B	0.91	0.90	0.81	0.91	0.86	*0.73*

(1) $N = 586$ for caseins and 778 for whey proteins. In heavy bold type: additional variants characterized in the present paper; x = another additional variant, not characterized.

method assumes a Hardy-Weinberg equilibrium, a requirement that was found to be satisfied at all three individual loci in the five populations. The allelic frequencies observed in the N'Dama and Baoulé samples are remarkably similar. To gauge this similarity, the allelic frequencies at the five polymorphic milk protein loci were used to calculate the genetic distances, according to Cavalli-Sforza or Nei, between a total of 23 populations (17 French breeds, 3 African *Bos indicus* and 3 African *Bos taurus* populations including N'Dama and Baoulé). Consensus trees were built using the UPGMA method and a bootstrap procedure was carried out (for references of the methods, see Moazami-Goudarzi et al. [26]). Among all pairwise comparisons, the closest distance was indeed observed between the N'Dama and Baoulé, the bootstrap value being as high as 97 % (not shown). On the contrary, the frequencies observed in N'Dama and Baoulé showed a marked contrast to those of the two true zebu populations, Shuwa Arab and Madagascar zebu. In N'Dama and Baoulé, α_{s1}-Cn^B, β-Cn^{A1} and κ-Cn^B are the most frequent alleles compared to α_{s1}-Cn^C, β-Cn^{A2} and κ-Cn^A in zebus.

Coherently, haplotype BA^1B (a simplified designation for α_{s1}-Cn^B, β-Cn^{A1}, κ-Cn^B) is the most frequent in taurines, in contrast to CA^2A in zebus. The

Table II. Haplotypic frequencies at the α_{s1}-, β-, κ-casein cluster of genes in the same population samples as in *table I.*

(1) Haplotypes	Populations (number of cows)					
	Taurines				Zebus	
	Baoulé	N'Dama	Kuri	Sudanese Fulani	Shuwa Arab	*Madagascar*
	(96)	(75)	(169)	(1 022)	(95)	(*586*)
BA^1A	0.07	0.05	0.16	0.04	0.02	*0.02*
BA^1B	0.60	0.56	0.12	0.18	< 0.01	*0.01*
BA^1C	–	–	0.01	–	–	–
BA^2A	0.20	0.14	0.09	0.07	0.13	*0.09*
BA^2B	0.02	0.10	0.05	0.05	0.02	*0.05*
BA^2J	0.02	–	–	–	–	–
BBA	< 0.01	0.02	0.07	0.03	0.05	–
BBB	< 0.01	0.01	0.02	0.02	–	–
BDB*	–	–	–	< 0.01	–	
CA^1A*	–	–	–	< 0.01	–	*0.06*
CA^1B*	–	–	–	–	0.06	*0.02*
CA^2A	0.07	0.06	0.38	0.53	0.63	*0.59*
CA^2B	0.01	0.06	0.05	0.05	0.09	*0.17*
CBA*	–	–	–	< 0.01	–	–
CBB*	–	–	–	< 0.01	–	–
CDA	–	–	–	0.02	–	–
CDB	–	–	–	< 0.01	–	–
HA^2A	–	–	0.04	–	–	–
xA^2A	–	–	–	< 0.01	–	–
xA^2B	–	–	0.01	< 0.01	–	–

(1) For BA^1A, read α_{s1}-Cn^B, β-Cn^{A1}, κ-Cn^A, etc.; * recombinant haplotypes between the tightly linked α_{s1}-Cn and β-Cn alleles.

values in Sudanese Fulani cattle show a zebu-like pattern, but the rather high frequency of haplotype BA^1B may be considered as revealing the influence of *Bos taurus* genes in the origin of this cattle type. In Kuri, allele α_{s1}-Cn^B prevails over α_{s1}-Cn^C, which is a taurine feature. The predominant haplotype is, however, CA^2A, as in zebus, and overall, the Kuri appears as an almost perfect intermediate between taurines and zebus. In contrast with a majority of west European breeds, α-lactalbumin is also polymorphic in taurines, but the frequencies of α-La^A are significantly lower than in zebus.

The occurrence of three additional variants was suspected at the α_{s1}-Cn and κ-Cn loci. Two of them could be characterized by biochemical analyses summarized hereafter and were given a regular designation: variant α_{s1}-Cn H was found in 16 Kuri cows (nine C/H, six B/H, one H/H), and variant κ-Cn J in three Baoulé cows (one A/J, two B/J) and one Fulani, not belonging to the population sample (B/J) (*figure 2*).

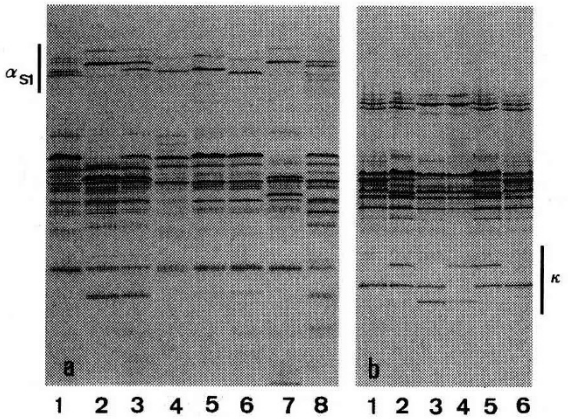

Figure 2. Isoelectric focusing of individual milk samples showing the α_{s1}-Cn H and κ-Cn J variants. **a**) The α_{s1}-casein genotypes of samples 1–8 are as follows: 1, C/H; 2, B/B; 3, B/C; 4 and 5, C/C; 6, H/H; 7, B/B; 8, B/x. **b**) The κ-casein genotypes of samples 1–6 are as follows: 1, B/B; 2, A/B; 3, B/J; 4, A/J; 5, A/B; 6, B/B.

3.2. Characterization of variant α_{s1}-Cn H

Mass spectrometry analysis of purified α_{s1}-caseins B and H gave a value of 22 691.6 Da for variant H as compared to 23 613 Da for variant B (theoretical value: 23 615.8). The difference of 921.4 Da was indicative of a deletion of about eight amino acid residues. RP-HPLC elution patterns of tryptic hydrolysates of α_{s1}-caseins B and H showed the absence of one peak in α_{s1}-casein H as the only difference. The fraction corresponding to this missing peak was identified, by Edman degradation, as the peptide 43–58 of α_{s1}-casein B, which suggests that the difference is located in this region (*figure 3*).

The sequence of the first 52 residues of the α_{s1}-casein H protein was established unambiguously by Edman degradation. This sequence was identical

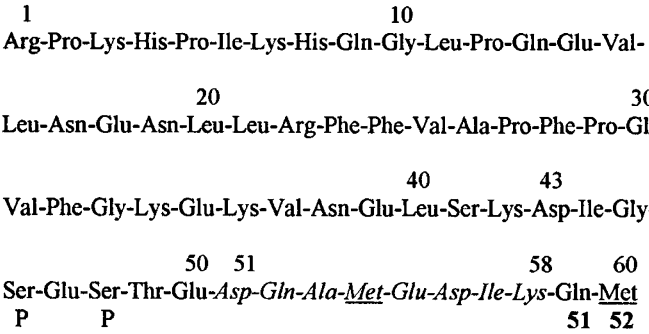

Figure 3. Sequence of the first 60 amino acid residues of α_{s1}-casein B. The Met residues are underlined. The sequence of eight residues (51–58) which is deleted in variant H is in italics. After residue 58, the sequence of the H protein is numbered below, in heavy bold type.

to that of variant B, up to residue 50, but the next two residues (51–52) were Gln-Met, instead of Asp-Gln in variant B. This result indicated the deletion, in the H protein, of residues 51–58 (*figure 3*) since, in the reference variant B, the Gln-Met sequence occurs at positions 59–60. MALDI-MS analysis of the purified CNBr peptides, carried out for confirmation, showed a difference for only one peptide. The molecular mass found for this peptide was 6 300 Da in α_{s1}-Cn B, which corresponds to peptide 1–54 (theoretical mass: 6 285.12 Da). In the α_{s1}-Cn H variant, the measured molecular mass was 6 114 Da which is very close to the theoretical value obtained for the sequence 1–52 (6 099.1 Da). Altogether, these results established the deletion in variant H of the sequence of eight residues (51–58) which is coded by exon 8.

The non-deletion of exon 8 in the α_{s1}-Cn H gene could be ascertained by sequencing the product of PCR amplification of the corresponding region of the gene, including exons 7 and 8, carried out on the DNAs of two heterozygous cows, of genotypes α_{s1}-$Cn^{B/H}$ and α_{s1}-$Cn^{C/H}$. In both cases, only the normal, non-deleted sequence was obtained. This suggested that the deletion of peptide 51–58 was the consequence of exon skipping. The Edman degradation of the endo Asp-N C-terminal peptide of α_{s1}-Cn H showed that this variant had the same C-terminal sequence as α_{s1}-casein B (Glu in position 192 instead of Gly for α_{s1}-Cn C). Allele α_{s1}-Cn^{H} is thus derived from allele α_{s1}-Cn^{B}.

3.3. Characterization of variant κ-Cn J

The patterns observed in *figure 2* strongly suggest that κ-Cn J has either one negative charge less or one positive charge more than κ-Cn B. The para-κ-caseins obtained by chymosin hydrolysis of κ-caseins A, B and J showed an identical electrophoretic migration, indicating that the particularity of the type J casein was to be searched in the caseinomacropeptide. RP-HPLC elution patterns of B and J CMP0s digested with *S. aureus* protease V8 showed a single difference in the fractions (data not shown, available upon request). Partial Edman degradation of the fraction corresponding to the extra peak of type J gave a sequence corresponding to residues 152–164 of κ-casein B, except that Arg replaced Ser at position 155 (Val-Ile-Glu-*Arg*-Pro-Pro-Glu-Ile-Asn-Thr-Val-Gln-Val OH). The results from CPA degradation of κ0-Cn J showed no difference in its C-terminal part, which also corresponds to the C-terminal parts of CMPs and κ-caseins A and B [Thr: 2.68 (3); Ser: 1.25 (1): Ala: 0.67 (1); Val: 2.35 (3); Ile: 1.00 (1); Asn: 1.16 (1); Gln: 0.88 (1)].

In conclusion, the difference in charge between variants κ-Cn B and κ-Cn J is the consequence of a substitution of Ser 155 (B) → Arg (J) due to a mutation (AGC → AGA or AGG) having occurred in the κ-Cn^{B} allele. This result is consistent with the order of elution of variants A, B and J from an anion exchange mono Q column, as given above (J < B < A).

4. DISCUSSION

The number of individual milk samples analysed varied markedly in the populations studied. A large number of samples makes the detection of rare alleles and haplotypes possible. This is well exemplified in the Fulani sample where the BDB, $CA^{1}A$, CBB and CBA haplotypes are rare recombinants

between the α_{s1}-Cn and β-Cn loci. If a large number of samples had been available, such rare recombinants may also have been found in the other populations. Taking into account this disequilibrium in sample sizes, the comparison of populations should only be based upon the main alleles and haplotypes.

When comparing the protein polymorphism of geographically distant populations, the risk of homoplasy (different variants having the same electrophoretic behaviour) should be taken into consideration. In their work on Madagascar zebu, Grosclaude et al. [12] ascertained that the amino acid substitutions responsible for the differences in charge between variants α_{s1}-Cn B and C, β-Cn A^1 and A^2, as well as κ-Cn A and B, were the same in this zebu as in western humpless cattle. In the present study, it was also checked, by DNA typing [17], that the β-Cn A^1 variant of the N'Dama and the κ-Cn C variant of the Kuri were identical to their European counterparts (data not shown). Because κ-Cn C occurs at low frequencies in European breeds, its presence in the Kuri was unexpected. Since there is no record of any introduction of European cattle in the Kuri, κ-Cn C may be regarded, as κ-Cn A and B, as being common to European and African cattle, or at least *Bos taurus* populations.

The extension of research on the polymorphism of milk proteins to the so-far neglected African cattle populations was expected to lead to the discovery of additional variants, specific to these populations. Three such unknown variants were observed, and two of them were characterized, α_{s1}-Cn H and κ-Cn J, the amount of the casein sample available being insufficient to characterize the third one.

The α_{s1}-Cn H variant is the most interesting, since it is the fifth example of a casein variant having an internal deletion, most likely generated by exon skipping. Except for the D variant of bovine α_{s2}-casein [5], the other examples are all concerned with variants of either bovine or caprine α_{s1}-casein. The affected exons are exon 4 in bovine α_{s1}-casein A [27, 38] and caprine α_{s1}-casein G [23], and exons 9, 10 and 11 in caprine α_{s1}-casein F [15]. Still another exon, exon 8, is involved in bovine α_{s1}-casein H.

The difference in charge between variants κ-Cn B and κ-Cn J is due to a single amino acid substitution, Ser 155 (B) $\rightarrow$ Arg (J). Interestingly, position 155 is the same as that affected in the κ-Cn E variant [25] but, in this case, the mutation (AGC $\rightarrow$ GGC) occurred in the A allele, inducing the amino acid substitution Ser 155 (A) $\rightarrow$ Gly (E).

The most striking result in *tables I* and *II* is the exceptional similarity of allelic and haplotypic frequencies in the N'Dama and Baoulé samples. In both cases, sampling was carried out in two experimental herds and not in the field populations, but it is difficult to believe that this procedure, which is not quite appropriate, could account for such a similarity. Frequencies observed in the Shuwa Arab zebu are not very different from those of the Madagascar zebu. The inversion of allelic frequencies between taurines and zebus, already well documented in the literature for the α_{s1}-Cn locus [24, 28], may also be observed here at the β-Cn and κ-Cn loci. The polymorphism of casein loci thus appears to be particularly useful for population studies in the African continent, where, for a long time, taurines have coexisted and intercrossed with zebus. The fact that the frequencies in Sudanese Fulani cattle are intermediate between those

of taurines and zebus, and closer to the latter, is in good agreement with the above-mentioned theory about the origins of this cattle type.

The Kuri, which is humpless and has the small metacentric chromosome of *Bos taurus* [29], is thus considered as a taurine population. It was classed by Epstein [7], together with the N'Dama, in the humpless longhorn cattle group, descending from the first domesticated *Bos* introduced into Africa. Surprisingly, its allelic frequencies are not far from the mean of those of the N'Dama and of the true zebus, except for the κ-casein polymorphism which is closer to that of zebus. This situation may be due to introgression of zebu genes into the breed. MacHugh et al. [18] have indeed concluded to the existence of an east to west introgression gradient of microsatellite alleles of Indian *Bos indicus* into African populations, including sub-populations of the taurine type N'Dama, but the Kuri was not included in their study. As a matter of fact, crossbreeding with Shuwa Arab and M'Bororo zebus is a common practice in the areas fringing Lake Chad, while pure Kuri are restricted to the islands [22]. The frequencies of milk protein polymorphisms were, however, found to be exactly the same in a group of 103 cows considered, on the basis of phenotypic characters, to be pure Kuri, as in a group of 63 cows which, although humpless, could be suspected to carry zebu genes [37]. The possible gene flow between Kuri and zebus is thus not easily detectable in the present conditions.

When considering the allelic differences between N'Dama and Kuri, one should remember that these two populations most likely originated from two distinct routes of introduction of domesticated humpless longhorn cattle into Africa. The ancestors of the N'Dama probably spread through northern and north-western Africa, and those of the Kuri through the Sahara during the 'green period'. The genetic differences between N'Dama and Kuri could thus have a quite remote origin. It is conceivable that the Kuri remained genetically closer to the originally domesticated population of South-West Asia, the common origin of all taurines than the N'Dama did [7]. The gene frequencies in this population were probably closer to those of the zebus than those of the modern taurines of northern Europe and western Africa. This is supported by the distribution of the α-lactalbumin polymorphism in Europe. While the α-lactalbumin polymorphism is the rule in zebus, it is restricted, in Europe, to Podolic breeds, or breeds known to have been crossed with Podolic cattle, all of which are located in southern Europe [21, 24, 28]. Because the longhorn Podolic group is genetically and geographically the closest in Europe to the original domestic population of South-West Asia [20], it can be assumed that α-lactalbumin was polymorphic in this original population, which is more a *Bos indicus* than a *Bos taurus* feature. The phylogenetic status of the Kuri will be analysed with more genetic markers in a forthcoming paper.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to all those who took part in the collection of milk samples and to their institutions. They thank Patricia Anglade and Patrice Martin for their help, Georges Tacher for his personal communication regarding past history of the Kuri, and André Ng-Kwai-Hang for revision of the manuscript.

REFERENCES

[1] Aschaffenburg R., Genetic variants of milk proteins: their breed distribution, J. Dairy Res. 35 (1968) 447–460.

[2] Aschaffenburg R., Drewry J., Genetics of the β-lactoglobulins in cow's milk, Nature 180 (1957) 376–378.

[3] Aschaffenburg R., Michalak W., Simultaneous phenotyping procedure for milk proteins. Improved resolution of the β-lactoglobulins, J. Dairy Res. 51 (1968) 1849.

[4] Aschaffenburg R., Sen A., Thompson M.P., Genetic variants of casein in Indian and African Zebu cattle, Comp. Biochem. Physiol. 25 (1968) 177–184.

[5] Bouniol C., Printz C., Mercier J.C., Bovine α_{s2}-casein D is generated by exon VIII skipping, Gene 128 (1993) 289–293.

[6] Ceppellini R., Siniscalco M., Smith C.A.B., The estimation of gene frequencies in a random-mating population, Ann. Hum. Genet. 20 (1956) 97–115.

[7] Epstein H., The Origin of the Domestic Animals of Africa, Africana Publishing Corporation, New York, vol. I, 1971.

[8] Erhardt G., A new α_{s1}-casein allele in bovine milk and its occurrence in different breeds, Anim. Genet. 24 (1993) 65–66.

[9] Erhardt G., Detection of a new κ-casein variant in milk of Pinzgauer cattle, Anim. Genet. 27 (1996) 105–107.

[10] Grosclaude F., Le polymorphisme génétique des principales lactoprotéines bovines. Relations avec la quantité, la composition et les aptitudes fromagères du lait, Inra Prod. Anim. 1 (1988) 5–17.

[11] Grosclaude F., Aupetit R.Y., Lefebvre J., Mériaux J.C., Essai d'analyse des relations génétiques entre les races bovines françaises à l'aide du polymorphisme biochimique, Genet. Sel. Evol. 22 (1990) 317–338.

[12] Grosclaude F., Mahé M.F., Mercier J.C., Ribadeau-Dumas B., Comparaison du polymorphisme génétique des lactoprotéines du zébu et des bovins, Ann. Génét. Sél. Anim. 6 (1974) 305–329.

[13] Guillou H., Miranda G., Pélissier J.P., Analyse quantitative des caséines dans le lait de vache par chromatographie liquide rapide d'échange d'ions (FPLC), Lait 67 (1987) 135–148.

[14] Ikonen T., Ruottinen O., Erhardt G., Ojala M., Allele frequencies of the major milk proteins in the Finnish Ayrshire and detection of a new κ-casein variant, Anim. Genet. 27 (1996) 179–181.

[15] Leroux C., Mazure N., Martin P., Mutations away from splice sites recognition sequences might cis-modulate alternative splicing of goat α_{s1}-casein transcripts. Structural organization of the relevant gene, J. Biol. Chem. 267 (1992) 6147–6157.

[16] Lhoste P., Cattle genetic resources of West Africa, in: Neiman-Sorensen A., Tribe D.E. (Eds.), World Animal Science, Vol. B7, Elsevier, Amsterdam, 1991, pp. 73–89.

[17] Lien S., Aleström P., Klungland H., Rogne S., Detection of multiple β-casein (CAS B) alleles by amplification created restriction sites (ACRS), Anim. Genet. 23 (1992) 333–338.

[18] MacHugh D.E., Shriver M.D., Loftus R.T., Cunningham P., Bradley D.G., Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*), Genetics 146 (1997) 1071–1086.

[19] Mahé M.F., Grosclaude F., Polymorphism of β-casein in the Creole goat of Guadeloupe: evidence for a null allele, Genet. Sel. Evol. 25 (1993) 403–408.

[20] Manwell C., Baker C.M.A., Chemical classification of cattle. 2. Phylogenetic tree and specific status of the Zebu, Anim. Blood Grps Biochem. Genet. 11 (1980) 151–162.

[21] Mariani P., Russo V., Polimorfismo genetico della α-lattalbumina nelle razze bovine, Riv. Zootec. Vet. 5 (1977) 603–613.

[22] Martin A., Bonnet P., Bourzat D., Lancelot R., Souvenir Zafindrajaona P., Importance de l'élevage et sa place dans l'économie des pays de la Commission du Bassin du Lac Tchad, in: Atlas d'élevage du bassin du lac Tchad, CIRAD-EMVT, Montpellier, 1996, pp. 79–86.

[23] Martin P., Leroux C., Characterization of a further α_{s1}-casein variant generated by exon-skipping, in: Proc. XXIV Int. Conf. on animal genetics, Prague, 23–29 July 1994, p. 86.

[24] Mercier J.C., Grosclaude F., Génétique moléculaire des protéines du lait et de leurs gènes, in: Martinet J., Houdebine L.M. (Eds.), Biologie de la lactation, Inserm/Inra Paris, 1993, pp. 319–347.

[25] Miranda G., Anglade P., Mahé M.F., Erhardt G., Biochemical characterization of the bovine genetic κ-casein C and E variants, Anim. Genet. 24 (1993) 27–31.

[26] Moazami-Goudarzi K., Laloë D., Furet J.P., Grosclaude F., Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites, Anim. Genet. 28 (1997) 338–345.

[27] Mohr U., Koczan D., Linder D., Hobom G., Erhardt G., A single point mutation results in A allele-specific exon skipping in the bovine α_{s1}-casein mRNA, Gene 143 (1994) 187–192.

[28] Ng-Kwai-Hang K.F., Grosclaude F., Genetic polymorphism of milk proteins, in: Fox P.F. (Ed.), Advanced Dairy Chemistry, Elsevier Applied Science, London, 1992, pp. 405–455.

[29] Petit J.P., Queval R., Le Kouri : race bovine du lac Tchad. II. Étude biochimique : les hémoglobines et les constituants du sérum, Rev. Elev. Méd. Vét. Pays Trop. 26 (1973) 97–104.

[30] Printzenberg E.M., Anglade P., Ribadeau-Dumas B., Erhardt G., Biochemical characterization of bovine α_{s1}-casein F and genotyping with sequence-specific primers, J. Dairy Sci. 65 (1998) 223–231.

[31] Printzenberg E.M., Erhardt G., High-resolution SSCP analysis reveals new alleles at the κ-casein (CSN3) locus in *Bos taurus* and *Bos indicus* cattle, in: Proc. XXVI Int. Conf. on animal genetics, Auckland, 9–14 August, 1998, p. 17.

[32] Rando A., Mariani P., Fiorella A., Di Gregorio P., Ramunno L., Masina P., Un allele quantitativo della caseina α_{s1} di bovino, Atti Congr. Naz. Assoc. Sci. Prod. Anim. 11 (1995) 175–176.

[33] Rando A., Ramunno L., Di Gregorio P., Davoli R., Masina P., A rare insertion in the bovine α_{s1}-casein gene, Anim. Genet. 23 (1992) suppl. 1, p. 55.

[34] Rege J.E.O., Aboagye G.S., Tawah C.L., Shorthorn cattle of West and Central Africa. I. Origin, distribution, classification and population statistics, World Anim. Rev. 78 (1998) 2–13.

[35] Saiki R.K., Gelfand D.H., Stoffel S., Scharf S.J., Higuchi R., Horn G.T., Mullis K.B., Erlich H.A., Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase, Science 239 (1982) 487–491.

[36] Sanger F., Nicklen S., Coulson A.R., DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

[37] Souvenir Zafindrajaona P., Zeuh V., Volet de travaux génétiques, in: Projet « Conservation de la race bovine Kouri dans le bassin conventionnel du lac Tchad », rapport final, 16-66, Laboratoire de recherches vétérinaires et zootechniques de Farcha, BP 433, N'Djamena, Tchad, 1996.

[38] Wilkins R.J., Xie Tao., Two distinct gene mutations - one milk protein polymorphism. The example of the α_{s1}-casein A variant, in: Abstracts of the IDF Milk Protein Polymorphism II Seminar, Palmerston North, New Zealand, 24–27 February, 1997.

[39] Yvon M., Chabanet C. Pélissier J.P., Solubility of peptides in trichloracetic acid (TCA) solution: hypothesis on the precipitation mechanism, Int. J. Pept. Protein Res. 34 (1989) 166–176.