MILK PROTEIN POLYMORPHISM: DETECTION AND DIFFUSION OF THE GENETIC VARIANTS IN BOS GENUS

P. Formaggioni, A. Summer, M. Malacarne, P. Mariani

(1) The work was supported by the experimental program of the Emilia-Romagna region, with the technical-organising coordination of the Centro Ricerche Produzioni Animali of Reggio Emilia.

(2) Istituto di Zootecnica, Alimentazione e Nutrizione, Università degli Studi, Via del Taglio 8, 43100 Parma.

Introduction

Studies on the milk protein system have been in progress for more than 100 years, and constituted a rather difficult matter, because of the intrinsic complexity of the subject and because of experimental methods not ever adequate to verify preliminary hypothesis. The improving of the experimental procedures and the introduction of new analytical methods (e.g. new electrophoretical or chromatographic techniques), together with the contributions of sciences in progress like molecular biology, genetics and biochemistry, have lead great benefits to milk protein knowledge.

Milk proteins are usually divided into two great “historical” groups, depending on their behaviour by acidification at pH 4.6. The soluble fraction, named “whey protein”, is constituted by several different proteins, the most important ones are \( \alpha \)-lactalbumin (\( \alpha \)-La) and \( \beta \)-lactoglobulin (\( \beta \)-Lg). The fraction insoluble, named “whole casein”, is constituted of four different native caseins (Cn): \( \alpha s1 \)-Cn, \( \alpha s2 \)-Cn, \( \beta \)-Cn and k-Cn; these proteins are associated with a variable number of phosphate groups and, in the case of k-casein, with a carbohydrate moiety.

\( \alpha s1 \)-, \( \alpha s2 \)- and \( \beta \)-caseins, richer of phosphate groups, are distinguishing from k-casein for their more or less marked tendency to “precipitate” in presence of calcium ions. k-casein is constituted of two different moieties with regard to their solubility: one of these (1-105 aa.) is characterised by the presence of hydrophobic residues, the other (106-169 aa.), to which are attached also the carbohydrate groups, manifests a marked hydrophilic nature.

Because of these ampholitic properties, k-casein plays a role of colloid-protector towards the other caseins, and constitutes with them a dispersion in the milk by the formation of micelles; each of these is composed by several thousand molecules of all the four
caseins. The micelles maintain the hydrophobic portions protected inside, and expose outside the hydrophilic moieties of the k-casein.

**Genetic polymorphism**

The term "genetic polymorphism" defines the fact that each milk protein presents two or more forms genetically determined by autosomal and codominant alleles. The absence of dominance is very useful, because homozygous individuals present in the electropherogram only one variant for each protein, while heterozygous ones both variants, so that the count of the gene frequencies for a population results very easy.

Studies on milk protein polymorphism have been developed with various different finalities; to point out milk protein chemical evolution and find some eventual similarity with other proteins; to verify relationships between different species or breeds; to monitor variations that happen in the time or in the space for a particular animal population; to understand, and this is the most important aim, the biological significance of genetic variants.

Genetic polymorphism, however, has great significance also in applied fields, like zootechnical sciences or dairy industry: in particular the study is focused to clarify eventual associations between genetic variants and production traits, reproduction efficiency and adaptation capacity of the cattle, as well as detection of eventual influences on milk nutritional and technological properties [1]. The relationships between protein genetic types and milk composition and properties have been elucidated by several Authors [2-5]; in particular has been evidenced that milks characterised by B variants of $\beta$-lactoglobulin, k-casein and $\beta$-casein present a nitrogen composition and/or rennet-coagulation properties that are better with respect to those characterised by A types and, therefore, are more favourable for the cheesemaking.

**Some furthermore remarks**

The convention for the nomenclature of the variants for each protein is a progressive alphabetical order corresponding to the chronological order of the discoveries (at the beginning it was also referred to the electrophoretical mobility of the bands). The nomenclature is unified for the four species of *Bos* genus, i.e. *B. taurus* ("common" bovine), *B. indicus* (zebu), *B. grunniens* (yak), *B. javanicus* (banteng of Bali).

Some exceptions are made for the progressive alphabetical nomenclature: sometimes it's difficult to reconstruct the history of some variants, because different variants are named with the same letter or, on the contrary, the same variant had received two
different "letters"; there are several variants not well or not still characterised, for which is not known the aminoacid substitution or its position in the molecule.

The term "genetic variant", traditionally used for the coded mature protein, actually is referred indifferently to the mature protein and to the coding gene; how must be considered those cases where the mutation is in a not-coding moiety of the gene (e.g. not-coding exons, introns, 5' and 3' flanking regions)? The coded protein is equal to the common one, but often is synthesised in a quantitatively different way, because the mutation causes a change in a regulatory sequence of the gene. Some of them have been accepted in the official nomenclature (like \( \alpha_{s1}\)-casein G), some other of these are not considered variants in a true sense.

\( \alpha_{s1}\)-casein system

"\( \alpha_{s1}\)-casein" consists of one major and one minor component, both with the same aminoacid sequence; the minor component (\( \alpha_{s0}\)-casein) contains one additional phosphate group linked to the serine at position 41. The sequence of \( \alpha_{s1}\)-casein was first established by Mercier et al. [6] and Grosclaude et al. [7]: the primary structure of the most common variant B consists of 199 aminoacid residues, with a molecular weight of 23,614. The number of acid (7 Asp, 25 Glu, 8 P) residues is higher than that of basic (14 Lys, 6 Arg, 5 His) ones: so the isoelectric point is rather low (4.1-4.5). The high amount of non-polar residues makes the protein rather hydrophobic, but less than \( \beta\)-casein. The most hydrophilic region is between 45th and 89th residues; here are concentrated most of acid residues and 6 phosphate groups: in \( \beta\)-casein there is a similar disposition [8]. Probably because of its hydrophobic nature, attempts to crystallise \( \alpha_{s1}\)-casein (as been as for the other caseins) have been not successful: therefore, until now, secondary and tertiary structures of this protein are not known. Proline (17 residues) is very diffused and uniformly distributed: this makes less probable the presence of eventual regular secondary structures [9, 10].

Polymorphism – In 1962, by starch gel electrophoresis at alkaline pH, Thompson et al. [11] demonstrate for the first time a polymorphism for \( \alpha_{s1}\)-casein. The two electrophoretic bands were named A and B, in order of their decreasing mobility: B variant is diffused in all the species; A, instead, is a not common variant, found in Holstein breed; it is characterised by a deletion of 13 aminoacids, from 14 to 26, with respect to the B variant [12]. The deletion of the mature protein arises from a corresponding absence of 39 bp (the entire exon 4) in the mRNA, but not in the gene, as was suggested by McKnight et al. [13] and proved by Mohr et al. [14]. In fact, the deletion occurs at level of mRNA splicing after transcription, and is correlated with a base substitution at position +6 of the intron 4, in the splice donor sequence distal of
exon 4; this mutation gives rise to the exon skipping during the splicing of the A allele mRNA [14]. McKnight et al. [13] found also 5 silent mutations in the coding sequence of A variant with respect to B variant: all of them consist of C↔T conversions at level of the third base of the codon, silent mutations that don’t change the coded aminoacid. Only three of these mutations regard the mature protein: Pro(2), Ala(163) and Pro(185); the other two are in the region cut away in the post-translational events.

Recently Wilkins and Xie [15] found a mutation in αs1-Cn A gene, still in intron 4 but in a different position, at +4 instead that +6. This mutation gives rise to the same αs1-Cn A variant with the deletion of exon 6.

The phenomenon of deletion in protein variants is rare; until now three example are known in Bos genus: besides the “hystorical” ones αs1-Cn A and αs2-Cn D, recently Mahé et al. [16] found in African Kuri (B. taurus) another αs1-Cn variant with a deletion of 8 aminoacids (51-58: the entire exon 8), named αs1-Cn H. Further work is necessary to establish the causes, at mRNA and genomic level, that give rise to this deletion.

The principal variants of αs1-Cn (from A to H), are reviewed in Table 1; the identity of Eyak variant found in B. grunniens by Grosclaude et al. [19, 20] and EBali found in B. javanicus (banteng of Bali) by Bell et al. [21] is not well established: the aminoacid substitution of the latter is still unknown.
In 1992, Kawamoto et al. [43] found in Nepalese \textit{B. taurus} and \textit{B. taurus x B. grunniens} crosses two further alleles of $\alpha_{s1}$-casein that they named tentatively X and Y, respectively. In isoelectric focusing X variant migrates between B and C; Y faster than B. No more works were made to furthermore characterise these eventual variants.

A particular emphasis must be reserved to \textit{\$\alpha_{s1}$-Cn G}. This "variant", found in Italian Brown and Podolian breeds by Rando et al. [24-26], represents an \textit{unicum} in cow milk polymorphism; in fact the mature protein is not different from the common B variant, but is synthesised with a lower amount than $\alpha_{s1}$-Cn B. This quantitative major effect is caused by an insertion of 371 bp at level of the 19th (not-coding) exon, exactly between 17281 and 17282 bp of \textit{\$\alpha_{s1}$-Cn B} gene [44]. This inserted segment has the typical structure of relics of long interspersed elements (LINEs) of retropositional origin [45].
The bovine αs1-Cn G allele is the second example of insertion of a genetic mobile element that has a marked effect on the expression of a casein gene; in fact in goat milk was found [46, 47] an analogous allele, named αs1-Cn E, with an insertion at level of 19th exon, but in another position; the inserted segment, however, has a great homology with that of the cow, and the effects on protein synthesis are the same.

**Diffusion** – The most diffused αs1-casein variant is B, present in all breeds with a frequency of 90-95% and sometimes of 100%; only in some breeds like Jersey, Guernsey, Normande, Italian Brown, Reggiana and Modenese the frequency is a little lower (75-85%), favouring C variant. In zebu and yak, conversely, allele C is predominant with respect to B, with about 90% of frequency in the first and about 63% in the latter [20]. Surprisingly C variant has an high frequency in Swedish Holstein [48].

αs1-Cn D variant is not very common, but, after Flamande [18], was observed also in Red Danish and Red Polish [34], Jersey [35], Italian Brown [36, 37], Reggiana [38], Podolian [39] and other Italian breeds.

Also for the A allele, after the discovery in Holstein Friesian, other detections were made in Red Danish [27], Kostroma [28], other Friesian strains [28-32] and more recently in German Friesian [33].

αs1-Cn G allele was observed also in other Italian breeds, like Agerolese and Modicana [25, 40, 41] and recently in Reggiana, Italian Red Pied, Sarda and Bruna-Sarda [42].

**αs2-casein complex**

"αs2-casein" is a family of protein, including αs2, αs3, αs4, αs6, with the same aminoacid sequence but a different content (13, 12, 11 and 10, respectively) of phosphate groups. The sequence of αs2-Cn A, the most diffused variant, was first established by Brignon et al. [49]. It consists of 207 aminoacid residues, with a molecular weight from 25,150 to 25,390, depending on the relative number of phosphate groups. The presence of 2 cysteins, the scarce amount of proline and a lower hydrophobic nature characterise αs2 with respect to the other caseins. Its susceptibility to calcium ions in relation to its high phosphorylation degree is higher than that of the other caseins. Two segments (50-123 and 132-207) in the αs2-Cn peptide chain present a marked homology; this suggests that αs2 arises from the duplication of a primitive gene [50].

**Polymorphism and diffusion** – Only four variants are known for αs2-Cn: the first example of polymorphism was evidenced in 1976 by Grosclaude et al. [51] in Nepalese B. taurus and B. indicus populations; in these species both A and B variants were found. In the
same paper Grosclaude et al. [51] noticed the discovery of a third variant (C) in Mongolian yak. $\alpha_s2$-Cn C is peculiar of $B. grunniens$, while B variant has never been detected in this species (Table 2).

Table 2 – Discovery and diffusion of the genetic variants of $\alpha_s2$-casein.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Year</th>
<th>Species (country)</th>
<th>Breeds (frequency)</th>
<th>Methods</th>
<th>Diffusion (1)</th>
<th>Variations with respect to A</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>(*) A</td>
<td>1976</td>
<td>$B. taurus$ (H)</td>
<td></td>
<td></td>
<td>All species</td>
<td>All breeds</td>
<td>Grosclaude et al., 1976</td>
</tr>
<tr>
<td>*E</td>
<td>1976</td>
<td>$B. indicus$ (H)</td>
<td>$E. indicus$ (0.34)</td>
<td>SGE pH 8.6</td>
<td>Less common</td>
<td>$B. taurus$ (2) $B. indicus$ (4) $B. grunniens$</td>
<td>Grosclaude et al., 1976</td>
</tr>
<tr>
<td>C</td>
<td>1976</td>
<td>$B. grunniens$ (M)</td>
<td>(0.23)</td>
<td>SGE pH 8.6</td>
<td>Not common</td>
<td>$Glu$ (33) $\rightarrow$ $Gly$ $Ala$ (47) $\rightarrow$ $Thr$ $Thr$ (130) $\rightarrow$ $Ile$</td>
<td>Grosclaude et al., 1976</td>
</tr>
<tr>
<td>D</td>
<td>1976</td>
<td>$B. taurus$ (F)</td>
<td>Montbèrard (0.011), Vosgienne (0.088)</td>
<td>SGE pH 8.6</td>
<td>Less common</td>
<td>Other breeds (3) Deleteon 51-59 aa</td>
<td>Grosclaude et al., 1976 [52]</td>
</tr>
</tbody>
</table>

* Still not characterized variant
(1) Analysis not extended to $B. javanicus$.
(2) Reference (most common) variant.
(3) In Podolian with frequency 0.004 [39].
(4) In South African zebu [52, 53].
(5) In German breeds: Gelbvieh, Vorder Walder, Hinter Walder, Fleckvieh, Braunvieh [33]; in Finnish Ayrshire [54].

The first evidence of the presence of B variant in Western $B. taurus$ breeds was proved by Chianese et al. [39] that found this allele in Podolian cattle. No furthermore works were made to individuate the aminoacid substitutions that differentiate B variant from A: $\alpha_s2$-Cn B is still now a not characterised variant.

The last discovered variant, $\alpha_s2$-Cn D, found in two French bovine breeds [51, 52], is characterised by a deletion of 9 aminoacids (51-59), corresponding to the entire exon 8. As well as for $\alpha_s1$-Cn A, the deletion in mRNA is caused by a single point mutation that gives rise to the exon 8 skipping during mRNA splicing. Differently from $\alpha_s1$-Cn A, the single point mutation (G$\rightarrow$T) is not located in an intron, but in exon 8 last nucleotide, that represents the 5'-consensus splicing site: this place, by effect of mutation, cannot be recognised [55]. D allele was found also in some German breeds [33] and in Finnish Ayrshire [54].

There are other cases of probable polymorphism for $\alpha_s2$-Cn, that have not been substantiated by further information. In 1967, for example, Michalak [56] found in Red Danish three individual samples characterised by the complete absence of "bands 1.00
and 1.04", that correspond to the actual $\alpha s2$-Cn. Is this possibly the first example of null-allele in cow? No further investigations were made.

In 1979, Merlin and Di Stasio [57] found in Pinzgauer cattle $\alpha s2$-Cn bands with a lower mobility in comparison to those of $\alpha s2$-Cn A. It is possible that such bands were those of B variant, that in those times seemed to be absent in Western B. taurus; also in this case no further investigations were made.

**$\beta$-casein system**

The primary sequence of $\beta$-casein has been elucidated by Ribadeau-Dumas *et al.* [58] and Grosclaude *et al.* [7]. This protein consists of 209 residues with a molecular weight of 23,983. Several are the analogies with $\alpha s1$-casein: there are not cystein residues, while proline residues are very common; the molecular weight is similar. It shows a marked hydrophobic character and, at room temperature, it is sensitive to calcium ions. There is also an high homology between the sequences of the two proteins; in particular two peptides of eight aminoacids (63-70 for $\alpha s1$-casein; 14-21 for $\beta$-casein) are very similar [8]. By action of plasmin, $\beta$-casein can be cleaved in three different positions, to give rise to the $\gamma$-caseins and to their complementary fragments, named proteosepeptones.

*Polymorphism* – Polymorphism of $\beta$-casein is quite complex, due to its high genetic variability and to the presence of a large number of cases of not characterised or not well clarified variants. The alphabetical order itself is often not respected. The first evidence of polymorphism in $\beta$-casein dates to 1961, when Aschaffenburg discovered, by paper electrophoresis at alkaline pH, three different $\beta$-casein bands, named A, B and C, in order of their decreasing mobility [59].

Some years later, Peterson and Kopfler [60] and Kiddy *et al.* [61] demonstrate, by polyacrylamide gel electrophoresis at acid pH, that "A" band was not a unique casein but three different variants, named A1, A2 and A3.

The most important variants of $\beta$-casein are reviewed in Table 3; in the table also some not characterised variants are shown, like A', A3Mongolie, B2, and A4 of Bell, because they are now accepted from many Authors [21, 81, 82].
Table 3 – Discovery and diffusion of the genetic variants of β-casein.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Year</th>
<th>Species (country)</th>
<th>Breeds (frequency)</th>
<th>Methods of discovery</th>
<th>Diffusion</th>
<th>Variations with respect to A&lt;sup&gt;1&lt;/sup&gt;</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2) A&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1966</td>
<td><em>E. taurus</em> (UK)</td>
<td>Jersey, Guernsey, Ayshire, Holstein, Brown Swiss</td>
<td>PAGE pH 3</td>
<td>All species; all breeds</td>
<td>Peterson &amp; Kopfler, 1968 [60]; Peterson &amp; Kopfler, 1969 [61]; Kiddy et al., 1966 [62]</td>
<td></td>
</tr>
<tr>
<td>A&lt;sup&gt;3&lt;/sup&gt;</td>
<td>1961 1963</td>
<td><em>E. taurus</em> (UK)</td>
<td>Holstein</td>
<td>PAGE pH 3</td>
<td>Not common; other breeds</td>
<td>Pro (67) → Gln</td>
<td>Kiddy et al., 1966 [61]</td>
</tr>
<tr>
<td>D</td>
<td>1968</td>
<td><em>E. indicus</em> (India &amp; East Africa)</td>
<td>Indian Deshi, East African Boran</td>
<td>PAGE pH 3</td>
<td>Rare</td>
<td>Ser P (18) → Lys</td>
<td>Voglin, 1972 [64]</td>
</tr>
<tr>
<td>E</td>
<td>1972</td>
<td><em>E. taurus</em> (Italy)</td>
<td>Piemontese</td>
<td>PAGE pH 3</td>
<td>Rare</td>
<td>Glu (36) → Lys</td>
<td><em>&lt;sup&gt;6&lt;/sup&gt;</em></td>
</tr>
<tr>
<td>(6) A&lt;sup&gt;4&lt;/sup&gt;</td>
<td>1975</td>
<td><em>E. taurus</em> (Japan)</td>
<td>Japanese Brown (0.011)</td>
<td>PAGE pH 1.7</td>
<td>Rare</td>
<td>Pro (67) → His</td>
<td>Abo et al., 1975 [65]</td>
</tr>
<tr>
<td>(7) A&lt;sup&gt;5&lt;/sup&gt;</td>
<td>1975 1976</td>
<td><em>E. taurus</em> (Mongolia)</td>
<td>Mongolian cattle (0.04)</td>
<td>PAGE pH 3</td>
<td>Rare</td>
<td>Pro (67) → His</td>
<td>Groscleude, 1975 [66]</td>
</tr>
<tr>
<td>*B&lt;sup&gt;3&lt;/sup&gt;</td>
<td>1975</td>
<td><em>E. taurus</em> (N. Zealand)</td>
<td>PAGE pH 3.5</td>
<td>Rare</td>
<td>Pro (67) → His</td>
<td>Groscleude, 1975 [66]</td>
<td></td>
</tr>
<tr>
<td>*A&lt;sup&gt;6&lt;/sup&gt;</td>
<td>1981</td>
<td><em>E. javanicus</em> (Australia)</td>
<td>PAGE pH 3</td>
<td>Rare</td>
<td>Pro (67) → His</td>
<td>Bell et al., 1981 [21]</td>
<td></td>
</tr>
<tr>
<td>A&lt;sup&gt;5&lt;/sup&gt;</td>
<td>1983 1996</td>
<td><em>E. taurus</em> (Korea)</td>
<td>Korean cattle (0.008)</td>
<td>PAGE pH 3</td>
<td>Rare</td>
<td>Pro (67) → His</td>
<td>Han et al., 1983 [68]; Han et al., 1995 [69]; Han et al., 1996 [70]</td>
</tr>
<tr>
<td>F</td>
<td>1991</td>
<td><em>E. taurus</em> (Netherlands)</td>
<td>Meuse-Rhine-Yssel</td>
<td>PAGE-HPLC</td>
<td>Rare</td>
<td>Pro (67) → His</td>
<td>Visser et al., 1991 [71]; 1995 [72]</td>
</tr>
<tr>
<td>A&lt;sup&gt;3&lt;/sup&gt;</td>
<td>1993</td>
<td><em>E. taurus</em> (Norway)</td>
<td>Norwegian cattle</td>
<td>PAGE-HPLC</td>
<td>Rare</td>
<td>Pro (67) → His</td>
<td>Lien &amp; Rogne, 1993 [73]</td>
</tr>
<tr>
<td>C</td>
<td>1997</td>
<td><em>E. taurus</em> (Canada)</td>
<td>Holstein</td>
<td>PAGE-HPLC, MS</td>
<td>Rare</td>
<td>Pro (67) → His</td>
<td>Chin &amp; Ng-Kway-Han, 1997 [74]; Dong &amp; Ng-Kway-Han, 1998 [75]</td>
</tr>
</tbody>
</table>

<sup>1</sup> Still not characterized variant.

<sup>2</sup> At alkaline pH, A<sup>1</sup>, A<sup>2</sup>, and A<sup>3</sup> variants have the same mobility, so in a first time, appeared as a unique variant, named A.<n>3</n>

<sup>3</sup> Reference (most common) variants.

<sup>4</sup> Found with very low frequency in Italian Friesian [76] and German Friesian, Jersey and other German breeds [23], in Simmental [77] and Grey Alpine [78].

<sup>5</sup> In Sahiwal [79].

<sup>6</sup> In Hariana [63] and Choa [19]. It was named Bz, but it is identical to B variant.

<sup>7</sup> β-CN A<sup>1</sup> and β-CN H(A<sup>4</sup>) are probably the same variant.

<sup>8</sup> A<sup>5</sup> = A<sup>5</sup> (Algeria).

<sup>9</sup> Chromatography and trypsin hydrolysis: the presence of His-Lys peptide indicates that A<sup>3</sup>, nonet, is not equal to A<sup>3</sup>. (10) It is not specified if β-CN H variant arise from A<sup>1</sup> or A<sup>2</sup>.

<sup>11</sup> Italian Friesian [80].

FE: Paper electrophoresis; PCR: Polymerase chain reaction.
The variant named A' was found in 1975 by Abe et al. [65] in Japanese Brown cattle; in SGE at pH 1.7 it had a very low mobility. In 1983, Han et al. [68] found in Korean cattle (B. taurus) a variant, named A4, that seems to be the same as β-Cn A', according to comparisons of measured distances of the electrophoretic migration. On the base of this identity the Authors concluded a "phylogenesis" of Japanese Brown cattle from Korean cattle. They reproposed another time this discovery in 1995 [69], in which is precised that β-Cn A4 shows in SGE at acid pH a much slower electrophoretical mobility than β-Cn A3.

In 1996, Han and Shin [70] announced the discovery of β-Cn H variant and its aminoacid substitution. Most likely, also if it is not clearly declared, this is still the same variant that the Authors previously called β-Cn A4: in fact, in starch gel electrophoresis at acid pH, it migrates much slower than the other variants. It is not specified if H variant arises, phylogenetically, from A1 or A2.

The β-Cn A4 variant, that Bell et al. [21] found in 1981 in the Australian Banteng (B. javanicus), is probably not the same variant as Han's one. In fact, its electrophoretical mobility at acid pH is only slightly lower than that of the A3 variant. Also this variant, until now, has not been characterised.

In 1975, Grosclaude [66] found in Mongolian cattle (B. taurus) a variant whose electrophoretical properties were equal to A3 ones; the results of the triptic hydrolysis, however, evidenced a His-Lys dipeptide, that is not present in the β-Cn A3 sequence; it is, therefore, a new genetic variant, named A3Mongolie, whose characteristics are not yet known.

In the same year, Creamer and Richardson [67] discovered in New Zealand the variant B2, with an higher mobility than β-Cn B at pH 5.5, but a lower one at pH 3.5; the Authors suggested that the difference between B and B2 variants could involve a charged group, with a pKa between 5.5 and 8.5; it is more probable that this group is histidine and not a phosphate group [67]. Furthermore works are necessary to characterise it.

Is well established, instead, that Bz variant, that Aschaffenburg et al. [63] and Grosclaude et al. [19], independently from each other, found in zebu, is not different from the B. taurus B variant.

In 1986, Carles [83] evidenced by RP-HPLC a variant with the same electrophoretical mobility of A1, but a different chromatographic behaviour, due to a Pro©Leu substitution in region 114-169. Some years later Visser et al. [71, 72] found with the same analytical method an analogous variant in Meuse-Rhine-Yssel breed, that named β-Cn F. The
characterisation showed a Pro\(\rightarrow\)Leu substitution in position 152, therefore inside the region indicated by Carles. The identity of the experimental methods and substitution and the congruence of the position allowed the Authors to conclude that \(\beta\)-Cn F and \(\beta\)-Cn A1Carles were the same variant.

This fact seemed to be assumed until 1997, when Chin and Ng-Kway-Hang [74] discovered in Holstein another A1-similar variant, named G, by RP-HPLC and MS, still characterised by a Pro\(\rightarrow\)Leu substitution in position 137, therefore another time inside the Carles’ region. Now the question is: A1Carles variant corresponds to \(\beta\)-Cn F or to \(\beta\)-Cn G or, maybe, to neither of them? Since \(\beta\)-Cn region from 114 to 169 includes 9 proline residues, then the conclusion is not easy. The comparison between RP-HPLC spectra for the three cases shows rather similar behaviour, as is expected being the substitution in all the cases the same.

Another particular \(\beta\)-Cn “variant” is A5, found by Lien and Rogne in 1993 [73]; direct sequencing of PCR products showed a silent C\(\rightarrow\)T mutation in the third position of the codon coding for Pro(110). The mature protein is not different from \(\beta\)-Cn A2; nevertheless this fact it is possible to catalogue A5 in the list of \(\beta\)-Cn variants: it is a “genetic variant” in a literal sense, i.e. at level of the gene.

At least, Merlin and Di Stasio [84], found in Grey Alpine a \(\beta\)-Cn A3 linked to \(\alpha_s1\)-Cn B and supposed that it could be a new \(\beta\)-casein variant, on the base that in Western cattle the A3 allele is always linked to \(\alpha_s1\)-Cn C [85]. No further works were made to confirm this hypothesis.

**Diffusion** – A1 and A2 are the most diffused \(\beta\)-casein variants, with a slight prevalence of the latter in most breeds. B variant also is very diffused, but generally with a lower frequency with respect to A1 and A2. Normande and Jersey breeds have the highest \(\beta\)-casein B frequency values (30-45%), followed by Montbéliarde, Italian Brown, Reggiana, Modenese and Italian Red Pied (10-25%); in most breeds the frequency is near 10%. Also \(\beta\)-casein C is rather diffused, but less than B variant and with low frequency; in yak and zebu, C variant has never been detected.

\(\beta\)-casein A3 is a not common variant, but with very low frequency was found in several breeds: Italian Friesian [76] and German Friesian, Jersey and other German breeds [33], Simmental [77] and Grey Alpine [78]. The other \(\beta\)-Cn variants are rare and detected only in one breed, except \(\beta\)-Cn F that was found also in Italian Friesian [80].

**k-casein system**
The k-casein family consists of a major carbohydrate-free component and several "minor" glycosylated ones, with the same aminoacid sequence but different for the nature and the number of the carbohydrate groups [86]. The primary sequence of the B variant was first determined by Mercier et al. [87]; it has 169 aminoacids with a molecular weight of 19,007. There are two cystein residues that can form intra- or inter-molecular S-S bonds, giving rise to several polymeric forms. These cystein residues, by effect of heating, can form disulphide bonds with free SH of β-lactoglobulin. k-casein is completely soluble in presence of calcium ions. It's the only casein that can be associated with a carbohydrate co-factor; the most common glucides that take part to this complex are galactose, galactosamine and N-acetylneuraminic acid. All k-casein fractions are cleaved by chymosin: the specific attack site of this enzyme is the peptide bond between Phe (105) and Met (106); this cleavage is the first step of the rennet-coagulation process. In a phylogenetic point of view, k-casein represents an exception; in fact, while the other caseins seem to be derived from a common ancestor (they have an high sequence homology), k-casein is quite different from them and is similar to fibrinogen γ of blood serum [88], as can suggest also the affinity of their biological functions.

Polymorphism – k-casein was the last of the principal milk proteins for which polymorphism has been detected. In fact only the employment of appropriate substances, such as mercaptoethanol and cystein, able to break S-S bonds and reduce polymers to monomers, allowed to separate two genetic variants, named A and B [89-93] (Table 4).
<table>
<thead>
<tr>
<th>Variant</th>
<th>Year</th>
<th>Species (country)</th>
<th>Breeds (frequency)</th>
<th>Methods</th>
<th>Diffusion</th>
<th>Variations with respect to A or B</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) A, B</td>
<td>1964</td>
<td>B taurus (Canada)</td>
<td>Holstein</td>
<td>SGE alkal. pH</td>
<td>All species All breeds</td>
<td>A B</td>
<td>Neelin, 1964 [90]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B taurus (Holland)</td>
<td></td>
<td></td>
<td></td>
<td>Th(136) \rightarrow Ile</td>
<td>Schmidt, 1964 [91]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B taurus (Australia)</td>
<td></td>
<td></td>
<td></td>
<td>Asp(148) \rightarrow Ala</td>
<td>MacKinlay &amp; Wake, 1964 [92]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B taurus (USA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Woychik, 1964 [93]</td>
</tr>
<tr>
<td>C</td>
<td>1978</td>
<td>B taurus (Italy)</td>
<td>Grey Alpine (0.007%)</td>
<td>SGE pH 8.6</td>
<td>Less common Other breeds</td>
<td>B C</td>
<td>Di Stasio &amp; Merlino, 1978 [94], 1979 [95] Mariani, 1983 [96]</td>
</tr>
<tr>
<td></td>
<td>1979</td>
<td></td>
<td>Italian Brown</td>
<td></td>
<td></td>
<td>Arg(97) \rightarrow His</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1983</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1987</td>
<td>B taurus (Russia)</td>
<td></td>
<td>cDNA sequencing</td>
<td>Fare</td>
<td>B</td>
<td>Goroadezkij &amp; Kolesar, 1987 [97]</td>
</tr>
<tr>
<td>E</td>
<td>1989</td>
<td>B taurus (Germany)</td>
<td>Angler (0.026), German Red and White (0.022), German Friesian (0.06)</td>
<td>IEF</td>
<td>Not very common Other breeds</td>
<td>A E</td>
<td>Erhardt &amp; Seaf, 1989 [97] Erhardt, 1989</td>
</tr>
<tr>
<td>F</td>
<td>1992</td>
<td>B taurus (Russia)</td>
<td>Yakut</td>
<td>PCR</td>
<td>Fare</td>
<td>A F</td>
<td>Salimov et al., 1999 [98]</td>
</tr>
<tr>
<td>F</td>
<td>1996</td>
<td>B taurus (Finland)</td>
<td>Finnish Ayshire (&lt;0.001)</td>
<td>IEF</td>
<td>Fare</td>
<td>A F</td>
<td>Ikonez et al., 1996 [99]</td>
</tr>
<tr>
<td>G</td>
<td>1995</td>
<td>B grummiens (Russia)</td>
<td></td>
<td>PCR</td>
<td>Fare</td>
<td>A G</td>
<td>Salimov et al., 1999 [100]</td>
</tr>
<tr>
<td>G</td>
<td>1996</td>
<td>B taurus (Austria, Germany)</td>
<td>Pinzgauer (0.003)</td>
<td>IEF PAGE pH 8.9</td>
<td>Fare</td>
<td>A G</td>
<td>Erhardt, 1996 [101]</td>
</tr>
<tr>
<td>H</td>
<td>1998</td>
<td>B taurus (Austria, Germany), B taurus x B. indicus (Namibia)</td>
<td>Pinzgauer</td>
<td>SSCP</td>
<td>Fare</td>
<td>A I</td>
<td>Prinzenberg &amp; Erhard, 1998 [103]; Prinzenberg et al., 1999 [104]</td>
</tr>
<tr>
<td>I</td>
<td>1998</td>
<td>B taurus x B. indicus (Namibia)</td>
<td></td>
<td>SSCP</td>
<td>Fare</td>
<td>A I</td>
<td>Prinzenberg &amp; Erhard, 1998 [103]; Prinzenberg et al., 1999 [104]</td>
</tr>
<tr>
<td>A(1)</td>
<td>1999</td>
<td>B indicus, B taurus x B. indicus</td>
<td>Boran, Boran x N'Dama, Friesian x Sakival</td>
<td>SSCP</td>
<td>Fare</td>
<td>None</td>
<td>Prinzenberg &amp; Erhard, 1998 [102]; Prinzenberg et al., 1999 [103]</td>
</tr>
<tr>
<td>J</td>
<td>1999</td>
<td>B taurus (Ivory Coast, Burkina Faso)</td>
<td>Baoulé (0.02)</td>
<td>Triptic hydrosytes, IEF, RP-HPLC, MS</td>
<td>Fare</td>
<td>B J</td>
<td>Mahé et al., 1999 [105]</td>
</tr>
</tbody>
</table>

(1) Reference (most common) variants.
(2) German Simmental [77] and German Fleckvieh [93]; this variant was named k-Cn D, but it is identical to C. k-Cn C variant was found also in Muravey-Werdenfelsel [103] and Red Holstein [106].
(3) Finnish Ayshire (frequency 0.307) [54] and Italian Brown and Italian Friesian [107].
(4) The exact name of k-Cn A Az variant was “k-Cn A of zebu”.
SSCP: Single Strand Conformation Polymorphism.
For a long time these two variants seemed to be the only ones for k-casein, until 1978, when Di Stasio and Merlin [78, 94] found in Grey Alpine, by SGE at alkaline pH, another k-casein, named C. This variant was characterised by a faster migration with respect to A variant; at acid conditions A, B and C variants, instead, don’t show any difference in their electrophoretical mobility. No characterisation was made for Di Stasio and Merlin C variant. Some years later also Mariani [95] found a third k-casein in Italian Brown cattle. The Author, on the base of its electrophoretical mobility considered it as the same variant as Di Stasio and Merlin one. Only this latter C variant was characterised; the identity at level of aminoacid substitution between the two k-Cn C was never proved. k-Cn C, conversely with respect to A and B variants, presents an Arg→His substitution in position 97, i.e. in the para-k-casein portion [108]. This fact has negative repercussions on rennet-clotting time, that results longer, probably due to a change of conformation of the molecule that makes more difficult the interaction between substrate and chymosin [109, 110]. This substitution seems to have not a negative effect on curd firmness, that, conversely, appears to be good [111], probably because of a better interaction between the micelles of para-k-casein.

The k-Cn D variant found by Seibert et al. [77] in 1987 was proved [105] to be identical to k-casein C. The following k-casein allele discovered [97, 98] was called E.

Another allele, often forgotten, is k-Cn B2, found by Gorodetskij and Kaledin in 1987 [96]. This variant has been characterised in the same work, and shows an Ile→Thr substitution at position 153 with respect to k-Cn B. The Authors discovered this substitution sequencing for the new allele both mature protein both cDNA.

In 1989, Damiani et al. [112] found, by PCR, two k-casein A amplified fragments with a different behaviour towards the nuclease MboI and TaqI in a site located in a not-coding region of the gene. These eventual "variants" were named tentatively A1 and A2; no more works were made to furthermore characterise them.

One of the most complex knot in milk protein polymorphism is the case of k-casein F. In 1992, Sulimova et al. [99], found in Yakut (Russian B. taurus) by PCR a new k-casein allele, named F; the primary structure of the protein was reconstructed from the DNA sequence data, and revealed an Asp(148)→Val substitution with respect to A variant [113]. In 1995, Woollard and Dentine [114] announced the discovery, by PAGE and PCR, of a new k-casein allele, named F. Further investigations, however, have not confirmed this hypothesis: the new pattern observed was most likely a PCR artefact that was not evident at the time of the discovery [115]. In 1996, Ikonen et al. [54], by IEF, found in Finnish Ayrshire another k-casein allele, named F. This variant has been
characterised by Prinzenberg et al. [116] in the same year and showed a Arg(10)\(\Rightarrow\)His substitution.

k-Cn G represents another case of homonimy: in fact two different k-Cn variants, both named G, were discovered. One was found by Erhardt [101] in Pinzgauer and is characterised by a Arg(97)\(\Rightarrow\)Cys substitution. The other was been previously found by Sulimova et al. [100] in yak and is characterised by an Asp(148)\(\Rightarrow\)Ala substitution; the Authors cannot exclude the possibility (on the base of the IEF migration) that this variant could correspond to the k-casein X variant found by Kawamoto et al. [43] in Nepalese yak and never characterised. The k-Cn G of Sulimova et al. [100] is similar to the *Bison bonasus* k-Cn G allele previously found by Udina et al. [117] except for at least a single nucleotide substitution in the stop codon (TGA\(\Rightarrow\)TAA).

Another point is that concerning k-Cn "A of zebu" and k-Cn H variants. In 1974, Grosclaude et al. [19] announced the discovery in Madagascan zebu of a Ile(135)\(\Rightarrow\)Thr substitution that differentiates zebu A variant from the common *B. taurus* k-Cn A; this variant was indicated as "A of zebu" or simply k-Cn Az. The same substitution, Ile(135)\(\Rightarrow\)Thr, has been detected also in *B. taurus* (Pinzgauer) by Prinzenberg and Erhardt [102]; this latter was considered as a new variant and named k-Cn H.

Recently, Prinzenberg and Erhardt [104] evidenced by SSCP in *B. taurus* x *B. indicus* crosses the presence of a silent A\(\Rightarrow\)G transition in the third codon position for aminoacid Pro(150) (CCA\(\Rightarrow\)CCG), that creates an *MspI* restriction site. This mutation has not effect on the aminoacid sequence of the mature protein. This variant has been denoted A(1) in order to distinguish it from the A allele.

The last detected k-casein variant is k-Cn J that Mahé et al. [16] found in African Baoulé cattle (*B. taurus*), characterised by a Ser(155)\(\Rightarrow\)Arg substitution with respect to the B variant.

**Diffusion** – The most diffused k-casein alleles are A and B, present in all breeds with variable frequency. A variant prevails in Friesian, Ayrshire, Red Danish and Indian zebu; in Irish Kerry its frequency is near 93% [118]. B variant, instead, is prevalent in Jersey, Normande and African zebu. Beef cattle breeds have a marked prevalence of B variant [1].
k-casein C is less common, but was found in many breeds. Besides Grey Alpine and Italian Brown, it was detected in German Simmental [77], German Fleckvieh [98], Murnau-Werdenfelser [105] and Red Holstein [106].

k-casein E, nevertheless is considered a not very common variant, in Finnish Ayrshire was found also with high frequency (~30%) [54]; recently this variant has been detected also in Italian Brown and Italian Friesian breeds [107, 119].

\[\alpha\]-lactalbumin

\[\alpha\]-lactalbumin is a constituent of lactose-synthetase, the enzyme responsible of the synthesis of lactose, in the final step where glucose is linked to galactose. The primary sequence of \[\alpha\]-La was first determined by Brew et al. [120]. The most common B variant consists of 123 aminoacids with a molecular weight of 14,175. There are eight cysteins, variously connected with both inter- both intra-molecular bonds. There is a great percentage of homology with lysozime; furthermore, they have a similar molecular weight, the same number of S-S links, identical N and C terminal residues; all this similarity suggests that both proteins arise from a common ancestor. Also some glycosylated forms were found [121-124]. They can not be considered as genetic variants in true sense, since they have not any aminoacid substitution and any mutation was evidenced at level of the gene.

Polymorphism and diffusion – Nevertheless its less accentuated genetic variability (until now only a few variants are known, see Table 5), \[\alpha\]-lactalbumin was the second milk protein in which polymorfism was noticed: in 1958 Blumberg and Tombs [125] found in South African White Fulani (\(B. \text{indicus}\)), by paper electrophoresis, two genetic variants, named A and B, in order of their decreasing electrophoretical mobility. B variant was the common \[\alpha\]-La present in all Western breeds; \[\alpha\]-La A, instead, was found only in \(B. \text{indicus}\) populations and seemed to be peculiar of this species. After the work of Blumberg and Tombs, other detections were made of this allele by Aschaffenburg [127] and Bhattacharya [126], but always in \(B. \text{indicus}\) species.
The first evidence of the presence of $\alpha$-La A in *B. taurus* was proved by Osterhoff and Pretorius [128], that found this allele in European breeds imported in South Africa. Some Authors, however, objected that maybe A allele present in such breeds could arise from crosses with South African *B. indicus*. This controversy was resolved when $\alpha$-La A was found by several Authors also in bovine breeds reared in Europe [129-132]. However, a study to verify and confirm the identity between *B. taurus* A variant and *B. indicus* one has never been made.

Bettini and Masina [131], on the base of the occurrence of the $\alpha$-La A allele in breeds reared in the South of Italy (like Podolian cattle), and on the base also of some anatomical resemblances between these breeds and zebu, suggested a phylogenetic origin of such breeds from *B. indicus*.

Bell *et al.* [21] in Bali (banteng) cattle (*B. javanicus*) found a third $\alpha$-La variant, named C. Until now, there is no evidence of this third allele neither in *B. taurus* nor in *B. indicus*. For this latter variant the aminoacid substitution is known [21], but not its position (Table 5).

Another $\alpha$-lactalbumin polymorphism, named $\alpha$-La(+15) [133], was found in a not-coding sequence of the gene; this case will be examined later.

### $\beta$-lactoglobulin
ß-lactoglobulin is the major whey protein in cow milk, and after crystallisation by Palmer in 1934 [134] it was used for many years as a protein model for structural and enzymatic studies concerning denaturation and linkage between ions and proteins. ß-lactoglobulin is not present in all mammals: in human milk, for example, is not found. The biological functions of this protein are not still well-known; it could have a role on the metabolism of phosphate in the mammary gland and on the transport of retinol and fatty acids within the gut [135, 136]. The most common B variant consists of 162 aminoacids with a molecular weight of 18,277. The primary sequence was first established by Braunitzer et al. [137]. Some corrections are made to the original sequence [20, 138]: residues 155 and 156 have been changed from Leu-Gln to Gln-Leu, and residues 84 and 87 from Leu-Ile to Ile-Leu. The proposal [20] to change Asp 11 in Asn has not been substantiated [139, 140].

Polymorphism — ß-lactoglobulin was the first protein in which polymorphism was evidenced. In 1955, Aschaffenburg and Drewry [141] observed, by paper electrophoresis, two distinct bands of ß-Lg, that were named ß1 and ß2. On the base of the fact that milk arising from one-egg twin heifers always presented the same type of ß-Lg, the Authors suggested the genetic nature of the variation. In 1957 [142], when Aschaffenburg and Drewry confirmed the discovery, the name of the bands was changed in A and B, in order of their decreasing electrophoretical mobility. Until now, at least 12 variants are known for ß-lactoglobulin, from A to J plus ß-Lg W and ß-Lg Dr (Table 6). The variant named ß-Lg D was found by Grosclaude et al. [18] in Monbéliarde and by Meyer [145] in German breeds, independently from each other.
Table 6 – Discovery and diffusion of the genetic variants of β-lactoglobulin.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Year</th>
<th>Species (country)</th>
<th>Breeds (frequency)</th>
<th>Methods</th>
<th>Diffusion with respect to B</th>
<th>Variations</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1955 1957</td>
<td><em>B. taurus</em> (UK)</td>
<td>Shorthorn (0.61), Guernsey (0.22), Friesian (0.40), Ayrshire (0.31)</td>
<td>PE pH 8.6</td>
<td>All breeds B. indicus B. grumnens</td>
<td>Gly(64) → Asp(113) → Val</td>
<td>Aschaffenburg &amp; Drewry, 1955 [141], 1957 [142]</td>
</tr>
<tr>
<td>B</td>
<td>1955 1957</td>
<td><em>B. taurus</em> (UK)</td>
<td>Shorthorn (0.89), Guernsey (0.78), Friesian (0.60), Ayrshire (0.69)</td>
<td>PE pH 8.6</td>
<td>All breeds B. indicus B. grumnens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1962</td>
<td><em>B. taurus</em> (Australia)</td>
<td>Jersey</td>
<td>SOE pH 8.6 (1)</td>
<td>Not common Other breeds (2) B. indicus (3) B. grumnens (4)</td>
<td>Gln(59) → His</td>
<td>Bell, 1962 [143], B. McKenzie, 1964 [144]</td>
</tr>
<tr>
<td>D</td>
<td>1966</td>
<td><em>B. taurus</em> (France)</td>
<td>Montbéliarde (0.02)</td>
<td>SOE pH 8.6</td>
<td>Less common Other breeds (2)</td>
<td>Glu(45) → Gln</td>
<td>Croesclaude et al., 1963 [145]</td>
</tr>
<tr>
<td>E</td>
<td>1976</td>
<td><em>B. taurus</em> (Germany)</td>
<td>German Red &amp; White, German Brown, German Highland</td>
<td>SOE pH 8.6 (1)</td>
<td></td>
<td></td>
<td>Meyer, 1966 [146]</td>
</tr>
<tr>
<td>F</td>
<td>1976</td>
<td><em>B. taurus</em> (Australia)</td>
<td>Droughtmaster</td>
<td>SOE pH 8.5</td>
<td>Rare</td>
<td>Gly(64) → Asp(113) → Val Asp(28) → Asn-CHO</td>
<td>Bell et al., 1970 [147], 1981 [148]</td>
</tr>
<tr>
<td>G</td>
<td>1976</td>
<td><em>B. taurus</em> (Australia)</td>
<td>Trptic hydrolysis</td>
<td>Only <em>B. grumnens</em> and <em>B. javanicus</em></td>
<td>Gln(158) → Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1976</td>
<td><em>B. grumnens</em> (Nepal)</td>
<td>E. grumnens and E. javanicus</td>
<td>Glu(158) → Gly</td>
<td></td>
<td>Croesclaude et al., 1976 [149], Bell et al., 1981 [150]</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1976</td>
<td><em>B. javanicus</em> (Australia)</td>
<td>SOE pH 8.5 (1)</td>
<td>Only <em>E. javanicus</em></td>
<td>Glu(158) → Gln(158) → Gly Ile(78) → Met</td>
<td>Bell et al., 1981 [151]</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>1976</td>
<td><em>B. grumnens</em> (Germany)</td>
<td>Mumau-Werdenfleter (0.006), Jersey, Red Holstein x Simmental</td>
<td>SOE pH 8.5 (1)</td>
<td>Only <em>E. javanicus</em></td>
<td>Glu(158) → Gln(Ile(78) → Met)</td>
<td>Bell et al., 1981 [151]</td>
</tr>
<tr>
<td>W</td>
<td>1976</td>
<td><em>B. grumnens</em> (Germany)</td>
<td>ELAC.CE pH 8.9</td>
<td>Rare</td>
<td>Ile(56) → Leu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*H</td>
<td>1983</td>
<td><em>B. taurus</em> (Italy)</td>
<td>Italian Friesian</td>
<td>SOE pH 8.6</td>
<td>Rare</td>
<td></td>
<td>Davoli et al., 1987 [152], Krause et al., 1988 [153]</td>
</tr>
<tr>
<td>I</td>
<td>1998</td>
<td><em>B. taurus</em> (Poland)</td>
<td>Polish Red (0.014)</td>
<td>IEF</td>
<td>Rare</td>
<td>Glu(103) → Gln(103) → Gly</td>
<td>Erhardt et al., 1998 [154]</td>
</tr>
<tr>
<td>J(X)</td>
<td>1993 1996</td>
<td><em>B. taurus</em> (Hungary)</td>
<td>Hungarian Grey (0.03)</td>
<td>IEF</td>
<td>Rare</td>
<td>Pro(126) → Leu</td>
<td>Baranyi et al., 1993 [155], Godovac-Zimmermann et al., 1996 [156]</td>
</tr>
</tbody>
</table>

* Still not characterised variant.
(1) Evidence for β-Lg A and B alleles in *B. grumnens* are noticed by Croesclaude et al. [156] with a very low frequency. The authors state that this presence can be interpreted as a trace of crossbreeding with *B. taurus*. These two alleles, however, have been detected also in yak by Lozoya [157].
(2) Reference (most common) variant.
(3) Without presence of ura.
(4) German Jersey [33].
(5) Cubian zebu [153].
(6) Pamir yak [157].
(7) Danish Jersey [27], Polish Simmental [56], Italian Brown [36,37], Reggiana [38], Mozènes [159], Modicana [131], Rendena [160], Simmental [77].
(8) In 1966 was discovered the variant and in 1970 was evidenced the presence of a carbohydrate moiety with respect to A variant; in 1973 identified also an aminoacid substitution [161, 162].
(9) In 1995 was called β-Lg X. Only in 1996 it was named β-Lg J.
(10) ELAC.CE. Cellulose acetate electrophoresis: IEF-IEF: isoelectric focusing in immobilized pH gradients.
β-Lg Dr, discovered by Bell et al. [146] in 1966 is very singular. Analysing by starch gel electrophoresis some milks from Droughtmaster cows (*B. taurus* x *B. indicus*), the Authors found a new β-Lg that migrates more slowly than A, B and C variants. In 1970, it was characterised and resulted identical to the A variant, except for the presence of a covalently-attached carbohydrate, constituted prevalently by N-acetilneuraminic acid, hexosamine, mannose and galactose [147]. As noticed by Eigel et al. [140], this glycosylated form could not be considered a genetic variant, because the variation was not at level of the gene or the mRNA, but rather by post-translational modifications. Some years later, however, was announced [139, 161, 162] for the same β-Lg Dr also the presence of an aminoacid substitution in position 28, in which is involved just the residue linked to the carbohydrate. This variation makes the protein a genetic variant in true sense.

The variant named DYak was found by Grosclaude et al. [20] in this species in 1976. The name D is due to the fact that the new variant migrated, by PAGE and SGE, both at alkaline both at acid pH, exactly like β-Lg D found in *B. taurus*. By hydrolysis with trptic enzymes, instead, it presented some fragments different from the ones of β-Lg D. The coincidence of the electrophoretical mobilities can be explained on the base of the aminoacid substitution Glu(158)\(\rightarrow\) Gly, that leads to a charge variation analogous to that of β-Lg D.

Some years later, Bell et al. [139], found in Bali (banteng) cattle (*B. javanicus*) three new β-Lg variants. One of these, named E, presented the same aminoacid substitution as DYak variant. Eigel et al. [140] suggested to identify these two variants with a unique name (β-Lg E) because to avoid subscripts and superscripts.

The other two variants found by Bell et al. [139] in banteng cattle arise from E variant, and were named F and G; they are characterised by an electrophoretical mobility respectively slower and faster with respect to β-Lg Dr; the mobility of G is equal to that of E variant, because the substitution does not lead to a change in the charge.

In 1980, Weiß [148] and Buchberger et al. [149] announced the discovery of another β-lactoglobulin variant in the Murnau-Werdenfelser breed, called β-Lg W; the discovery was confirmed in 1982 [150]. This variant was characterised and showed an aminoacid substitution Ile(56)\(\rightarrow\)Leu with respect to the B variant and an electrophoretical mobility at pH 8.9 between B and D variants.

The discovery of β-Lg H was announced in 1987 by Davoli et al. [151, 152]. This variant at pH 8.6 results faster than β-Lg A; unfortunately, until now, only some preliminary
characterisations were made, like determination of molecular weight, pI and aminoacid composition [163].

The last discovered variant is $\beta$-Lg J, found in 1993 in Hungarian Grey cattle by Baranyi et al. [154] and indicated provisionally as $\beta$-Lg X; by isoelectric focusing it migrates between A and B variants. In 1996, when it was characterised [155], its name was definitively changed in $\beta$-Lg J.

In 1998, Zappacosta et al. [164], by isoelectric focusing, identified C-terminally truncated A and B $\beta$-lactoglobulin variants with missing N-terminal peptides, beyond residues in the range 100-103 and 136-147 respectively. Two of the minor components were related to $\beta$-Lg A and two to $\beta$-Lg B. The Authors suggested two hypotheses: they may be non-allelic forms or enzyme-mediated products of the mature protein (in this case they have not a genetic origin) or, maybe, they can arise from the occurrence of a stop-codon, which would result in the synthesis of a protein 33 aa. shorter. Since this phenomenon involves both A and B alleles, the first hypothesis seems to be most probable.

Diffusion – $\beta$-lactoglobulin A and B variants are diffused in all breeds; B prevails in some European breeds, like Ayrshire, Shortorn, Red Danish, and in Asian and African zebu (85-95%), as well as in Italian beef cattle breeds (70-80%). In Friesian and in several other breeds the two alleles have the same frequency. Evidence for $\beta$-Lg A and B alleles in B. grunniens, with a very low frequency, are noticed by Grosclaude et al. [156]. The Authors suggested that this presence can be interpreted as a trace of crossbreedings with B. taurus. These two alleles have been detected by Lozovaya also in Pamir yak [157].

$\beta$-Lg C is a not common allele, found only in Australian Jersey [143] and in German Jersey [33]. Occurrences of this allele in Cuban zebu [158] and Pamir yak [157] have been evidenced.

$\beta$-Lg D, instead, was observed in several other breeds: Danish Jersey [27], Polish Simmental [56], Italian Brown [36, 37], Reggiana [38], Modenese [159], Modicana [131], Rendena [160], German Simmental [77].

$\beta$-Lg E, F and G variants, until now, were observed only in Bali (banteng) cattle (B. javanicus). The other $\beta$-lactoglobulin alleles are rare and were found only in one breed.

Variant phylogenesis
For a further explanation of the exposed concepts, in Figure 1 is illustrated the phylogenesis of the genetic variants for the six milk proteins here considered.

In addition, in Table 7 are reported the references of the first complete sequencing for the mature protein and the gene.
Polymorphisms at the not-coding regions

The introduction of modern biomolecular methods, like RFLP, PCR and SSCP, opened up new opportunities to the study of milk protein polymorphism. They allowed the knowledge of the variations at the DNA level for the known protein variants, and, at the same time, the detection of new alleles. Several studies have been applied to the not-coding zones, in particular 5' and 3' flanking regions. 5' flanking region, compared to coding sequences, is a zone with an high genetic variability, and mutations in this part often have important repercussions on the expression of the gene and, in general, on milk production traits and composition. The case concerning the origin of $\alpha_{s1}$-Cn G has been already treated.

For $\alpha$-lactalbumin, Bleck and Bremel [133] identified in Holstein-Friesian three single bp polymorphisms within the 5' flanking region, at position +15, +21 and +54 relatively to the mRNA transcription start point. The +15 and +21 variations were in the zone encoding the 5'-untranslated region (5' UTR) of the mRNA sequence, while the +54 polymorphism was a silent mutation in the coding region of the gene. The $\alpha$-La(+15)A and $\alpha$-La(+15)B alleles are characterised respectively by an adenine and a guanine in position +15. These two forms were examined to investigate the effects of such a mutation on milk production and milk composition: $\alpha$-La(+15)A variant is associated with greater milk, protein and fat yields, while $\alpha$-La(+15)B allele is related to higher protein and fat percentages [171]. The (+15) polymorphism was detected also in Taiwan Holstein [172], Italian Fresian and Italian Red Pied [173], Swedish Red and White [174] and Italian Brown [175].
In 1997, Voelker et al. [176] found in Holstein, Simmental and Brown Swiss another single bp difference located in position -1689 from the transcription start point, also a variation adenine (form A) / guanine (form B). The Authors noticed a relationship between this polymorphism and the (+15) one: α-La(A+15) was always linked to α-La(-1689)A variant. The (-1689) polymorphism was found also in Korean Holstein population [177].

In 1994, Schild et al. [178], analysing milk from several bovine breeds (Holstein Friesian, Brown Swiss, German Simmental, Jersey, Galloway, Scottish Highland) and Ceylon Dwarf zebu, found 15 DNA "variants" in the 5' flanking region of the k-casein gene, some of which located within potential regulatory sites and possibly involved in the expression of the gene. It is possible that the different expression of A and B k-casein alleles is related to some of these mutations in the 5' flanking region of the gene [178, 179].

Recently at level of the 2nd intron of the k-casein gene, Damiani et al. [180] evidenced a furthermore polymorphism, regarding the short interspersed elements (SINEs) [181], precisely in the Bov-A2 sequence, an homodimer of Bov-A.

Also for β-casein was evidenced a 5' flanking region polymorphism by Bleck et al. [182]. The Authors detected in Brown Swiss and Jersey a deletion of a thymine at position -516 from the transcription start point.

β-lactoglobulin is the most studied milk protein in relation to the 5' flanking polymorphism. Several studies are focused to understand the causes of the different expression of β-Lg A and B variants: in fact, on average, 1.2 times more A than B protein was found in the milk [183]. Measurements of mRNA levels indicated that approximately 60% more A than B mRNA was present. Differences either at the level of transcription or in mRNA stability are probably responsible for the different rates of synthesis of the corresponding proteins [183]. To investigate possible differences in transcription, the 5' flanking region of the two variants was sequenced (from position -795 to +59 with respect to the transcription start point). Wagner et al. [184] identify 14 single bp allele substitutions within the 5' flanking region, and 2 in the untranslated region (5' UTR) of exon 1, that could influence the transcription level of the two alleles. Further research (sequencing from position -733 to +95 with respect to the transcription start point of the β-Lg gene) [185] confirmed the results obtained from Wagner et al. [184]. In particular the studies of Lum et al. [185] were focused to the binding site (from -436 to -429) for the activator protein 2 (AP-2), a transcription factor present in the mammary gland during lactation, where an allele specific single base substitution (G in the A allele, and C in the B allele) at position -430 is present. Lum et al. [185] demonstrated that AP-2 has a
different binding affinity for the two variants, which could be affected by the -430 mutation, and proposed a modulating role of AP-2 in the different allele expression.

Kaminski and Zabolewicz [186], by means of SSCP, found six variants in the β-Lg 5' flanking region, between position -501 and -293, that were no further characterised.

PCR was also used to amplify and clone a region from β-Lg locus, that spanned exon 4 and 5 (849 bp). Sequence analysis of the cloned region revealed two new single base substitutions that further differentiated A and B forms. The mutation were localised in the intron sequence between the two exons, at position 276 (T in the A allele, and C in the B allele) and 562 (T in the A allele, and G in the B allele) of the cloned fragment. These nucleotide substitutions resulted in an allele specific restriction profile, that could be used as a genetic marker [187].

**Haplotypes**

The analysis of Mendelian segregation of three casein locus (αs1-Cn, β-Cn, and k-Cn) showed that they were tightly linked together [27, 188, 189]. Further studies demonstrated that also the αs2-Cn locus was genetically linked to the other three casein loci [52].

These observations led to the conclusion that the four casein loci behave themselves as a one Genetic Unit, in which allele combinations at the casein locus (haplotypes) are tightly linked together. These genetic combinations, therefore, could be peculiar of a particular breed, and could be used as a tool for genetic marker [73].

Further experiments enabled to establish the physical association (and therefore genetic association) of casein loci, which are situated on the same chromosome (6th) [190], on a range of 185-250 kb [191, 192], in this order: αs1-Cn, β-Cn, αs2-Cn, and k-Cn. The genes coding for β-Lg and α-La, are located, in Bos, on 11th [193] and 5th [194, 195] chromosome respectively, justifying their independent segregation respect to the caseins loci.

**Conclusions**

The research on milk protein polymorphism is in full growth, with several aims: to discover further new variants, characterise them and, particularly, to understand the role that each variant can have on milk nutritional and technological properties. With the recent progress in biochemistry and molecular biology it is possible to inquire into the causes that determine these effects on the composition and on the coagulation parameters, to understand why a simple mutation can have such a relevance for the
properties of milk and, after all, if these effects arise only from the variation in the protein or rather are due to the action of promoters present in the regulatory not coding sequences of DNA. Often, an effect can arise from a complex connection and succession of several causes, all involved in the same singular phenomenon; actually this is the most accredited interpretation.

The knowledge of the biochemical and biomolecular processes can lead new contributions to important biotechnological applications, like genetic improvement and molecular engineering, with the aim to obtain breeds more and more suitable for the modern requirements.

**Key words:** Milk proteins, genetic polymorphism, *Bos* genus, variant discovery, variant diffusion.

**Parole chiave:** Proteine latte, polimorfismo genetico, genere *Bos*, scoperta varianti, diffusione varianti.

**Summary** – The Authors review the discovery and the diffusion of genetic variants of the six principal milk proteins (9 for αs1-Cn, 4 for αs2-Cn, 15 for β-CN, 13 for κ-Cn, 3 for α-La, 12 for β-Lg) in the species of *Bos* genus (*B. taurus, B. indicus, B. grunniens, B. javanicus*): the year and the discoverers, the species and the breeds, the analytical method and the aminoacid substitution. A particular attention is devoted to the "problematic" situations: the changes of name, the omitted letters, the cases of homonymy, the variants not well or not still characterised and the "possible" genetic variants never confirmed. In particular two κ-casein F (one by Sulimova *et al.*, 1992 [99] and the other by Ikonen *et al.*, 1996 [54]) and two κ-casein G (one by Sulimova *et al.*, 1996 [100] and the other by Erhardt, 1996 [101]) are noticed. In the paragraphs relative to the diffusion, the distribution of the genetic variants for each protein and the following discoveries in other *Bos* species or breeds are described.

**Résumé** – Le polymorphisme des protéines du lait: découverte et diffusion des variants génétiques dans le genre *Bos*. Les Auteurs ont décrit la découverte et la diffusion des variants génétiques des six principales protéines du lait (9 pour αs1-Cn, 4 pour αs2-Cn, 15 pour β-CN, 13 pour κ-Cn, 3 pour α-La, 12 pour β-Lg) dans les espèces du genre *Bos* (*B. taurus, B. indicus, B. grunniens, B. javanicus*): l’an et les découvreurs, l’espèce et la race, la méthode et la substitution d’acide aminé. Une consideration particulière est donnée à les cases problématiques: les changements de dénomination, les lettres sautées, les questions d’homonymie, les variants pas bien ou pas encore caractérisés et les “possibles” variants génétiques jamais confirmés. Dans les paragraphes relatives à la diffusion est décrite la distribution des variants génétiques pour chaque protéine et les suivantes découvertes dans autres espèces ou races du genre *Bos*.

**Riassunto** – Il polimorfismo delle proteine del latte: scoperta e diffusione delle varianti genetiche nel genere *Bos*. Gli Autori hanno descritto la scoperta e la
diffusione delle varianti genetiche delle sei principali proteine del latte (9 per αs1-Cn, 4 per αs2-Cn, 15 per β-Cn, 13 per κ-Cn, 3 per α-La, 12 per β-Lg) nelle specie del genere Bos (B. taurus, B. indicus, B. grunniens, B. javanicus): l’anno e gli scopritori, la specie e la razza, la metodica e la sostituzione aminoacidica. Una particolare attenzione è stata riservata alle questioni problematiche: i cambiamenti di denominazione, le lettere omesse, i casi di omonimia, le varianti non ben caratterizzate o non ancora caratterizzate e le “possibili” varianti genetiche mai confermate. In particolare sono state messe in luce due κ-Cn F (una scoperta da Sulimova et al., 1992 [99]; l’altra da Ikonen et al., 1996 [54]) e due κ-Cn G (una scoperta da Sulimova et al., 1996 [100]; l’altra da Erhardt, 1996 [101]). Nei paragrafi relativi alla diffusione viene descritta la distribuzione delle varianti genetiche di ogni proteina e le successive scoperte in altre specie o razze del genere Bos.

References


• 51) Grosclaude F., Mahé M.F., Mercier J.C., Bonnemaire J., Teissier J.H. (1976) – Polymorphisme des lactoprotéines de Bovinés Népalais. II. Polymorphisme des caséines "minor"; le locus $\alpha S_2$-Cn est-il lié aux loci $\alpha S_1$-Cn, $\beta$-Cn et k-Cn? "Annales de Génétique et de Sélection Animale", 8, 481-491.

• 52) Grosclaude F., Joudrier P., Mahé M.F. (1978) – Polymorphisme de la caséine $\alpha S_2$ bovine: étroite liaison du locus $\alpha S_2$-Cn avec les loci $\alpha S_1$-Cn, $\beta$-Cn et k-Cn; mise en évidence d'une délétion dans le variant $\alpha S_2$-CnD. "Annales de Génétique et de Sélection Animale", 10, 313-327.

• 53) Guérin G. (1976) – Personal communication, quoted by Grosclaude et al., 1978 [52].


