Variability in major milk protein genes in two autochthonous cattle breeds mainly reared in the Parmigiano-Reggiano cheese production area: Reggiana and Modenese

Variability in major milk protein genes in two autochthonous cattle breeds mainly reared in the Parmigiano-Reggiano cheese production area: Reggiana and Modenese

Stefania Dall'Olio[1], Giuseppina Schiavo[1], Samuele Bovo[1], Anisa Ribani[1], Francesca Bertolini[1], Luca Fontanesi[1]

[1]Animal and Food Genomics Group, Division of Animal Sciences, Department of Agricultural and Food Sciences, University of Bologna, Viale Fanin 46, 40127 Bologna, Italy

Corresponding author: Stefania Dall'Olio, E-mail: stefania.dallolio@unibo.it

Running head: Milk protein gene variants in local cattle breeds.

Highlights

- Reggiana and Modenese are autochthonous cattle breeds reared in the North of Italy.
- Their milk is mainly used to produce branded mono-breed Parmigiano-Reggiano cheeses.
- Variants in milk protein genes can have important effects on cheese making properties.
- A total of 3,298 cattle of the two breeds were genotyped at the *CSN2*, *CSN3* and *PAEP* genes.
- Breeding strategies can be oriented to increase the frequency of favourable alleles.

Abstract

Reggiana and Modenese (also known as Bianca Valpadana) are two iconic autochthonous cattle breeds mainly raised in the North of Italy, in the area where the Protected Designation of Origin Parmigiano-Reggiano cheese is produced. The monitoring of diffusion of milk protein gene variants in these cattle breeds is important to evaluate the impact of selection and conservation programs on the frequency of favourable alleles on milk quality and cheese-making properties and to avoid the increase of the frequency of alleles with unfavourable effects. In this study, we genotyped a total of ~3,300 cattle of the Reggiana and Modenese breeds with the GeneSeek® Genomic Profiler™ Bovine 150K Array and obtained an updated picture of the variability present in three milk protein genes (*CSN2*, *CSN3* and *PAEP*, also known as β-lactoglobulin). We then retrospectively evaluated the trend of allele frequency changes over time, going back ~50 years, comparing the information reported by previous studies in the same two breeds. The allele frequency trends could be considered either positive, according to the increase in frequency of alleles with favorable effects on cheese making properties and cheese yield (i.e., *CSN3**B in both breeds) or negative, based on the decrease in frequency of favorable alleles on the same traits (i.e., *PAEP**B in Reggiana). Therefore, some adjustments in the selection and conservation programs might be needed to maintain genetic properties that can positively contribute to addressing the production of high-quality Parmigiano-Reggiano cheese and support a sustainable conservation of these cattle genetic resources.

Keywords: *Bos taurus*, Cheese-making property, *CSN2*, *CSN3*, Genetic resource, *PAEP*, Polymorphism.

1. Introduction

Reggiana and Modenese (also known as Bianca Valpadana) are dual purpose autochthonous cattle breeds mainly raised in the Emilia-Romagna region, in the North of Italy, in the area where the Protected Designation of Origin (PDO) Parmigiano-Reggiano cheese is produced. These two breeds have small populations as they account for a total of 4,752 (2,705 cows for Reggiana) and 1335 (576 cows for Modenese) registered animals in their Herd Books, raised in a total of 104 and 54 herds, respectively (Associazione Nazionale Allevatori Bovini di Razza Reggiana, ANABoRaRe, 2023). Both the Reggiana and Modenese breeds are historically considered to be derived from the ancestral cattle populations that were raised in the Po Valley and from which Parmigiano-Reggiano cheese first originated. For these reasons, they are regarded as iconic cattle genetic resources, linked to the traditional agriculture of the region. Nowadays, almost all milk produced by Reggiana cattle and a large amount of the milk produced by Modenese cattle is processed into two mono-breed PDO Parmigiano-Reggiano cheeses. These cheeses are labelled with brand names that are linked to the coat colour of the two breeds: "*Vacche Rosse*" (Red Cattle), for the solid red coat colour (known as *fromentino*) of the Reggiana cattle; "*Vacche Bianche*" (White Cattle), for the solid white coat colour of the Modenese cattle (Bovo et al., 2021; Bertolini et al., 2022). These branded Parmigiano-Reggiano cheeses are sold at a higher price than those made from undifferentiated milk (Gandini et al., 2007). The resulting economic gain can compensate the production gap that Reggiana and Modenese cattle have in terms of average 305-day milk yield (year 2023: 6,020kg and 4,469 kg, respectively; Associazione Italiana Allevatori, 2023) compared to more productive dairy breeds such as Italian Holstein and Italian Brown cows (10,428 kg and 7,718 kg, respectively; Associazione Italiana Allevatori, 2023). A selection program is currently underway in the Reggiana breed for which a composite genetic index (known as the Parmigiano Reggiano Cheese Yield Genetic Index) focused on expected cheese yield, including both milk yield and milk protein percentage, is estimated (Bertolini et al., 2020). For Modenese breed, only a

conservation program managed by ANABoRaRe is currently in place. For Reggiana and Modenese breeds, which are considered small populations, monitoring and controlling inbreeding are integral parts of the selection and conservation programs, respectively (Schiavo et al., 2021).

Many studies have investigated genetic variability in milk proteins and their effects on quantitative and qualitative milk and cheese-making properties in cattle. Some variants have consistently been reported to impact a variety of traits, leading to changes in the economic and health-related properties of the resulting dairy products (Comin et al., 2008; Caroli et al., 2009; Gai et al., 2021; Bisutti et al., 2022). Among the milk protein genes, casein beta (symbol *CSN2*, encoding the β-casein protein with symbol β-CN), casein kappa (*CSN3* encoding the k-CN protein) and progestagen-associated endometrial protein (*PAEP*, also known as β-lactoglobulin gene with symbols *BLG* or *LGB*, encoding the β-lactoglobulin protein) have been reported to have the highest genetic variability in *Bos taurus* (Caroli et al., 2009). The *CNS2* and *CNS3* genes are closely linked on *Bos taurus* chromosome 6 (BTA6) and are members of the casein gene cluster; *PAEP* is localised on BTA11.

The monitoring of the diffusion of milk protein variants in these genes at the cattle population level is important to evaluate the impact of breeding strategies on the frequency of favorable alleles and to avoid the increase of the frequency of alleles with unfavorable effects, which could have negative economic impacts (Caroli et al., 2009). In Reggiana and Modenese breeds, milk protein polymorphisms were previously investigated by a few pioneering studies that analysed milk protein variants directly from the milk using electrophoretic techniques (Mariani and Russo, 1971; Russo and Mariani, 1972; Russo et al., 1984; Caroli et al., 2004). Fontanesi et al. (2015) studied the polymorphisms of *CSN3* and *PAEP* genes at the DNA level in Reggiana sires.

Recently, *CSN2* gene has attracted significant scientific and practical interests, this is due to the fact that some of its variants have been reported to potentially impact human health (Summer et al., 2020; Thiruvengadam et al., 2021; Giribaldi et al., 2022). Several studies, but with some contrasting results, have suggested that the digestion of bovine β-CN variants carrying the amino

99 acid histidine at position 67 of the mature secreted protein (namely A1, B, C, F, G and L variants,
100 collectively referred to as A1 “like” variants; Summer et al., 2020) could induce milk
101 allergy/intolerance, gastrointestinal discomfort, and increase the risk for several not-transmissible
102 syndromes, including neurological disorders (Caroli et al., 2009; Brooke-Taylor et al., 2017;
103 Summer et al., 2020; Thiruvengadam et al., 2021; Cieślińska et al., 2022; Fernández-Rico et al.,
104 2022; Giribaldi et al., 2022; Jiménez-Montenegro et al., 2022). Conversely, the digestion of bovine
105 β-CN variants with proline at position 67 (A2, A3, D, E, H1, H2 and I variants, referred to as A2
106 “like” variants or A2 milk) could improve milk digestion efficiency and have antibacterial and
107 antioxidant activities. In view of possible human health benefits of A2 “like” milk, in addition to
108 avoid the A1 “like” adverse effects, some farmers have started adopting strategies to increase the
109 frequency of the β-CN A2 variant in their herds (Jiménez-Montenegro et al., 2022). From another
110 perspective, it is also important to underline that the β-CN A2 variant has been associated with
111 negative effects on cheese-making properties, positioning this variant at the crossroad of two
112 potential contrasting interests (Bittante et al., 2012; Gustavsson et al., 2014; Gai et al., 2021; Bisutti
113 et al., 2022). Among the variants of the *CSN2* gene, the β-CN B has been suggested to provide the
114 most positive cheese-making properties to milk, especially when in haplotype with the κ-CN B
115 variant (Comin et al., 2008).
116 Although many genetic variants have been identified in the *CSN3* gene, the two most
117 common forms in nearly all cattle breeds are variants A and B. In most European breeds the A
118 variant is prevalent than the B variant (Caroli et al., 2009; Chessa et al., 2020; Mahmoudi et al.
119 2020; Salmasi et al., 2020). The impact of *CSN3* polymorphisms on cheese-making properties is
120 well established (Di Stasio and Mariani, 2000). Variant κ-CN B has been associated with higher
121 levels of both glycosylated fraction and total k-casein as well as casein number (weight percentage
122 of casein to total protein which is an indicator of cheese yield), micelles with a smaller average
123 diameter (size), increased and faster reactivity with rennet enzyme and superior rheological
124 characteristics of the curd compared to the A variant (Di Stasio and Mariani, 2000). Variant κ-CN

E, found only in a few cattle breeds, and not yet identified in Reggiana and Modenese breeds, has been associated with lower milk protein content, and unfavorable milk coagulation properties, compared to the other two mentioned alleles, likely due to micelles with a lower percentage of κ-CN (Caroli et al., 2000; Hallén et al., 2008).

At least eight variants of the *PAEP* gene (which encodes the most important whey protein known as β-lactoglobulin, BLG or LGB) have been reported thus far in *Bos taurus*. Variants A and B are the most common in many breeds (Caroli et al., 2009). The BLG-A variant has been associated with higher concentrations of BLG in milk, as a result of increased transcriptional activity of the *PAEP* gene, total whey proteins and total milk proteins compared to the B variant (Caroli et al., 2000; Bonfatti et al., 2010). On the other hand, the BLG-B variant has been associated with improved rennet coagulation properties and higher milk fat content, enhancing cheese-making properties (e.g., Schaar et al., 1985; Aleandri et al., 1990).

In this study we genotyped a total of about 3,300 cattle from the Reggiana and Modenese breeds. Our objectives were i) to provide an updated overview of the variability in the *CSN2, CSN3* and *PAEP* genes, and ii) retrospectively analyse the trend of allele frequency changes at these major milk protein genes over time, going back to ~50 years, comparing our findings with the data from previous studies conducted in the same two breeds.

2. Material and methods

2.1. Investigated cattle populations

We analysed 3,298 cattle including 2,734 Reggiana cattle, born between 2000 and 2020, and kept on 100 farms located in the North of Italy and 564 Modenese cattle, born between 2005 and 2021, and kept on about 60 farms, distributed in the North of Italy. All analysed animals were registered in their herd books managed by ANABoRaRe. Individual biological samples (blood or

149 buccal swabs) were collected from 2019 to 2021 for routine monitoring analyses, not specifically
150 for this study. Therefore, no ethical issues were relevant for this study.
151
152 *2.2. Genotyping data*
153 DNA was extracted using the Wizard Genomic DNA Purification kit (Promega Corporation,
154 Madison, WI, USA). All animals were genotyped with the GeneSeek® Genomic Profiler™ (GGP)
155 Bovine 150K Array using standard genotyping protocols, based on the supplier's recommendations.
156 A total of 12 single nucleotide polymorphism (SNP) markers were included in this array. As this
157 array did not provide free access to the SNP that can distinguish between the A1 and A2 "like"
158 alleles of the *CSN2* gene, for only a subset of the genotyped cattle (600 Reggiana and 300
159 Modenese) this information was disclosed (BCN A2 test, Table 1). All SNPs were mapped on the
160 ARS-UCD2.0 assembly of the *Bos taurus* genome. More details of the considered SNPs are
161 reported in Table 1.
162
163 *2.3. Gene and protein nomenclature and correspondence between DNA (SNP) alleles and milk*
164 *protein variants*
165 In this study we used the official gene nomenclature and related gene symbols when the
166 information directly referred to the genes (casein beta, *CSN2*; casein kappa, *CSN3*; and
167 progestagen-associated endometrial protein, *PAEP*). The corresponding proteins were referred as β-
168 CN, κ-CN and BLG, respectively (as also reported above) and the protein variants were indicated
169 with the letters (Caroli et al., 2010). Gene alleles were indicated with the gene symbol followed by
170 an asterisk and the letter of the considered variant (for instance: *CSN2**A1 for the β-CN A1 variant;
171 *CSN2**A2 for the β-CN A2 variant; *CSN3**A for the κ-CN A variant; *CSN3**B for the κ-CN B
172 variant; *PAEP**A for the BLG A variant; *PAEP**B for the BLG B variant; and so on). The reported
173 correspondences between gene alleles and protein variants can make possible to align our results
174 with the results reported by previous studies (see below).

175 The SNP array used allows for genotyping of missense mutations in the *CSN2*, *CSN3* and
176 *PAEP* genes. The correspondence of SNP alleles with encoded milk protein variants is detailed in
177 Table 1. To align the results obtained in our study with the information reported in previous studies,
178 we utilized the nomenclature of milk protein variants (with the specific information at the protein
179 level). SNP tests on the chip covered eight variants/variant combinations (A3, B, C, E, F, H1, H2+I
180 and L) of the β-CN protein (Table 1). The samples that did not carry any of the eight
181 variants/variant combinations covered in the chip, were classified as carrying allele *A*, including
182 alleles A1 and A2 (or protein variants β-CN A1 and A2). Only for the subset of cattle genotyped
183 with the BCN A2 test, differentiation between A1 and A2 alleles was possible (Table 1).
184 Genotyping tests at the *CSN3* gene distinguished the three *CSN3* alleles C, D and E (corresponding
185 to the κ-CN C, D and E variants; Table 1) and two allele groups or protein variant groups (A and B;
186 Table 1). Similarly, at the *PAEP* gene, two allele groups (referred to as *PAEP**A and *PAEP**B
187 alleles in the manuscript, and corresponding to the BLG-A and BLG-B variants) were
188 distinguishable.
189
190 *2.4. Data editing and statistical analyses*
191 Genotyping data for the investigated genes were pruned from all SNP chip genotyping results
192 and analysed using PLINK 1.9 software (Chang et al., 2015) for quality control. Only data with an
193 animal genotyping call rate > 0.95, SNP call rate > 0.90 and minor allele frequency (MAF) ≥ 0.001
194 were retained for further statistical analyses. Allele and genotype frequencies, as well as Hardy-
195 Weinberg equilibrium for each SNP/breed were estimated using the R statistical program. The
196 relative frequencies of the *CSN2**A1 and A2 alleles (i.e., A1/A1+A2 and A2/A1+A2) were
197 calculated in the subset of cattle (Reggiana, no. 600; Modenese, no. 300) from which the
198 corresponding information was disclosed from the array. These relative frequencies were then used
199 to estimate the frequencies of the *CSN2**A1 and *CSN2**A2 alleles in the entire dataset (2,734

Reggiana and 564 Modenese cattle), as a proportion based on the observed frequency of the A1+A2 alleles in the genotyped animals.

In the subset of Reggiana and Modenese cattle, the *CSN2-CSN3* haplotypes were reconstructed using PHASE (version 2.1.1) software (Stephens and Donnelly, 2003) and the frequencies of haplotypes were estimated for each breed. The composite *CSN2-CSN3* genotypes were determined using the FREQ procedure of SAS (version 9.4; SAS Inst. Inc. Cary, NC, USA).

The obtained results on milk protein variant frequencies were compared with results published by other authors in previous studies (Mariani and Russo, 1971; Russo and Mariani, 1972; Russo et al., 1984; Caroli et al., 2004; Fontanesi et al., 2015). The years of publication of these studies, along with the year of the current study (set as 2024), defined the time points considered for the evaluating respective allele frequencies and related comparisons across time points. Frequencies of the same alleles obtained in different studies were compared using Fisher's exact tests or Chi square tests, as appropriate. Considering the sparse data and the limited number of data points, only when allele frequencies were available in at least four time points (based on the mentioned references, in addition to the current study), simple linear regression models were fitted to evaluate the trend of allele frequency differences. Frequencies of *CSN2**A1 "like" alleles (β-CN A1 "like" variants) and A2 "like" alleles (β-CN A2 "like" variants) were also estimated based on allele frequencies published from previous studies and those obtained in our subsamples.

3. Results

3.1. Allele and genotype frequencies of protein variants as inferred from SNP genotyping data

Allele frequencies for the milk protein genes under investigation are provided in Table 2. Supplementary Table S1 presents the frequencies for the subset of genotyped animals analysed with the BCN A2 test. All SNPs were found to be in Hardy-Weinberg equilibrium

At the *CSN2* gene, out of the 10 alleles/allele groups that could be analysed, seven were detected in Reggiana cattle, with four also observed in the Modenese breed (Table 2). Alleles A3, F

and H2+I were very rare (~1% or lower) and were only identified in the Reggiana breed, which had a larger number of genotyped animals compared to the other breed. The most frequent allele in both breeds was *CSN2**A2 (Table 2). Opposite frequencies were instead observed when we considered the A1 "like" and the A2 "like" alleles in the two breeds (Table 2): A2 "like" was the most common in the Reggiana breed (58.8%) and the A1 "like" was the most common in the Modenese breed (62.7%). A total of 11 different *CSN2* genotypes were found in the Reggiana population, with five having a frequency ≥5%: 37.7% were homozygous A2/A2 and about 50% were A2 heterozygous carriers (Table S2). In the Modenese breed, a total of 10 different *CSN2* genotypes were identified, with one having a frequency lower than 5%, and the most common genotype was A2/B, accounting for ~26% of the genotyped animals (Table S3).

At the *CSN3* gene, the C and D alleles were not observed in any Reggiana and Modenese cattle. For both breeds, allele group B was the most common, accounting ~56% in Reggiana and ~70% in Modenese breeds. Allele group A was the second most frequent in both Reggiana and Modenese breeds, representing 41.8% and 29.7%, respectively (Table 2). *CSN3**E was the rarest allele in both breeds, with frequencies of 2.7% in Reggiana and 0.3% in Modenese cattle.

At the *PAEP* gene, the frequency of A and B allele groups was opposite in the two breeds: A was the most frequent in Reggiana (61%) and B was the most frequent in Modenese (~68%) breeds.

As the *CNS2* and *CNS3* genes are closely linked on BTA6, we estimated their haplotypes and evaluated their haplotype frequencies in Reggiana and Modenese cattle, using a subset of 900 cattle that were genotyped to distinguish the *CSN3**A1 and A2 alleles. The frequencies of the *CSN2-CSN3* haplotypes are shown in Table 3. We identified a total of seven haplotypes, with a frequency ≥ 5% in at least one breed. The most common *CSN2-CSN3* haplotypes were A2-B (40.7%), A2-A (21.8%) and B-A (9.1%) in Reggiana cattle, and A2-B (29.8%), A1-B (13.9%) and B-B (13.6 %) in Modenese cattle. The composite *CSN2-CSN3* genotypes for the Reggiana and Modenese breeds are summarised in Tables S2 and S3, respectively.

252 *3.2. Over time evaluation of milk protein variant frequencies in Reggiana and Modenese*

253 A few other studies focusing on the Reggiana and Modenese breeds, published starting from
254 more than 50 years ago, have already reported allele frequencies in the three milk protein genes
255 under investigation (Mariani and Russo, 1971; Russo and Mariani, 1972; Russo et al., 1984; Caroli
256 et al., 2004; Fontanesi et al., 2015). We, therefore, have harmonized allele nomenclature, referred to
257 the protein variants, and used the reported results to obtain a comprehensive picture of allele
258 frequency distribution in these breeds over time. The year of publication of these studies served as
259 the reference time point for our population genetic analyses.

260 Combining the previously reported results with the results obtained in this study, we were
261 able to cover three to five time points for each gene/allele group combination in the Reggiana breed,
262 spanning approximately 40 to 50 years (Figure 1; Table S4). In the Modenese breed, which has
263 been less studied than the Reggiana breed, we could only cover two to three time points (including
264 the current study), but that they still encompassed a similar time frame to that reported for the
265 Reggiana cattle (Table S4).

266 In Reggiana, the frequency of the *CSN2**A2 allele has increased since Russo et al. (1984),
267 when it was first reported (45.7%; Figure 1a, Table S4), reaching 62.5% in the current study, with a
268 strong significant statistical difference between the two time points (P <0.00001). Allele *CSN2**B
269 has consistently increased in Reggiana (Figure 1b) over the explored time window of more than 50
270 years (R^2 of the intercept line was 0.8437), with a corresponding decrease of allele *CSN2**A.
271 Starting from Mariani and Russo (1971), who reported a frequency of the *CSN3**B allele at 46.6%,
272 the difference with our current study (57.7%) was highly significant (P <0.00001). However, the
273 increasing trend of this allele slowed down when comparing the frequency of 55.9% reported by
274 Fontanesi et al. (2015) with the frequency obtained in this study, with no statistical differences
275 between these two time points (P = 0.23). A similar increasing trend was observed for allele A of
276 the *PAEP* gene in Reggiana (R^2 = 0.9286), and in turn, a corresponding reduction in the frequency
277 of allele B at this gene, for which four time points were available (Figure 1c). When comparing the

two extreme time points (Mariani and Russo, 1971; and the current study) for *PAEP*, the allele frequency difference was highly significant ($P<0.00001$). However, when comparing the allele frequency reported by Fontanesi et al. (2015) to the allele frequency obtained in this study, the difference was not significant ($P = 0.99$).

In the Modenese breed, the frequency of the *CSN2**A2 allele decreased slightly from Russo et al. (1984) to the current study (from 41.8% to 39.3%), but the difference was not statistically significant ($P = 0.48$). At the *CSN3* gene, the Modenese population showed an increasing trend in the frequency of the *CSN3**B allele: the frequency obtained in this study (68.8%) was statistically different ($P<0.00001$) from what was reported in both previous studies (Russo and Mariani, 1972: 49.5%; Russo et al., 1984: 41.0%). Additionally, a small increase in the frequency of the *PAEP**A allele was observed in this breed, although it was not statistically significant ($P = 0.11$), when compared to the results reported by Russo and Mariani (1972) and the results obtained in this study (Table S4).

4. Discussion

Over the last decades, the study of milk protein genetic variants in cattle has received great attention for their effect on quantitative and qualitative milk and cheese-making properties, as well as on human health. In our study we analyzed SNPs in the *CSN2*, *CSN3* and *PAEP* cattle genes and translated these polymorphisms included in a SNP array into milk protein variants. This allowed us to compare the information obtained in Reggiana and Modenese breeds, with studies reported in these breeds a few decades ago, as well as with other breeds analysed using different methods/approaches. The time window that we covered by considering previous studies on Reggiana and Modenese was about five decades (Mariani and Russo, 1971; Russo and Mariani, 1972; Russo et al., 1984; Caroli et al., 2004; Fontanesi et al., 2015). It should also be noted that cattle investigated in these studies might have been born several years before the publication year, expanding the time window of this temporal comparative analysis. Allele frequencies within

(considering an overtime analysis) and among breeds may vary according to the combination of different breeding objectives, genetic history, possible admixture events and random genetic drift (Caroli et al., 2004; Chessa et al., 2020; Hohmann et al., 2021).

In both Reggiana and Modenese breeds, the *CSN2**A2 allele, considered the primitive allele from which other alleles originated (Caroli et al., 2009; Brooke-Taylor et al., 2017), was the most common (58.6% and 37.3%, respectively; Table 2). A high frequency for this allele has been also previously found in other dairy and dual-purpose cattle breeds, such as Danish Jersey, Italian Holstein-Friesian, Norwegian Red, and Simmental (Bonfatti et al., 2010; Poulsen et al., 2013; Massella et al., 2017; Chessa et al., 2020; Sanchez et al., 2020; Sebastiani et al., 2020). Our study also identified for the first time the rare *CSN2**F allele in the Reggiana population (1.1%; Table 2). It is worth mentioning that previous studies in this breed that investigated milk protein variants (Mariani and Russo, 1971; Russo et al., 1984; Caroli et al., 2004) used electrophoretic analyses (i.e., starch urea gel and isoelectric focusing), which are unable to identify this variant. Therefore, we cannot exclude the possibility that this allele was also present in the past Reggiana population. This rare allele has also been identified in Italian Holstein cattle, with a very low frequency (0.6%; Massella et al., 2017) and in two local Danish dairy breeds, but with a higher frequency (~20%; Poulsen et al., 2017). Additionally, at the *CSN2* gene, in Reggiana cattle, we identified another very rare variant profile (0.1%), which could be classified as derived from H2 or I variants. Since the H2 variant is rare and has not been identified in European cattle breeds, except in Normande breed (Caroli et al., 2009), we can assume that the SNP test BCN_8178 (putatively for H2 + I variants discovery; Table 1) can identify the *CSN2**I allele. This variant has already been found in several European cattle breeds, including Italian Simmental and Italian Holstein (Caroli et al., 2009; Sanchez et al., 2020; Sebastiani et al., 2020). Particularly, the I allele has been reported to be one of the most frequent alleles (~31%) in the Normande breed and with quite high frequencies (from ~8 to 17%) in other French breeds (Sanchez et al., 2020). Other rare *CSN2* alleles (E, H1, and L), that could be identified by the genotyping assays used here, were not present in Reggiana and Modenese

breeds. The *CSN2**E allele is known to be present only in the Piedmontese breed, and the H1 and L alleles have never been observed in any European cattle breeds (Caroli et al., 2009; Gallinat et al., 2013). The *CSN2**D and G alleles (not included in the genotyping assays used here) are known to be rare and it is unlikely to find them in the studied populations. In particular, the D allele has only been identified in *Bos indicus*, while the nucleotide sequence of the G variant remains controversial (Caroli et al., 2009).

As mentioned earlier, β-CN proteins have been suggested to have some effects on human health (Thiruvengadam et al., 2021; Giribaldi et al., 2022). The putative adverse effects of the β-CN A1 "like" variants (with p.67His) have been related to the formation of the bioactive opioid peptide of seven amino-acid (fragment 60–66 of the protein), known as β-casomorphin-7 (BCM7). This peptide, upon acid hydrolysis in the gastric environment, enzymatic hydrolysis in the small intestine, microbiome activity, or technological food processing, is released at a higher concentration from A1 "like" variants than from A2 "like" variants (Brooke-Taylor et al., 2017; Cieślińska et al., 2022). Due to the potential negative impact on human health of the β-CN A1 "like" milk, some farmers and dairy industries have implemented strategies to produce A2 "like" milk, usually for fresh consumption (Jiménez-Montenegro et al., 2022). However, many conflicting studies on this topic have not provided enough scientific evidence to confirm any causative effect between the consumption of β-CN A1 "like" milk and health problems in human, justifying the introduction of breeding strategies and changes in dietary habits (Caroli et al., 2009; Summer et al., 2020; Giribaldi et al., 2022). It is also worth mentioning that almost all milk produced by Reggiana and Modenese cows is used to obtain long-ripened PDO Parmigiano-Reggiano cheese, whose maturation steps may drastically reduce the level of released BCM7 below any detectable *in vivo* opioid activity (De Noni et al., 2015). On the other hand, several studies have provided clear evidence that the β-CN A2 variant confers worse coagulation properties to the milk (i.e., higher rennet coagulation time and curd-firming time, and lower curd firmness) and lower cheese yield than the A1, B and C variants (Amalfitano et al., 2019; Bittante et al., 2012; Gustavsson et al.,

2014; Gai et al., 2021; Bisutti et al., 2022). Therefore, in this context, the relatively recent increase in the frequency of the *CSN2**A2 allele in Reggiana should be considered with caution. It is also not known why over the last ~20 years this breed has had a significant increase in the frequency of the *CSN2**A2 allele, as there has been no direct selection for this allele. Breeding objectives have been focused on improving cheese-making properties of Reggiana milk, which, in turn, may have favoured allele A1. It is possible that genetic drift, common in small populations, resulted from the unconscious more frequent use of sires carrying allele A2, contributed to this recent trend. Another possible explanation for this trend, at least in part, could be the linkage disequilibrium of this allele with allele B at the *CSN3* gene (the most frequent *CSN2-CSN3* haplotype was A2-B in both breeds).

In both Reggiana and Modenese breeds, we observed a higher frequency of allele B than allele A at the *CSN3* gene, with a significant increase in its frequency over the last five decades. This increase seems primarily due to direct actions taken since the first demonstration of its importance in cheese-making properties (Mariani and Russo, 1972; Russo and Mariani, 1972; Russo et al., 1984), shaping the trend observed in our study. With more previous studies setting time points in Reggiana than in Modenese, we can see that the recent increasing trend in Reggiana has slowed down, compared to what was reported by Fontanesi et al. (2015), almost a decade ago. The reason for this recent change is unclear, especially considering that Reggiana sires have been routinely genotyped for *CSN3* since the beginning of this century, with a preference for sires carrying allele B. To prevent the reversal of this trend, a new composite genetic index has been approved by the Reggiana Cattle Breed Herd Book Technical Commission (ANABoRaRe, 2023), which includes an additional weight then the animals carry the *CSN3**B allele. In this study we also identified, for the first time, in both Reggiana and Modenese breeds, the rare *CSN2**E allele (with frequencies of 2.7% and 0.3%, respectively). This allele was found to be in complete linkage disequilibrium (i.e., forming the haplotype, data not shown) with the *CSN2**A1 allele, confirming previous reports in other breeds (Gai et al., 2021). The *CSN2**E allele was initially discovered in German taurine breeds and has since then detected in various other breeds including Italian Brown

and Italian Friesian (Caroli et al., 2009; Gai et al., 2021). The presence of the *CSN2**E allele in both Reggiana and Modenese breeds highlights the need for specific monitoring programs to prevent an increase in its frequency, given the negative impact of this variant on several cheese-making properties of the milk (Caroli et al., 2000; Hallén et al., 2008).

For the *PAEP**B allele, determining a whey protein variant positively associated with cheese making properties (Schaar et al., 1985; Aleandri et al., 1990), we have observed a trend of reduction in its frequency in Reggiana over the last five decades, with a slowdown in the last decade. Information on the genotype at this gene is not routinely generated for Reggiana and Modenese cattle, so breeders cannot base their selection strategies on this locus. Therefore, it would be important to establish a genotyping service for candidate sires for the *PAEP* gene to provide breeders with information and evaluate how to reverse the decreasing trend of allele B in Reggiana. It is also possible that the genotype at this gene may not have a significant effect on cheese-making properties in Reggiana, as also reported in other studies (Bonfatti et al., 2010), and the decreasing trend may be better explained by genetic drift in a small population. The frequencies of *PAEP* alleles have remained stable in the Modenese breed over the same time period, suggesting that genetic drift may not have influenced allele frequency at this gene in this case.

5. Conclusions

In this study we genotyped nearly half of the total cattle populations of two iconic autochthonous Italian breeds, Reggiana and Modenese, to provide an updated allele frequency for three major milk protein genes: *CSN2*, *CSN3* and *PAEP*. These genes encode important milk proteins: β-casein, κ-casein and β-lactoglobulin, respectively. The results obtained were compared to those reported in previous studies that investigated the same milk protein genes in the same breeds. This comparative analysis provided information useful to evaluate the effects of the selection and conservation programs in Reggiana and Modenese at the population genetic levels. The time frame that we reconstructed was of ~50 years, corresponding to about 8-10 generations.

For the *CNS2* and *CSN3* genes, we identified a few alleles that had not been previously detected in these two breeds. The trends in allele frequencies, especially in the Reggiana breed, could be seen as either positive, with an increase in the frequency of alleles with favorable effects on cheese yield, or negative, with a decrease in the frequency of favorable alleles for the same traits. Therefore, adjustments to breeding strategies and conservation programs may be needed to maintain genetic properties in the two breed populations that can positively contribute to further address the production of high-quality breed-branded Parmigiano-Reggiano cheeses.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

CRediT authorship contribution statement

S. Dall'Olio: Conceptualization, Methodology, Formal analysis, Investigation, Funding acquisition, Writing – original draft, Writing – review & editing. **G. Schiavo:** Conceptualization, Methodology, Writing – review & editing. **S. Bovo:** Methodology, Investigation, Writing – review & editing. **A. Ribani:** Formal analysis, Investigation. **F. Bertolini**: Formal analysis, Investigation. **L. Fontanesi:** Conceptualization, Methodology, Supervision, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing.

Fundings

This study was funded by the PSRN (Programma di Sviluppo Rurale Nazionale) Dual Breeding 2 (co-funded by the European Agricultural Fund for Rural Development of the European Union and by the Italian Ministry of Agriculture - MASAF) and by the European Union - NextGenerationEU under the National Recovery and Resilience Plan (PNRR) - Mission 4 Education and research - Component 2 From research to business - Investment 1.1 Notice Prin 2022 - DD N. 104 of

2/2/2022, from title Integration of novel phenotyping and omics-approaches for the sustainable management and breeding of autochthonous cattle genetic resources, proposal code 2022 2022FFR89X – CUP J53D23009990006.

Ethics statement

All animals were not raised for the purpose of this study and were not treated in any way. Sampling was performed in accordance with National Animal Protection Law (Legislative Decree No26/2014; https://www.gazzettaufficiale.it). Therefore, no ethical issues were involved in this study. The cows included in this study belonged to commercial herds and were not subjected to any invasive procedures. Blood samples were collected during the routine milk recording coordinated by technicians working at the National Reggiana Cattle Breeders Association (ANABoRaRe, Mancasale, Reggio Emilia, Italy) and therefore authorized on their duties.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

This work is dedicated to the memory of Prof. Vincenzo Russo, emeritus of the University of Bologna. We thank ANABoRaRe personnel, all Reggiana and Modenese farmers, and technical staff that collaborated in this study.

References

Aleandri, R., Buttazzoni, L. G., Schneider, J. C., Caroli, A., Davoli, R., 1990. The effects of milk protein polymorphisms on milk components and cheese-producing ability. J. Dairy Sci. 73, 241-255. https://doi.org/10.3168/jds.S0022-0302(90)78667-5

459 Amalfitano, N., Cipolat-Gotet, C., Cecchinato, A., Malacarne, M., Summer, A., Bittante, G., 2019.
460 Milk protein fractions strongly affect the patterns of coagulation, curd firming, and syneresis.
461 J. Dairy Sci. 102, 2903-2917. https://doi.org/10.3168/jds.2018-15524
462 Associazione Italiana Allevatori, 2022. Official statistics of milk recording activity by the Italian
463 Breeders Association, AIA; http://bollettino.aia.it/, accessed on the 15th February 2024.
464 Associazione Nazionale Allevatori Bovini di Razza Reggiana, 2023. ANABoRaRe,
465 http://www.razzareggiana.it, accessed on the 15th of February 2024.
466 Bertolini, F., Moscatelli, G., Schiavo, G., Bovo, S., Ribani, A., Ballan, M., Bonacini, M., Prandi,
467 M., Dall'Olio, S., Fontanesi, L., 2022. Signatures of selection are present in the genome of
468 two close autochthonous cattle breeds raised in the North of Italy and mainly distinguished for
469 their coat colours. J. Anim. Breed. Genet. 139, 307-319. https://doi.org/10.1111/jbg.12659
470 Bertolini, F., Schiavo, G., Bovo, S., Sardina, M.T., Mastrangelo, S., Dall'Olio, S., Portolano, B.,
471 Fontanesi, L., 2020. Comparative selection signature analyses identify genomic footprints in
472 Reggiana cattle, the traditional breed of the Parmigiano-Reggiano cheese production system.
473 Animal 14, 921-932. https://doi.org/10.1017/S1751731119003318
474 Bisutti, V., Pegolo, S., Giannuzzi, D., Mota, L.F.M., Vanzin, A., Toscano, A., Trevisi, E., Ajmone
475 Marsan, P., Brasca, M., Cecchinato, A., 2022. The β-casein (CSN2) A2 allelic variant alters
476 milk protein profile and slightly worsens coagulation properties in Holstein cows. J. Dairy
477 Sci. 105, 3794-3809. https://doi.org/10.3168/jds.2021-21537
478 Bittante, G., Penasa, M., Cecchinato, A., 2012. Invited review: Genetics and modeling of milk
479 coagulation properties. J. Dairy Sci. 95, 6843-6870. https://doi.org/10.3168/jds.2012-5507
480 Bonfatti, V., Di Martino, G., Cecchinato, A., Vicario, D., Carnier, P., 2010. Effects of beta-kappa-
481 casein (CSN2-CSN3) haplotypes and beta-lactoglobulin (BLG) genotypes on milk production
482 traits and detailed protein composition of individual milk of Simmental cows. J. Dairy Sci. 93,
483 3797-3808. https://doi.org/10.3168/jds.2009-2778

484 Bovo, S., Schiavo, G., Kazemi, H., Moscatelli, G., Ribani, A., Ballan, M., Bonacini, M., Prandi, M.,
485 Dall'Olio, S., Fontanesi, L., 2021. Exploiting within-breed variability in the autochthonous
486 Reggiana breed identified several candidate genes affecting pigmentation-related traits,
487 stature and udder defects in cattle. Anim. Genet. 52, 579-597.
488 https://doi.org/10.1111/age.13109
489 Brooke-Taylor, S., Dwyer, K., Woodford, K., Kost, N., 2017. Systematic Review of the
490 Gastrointestinal Effects of A1 Compared with A2 β-Casein. Adv. Nutr. 8, 739-748.
491 https://doi.org/10.3945/an.116.013953
492 Caroli, A., Bolla, P., Budelli, E., Barbieri, G., Leone, P., 2000. Effect of κ-casein E allele on
493 clotting aptitude of Italian Friesian milk. Zoot. Nutr. Anim. 26, 127-130.
494 Caroli, A., Chessa, S., Bolla, P., Budelli, E., Gandini, G.C., 2004. Genetic structure of milk protein
495 polymorphisms and effects on milk protein production traits in a local dairy cattle. J. Anim.
496 Breed. Genet. 121, 119-127. https://doi.org/10.1111/j.1439-0388.2003.00443.x
497 Caroli, A.M., Chessa, S., Erhardt, G.J., 2009. Invited review: milk protein polymorphisms in cattle:
498 effect on animal breeding and human nutrition. J. Dairy Sci. 92, 5335–5352.
499 https://doi.org/10.3168/jds.2009-2461
500 Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015.
501 Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience
502 4, 7. https://doi.org/10.1186/s13742-015-0047-8
503 Chessa, S., Gattolin, S., Cremonesi, P., Soglia, D., Finocchiaro, R., Van Kaam, J.T., Marusi, M.,
504 Civati, G., 2020. The effect of selection on casein genetic polymorphisms and haplotypes in
505 Italian Holstein cattle. Ital. J. Anim. Sci. 19, 833-839.
506 https://doi.org/10.1080/1828051X.2020.1802356
507 Cieślińska, A., Fiedorowicz, E., Rozmus, D., Sienkiewicz-Szłapka, E., Jarmołowska, B., Kamiński,
508 S., 2022. Does a little difference make a big difference? Bovine β-casein A1 and A2 variants

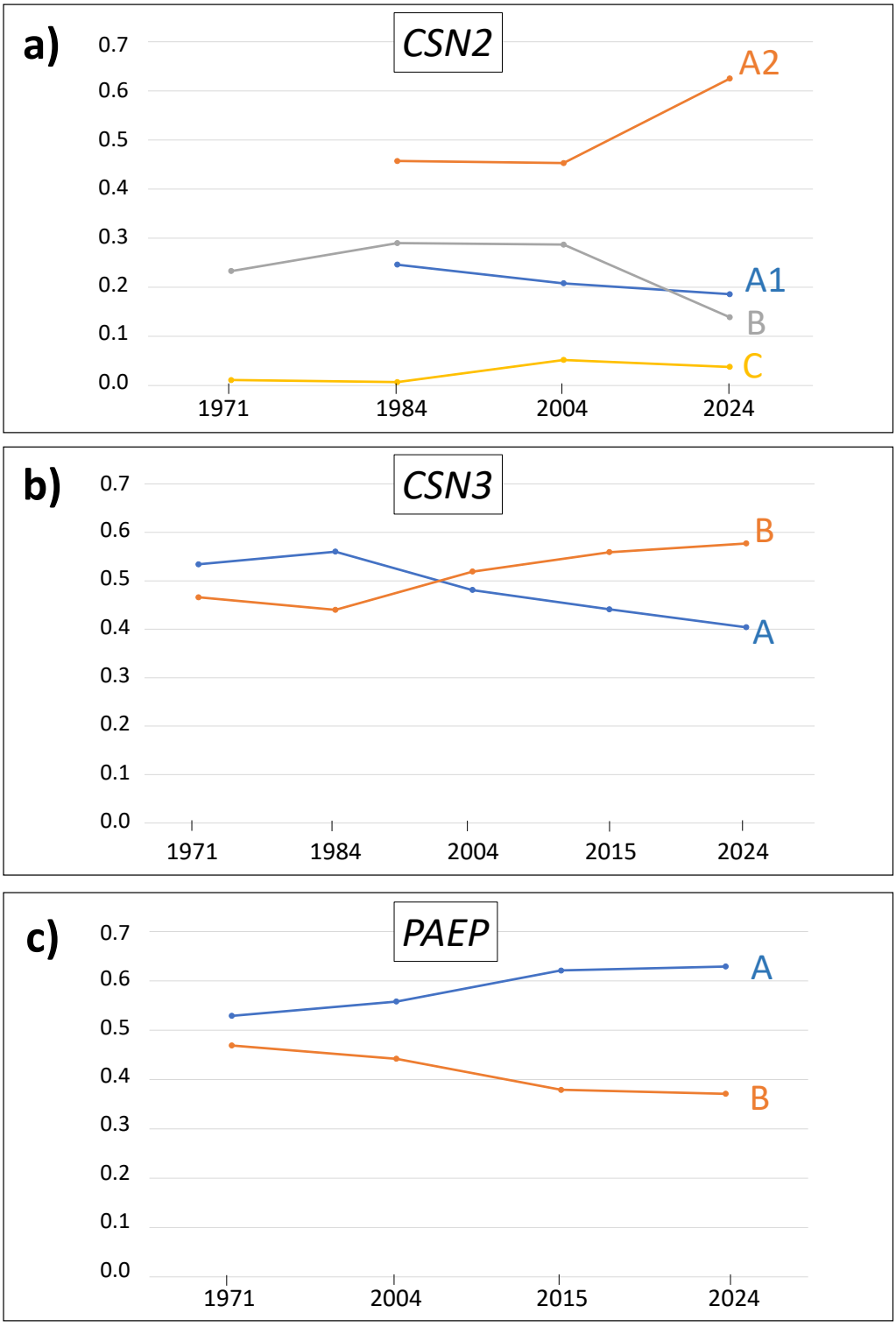
509 and human health-an update. Int. J. Mol. Sci. 23, 15637.
510 https://doi.org/10.3390/ijms232415637
511 Comin, A., Cassandro, M., Chessa, S., Ojala, M., Dal Zotto, R., De Marchi, M., Carnier, P., Gallo,
512 L., Pagnacco, G., Bittante, G., 2008. Effects of composite beta- and kappa-casein genotypes
513 on milk coagulation, quality, and yield traits in Italian Holstein cows. J. Dairy Sci. 91, 4022-
514 4027. https://doi.org/10.3168/jds.2007-0546
515 De Noni, I., Stuknyte, M., Cattaneo, S., 2015. Identification of β-casomorphins 3 to 7 in cheeses
516 and in their in vitro gastrointestinal digestates. LWT - Food Sci, Technol, 63, 550-555.
517 https://doi.org/10.1016/j.lwt.2015.03.036
518 Di Stasio, L., Mariani, P., 2000. The role of protein polymorphism in the genetic improvement of
519 milk production. Zoot. Nutr. Anim. 26, 69-90.
520 Fernández-Rico, S., Mondragón, A.D.C., López-Santamarina, A., Cardelle-Cobas, A., Regal, P.,
521 Lamas, A., Ibarra, I.S., Cepeda, A., Miranda, J.M., 2022. A2 Milk: New perspectives for food
522 technology and human health. Foods 11, 2387. https://doi.org/10.3390/foods11162387
523 Fontanesi, L., Scotti, E., Samorè, A.B., Bagnato, A., Russo, V., 2015. Association of 20 candidate
524 gene markers with milk production and composition traits in sires of Reggiana breed, a local
525 dairy cattle population. Livest. Sci. 176, 14-21. https://doi.org/10.1016/j.livsci.2015.03.022
526 Gai, N., Uniacke-Lowe, T., O'Regan, J., Faulkner, H., Kelly, A.L., 2021. Effect of protein
527 genotypes on physicochemical properties and protein functionality of bovine milk: A review.
528 Foods 10, 2409. https://doi.org/10.3390/foods10102409
529 Gallinat, JL, Qanbari, S, Drögemüller, C, Pimentel, EC, Thaller, G, Tetens J., 2013. DNA-based
530 identification of novel bovine casein gene variants. J Dairy Sci. 96(1):699-709.
531 https://doi.org/10.3168/jds.2012-5908
532 Gandini, G., Maltecca, C., Pizzi, F., Bagnato, A., Rizzi, R., 2007. Comparing local and commercial
533 breeds on functional traits and profitability: the case of Reggiana dairy cattle. J. Dairy Sci. 90,
534 2004-2011. https://doi.org/10.3168/jds.2006-204

535 Giribaldi, M., Lamberti, C., Cirrincione, S., Giuffrida, M.G., Cavallarin, L., 2022. A2 milk and
536 BCM-7 peptide as emerging parameters of milk quality. Front. Nutr. 9, 842375.
537 https://doi.org/10.3389/fnut.2022.842375
538 Gustavsson, F., Buitenhuis, A.J., Johansson, M., Bertelsen, H.P., Glantz, M., Poulsen, N.A.,
539 Lindmark Månsson, H., Stålhammar, H., Larsen, L.B., Bendixen, C., et al., 2014. Effects of
540 breed and casein genetic variants on protein profile in milk from Swedish Red, Danish
541 Holstein, and Danish Jersey cows. J. Dairy Sci. 97, 3866-3877.
542 https://doi.org/10.3168/jds.2013-7312
543 Hallén, E., Wedholm, A., Andrén, A., Lundén, A., 2008. Effect of beta-casein, kappa-casein and
544 beta-lactoglobulin genotypes on concentration of milk protein variants. J. Anim. Breed.
545 Genet. 125, 119-129. https://doi.org/10.1111/j.1439-0388.2007.00706.x
546 Hohmann, L.G., Weimann, C., Scheper, C., Erhardt, G., König, S., 2021. Genetic diversity and
547 population structure in divergent German cattle selection lines on the basis of milk protein
548 polymorphisms. Arch. Anim. Breed. 64, 91-102. https://doi.org/10.5194/aab-64-91-2021
549 Jiménez-Montenegro, L., Alfonso, L., Mendizabal, J.A., Urrutia, O., 2022. Worldwide research
550 trends on milk containing only A2 β-casein: A bibliometric study. Animals (Basel) 12, 1909.
551 https://doi.org/10.3390/ani12151909
552 Mahmoudi, P., Rostamzadeh, J., Rashidi, A., Zergani, E., Razmkabir, M., 2020.A meta-analysis on
553 association between CSN3 gene variants and milk yield and composition in cattle. Anim.
554 Genet. 51, 369-381. https://doi.org/10.1111/age.12922
555 Mariani, P., Russo, V., 1971. Distribuzione delle varianti genetiche delle caseine e della *β*-
556 lattoglobulina nelle vacche di razza Reggiana. Riv. Zoot. 44, 310-322.
557 Massella, E., Piva, S., Giacometti, F., Liuzzo, G., Zambrini, A.V., Serraino, A., 2017. Evaluation of
558 bovine beta casein polymorphism in two dairy farms located in northern Italy. Ital. J. Food
559 Saf. 6, 131-133. https://doi.org/10.4081/ijfs.2017.6904

560 Poulsen, N., Rosengaard, A., Szekeres, B., Gregersen, V.R., Jensen, H.B., Larsen, L.B., 2017.
561 Protein heterogeneity of bovine β-casein in Danish dairy breeds and association of rare β-
562 casein F with milk coagulation properties. Acta Agr. Scand. Sec. A Anim. Sci. 66, 190-198.
563 https://doi.org/10.1080/09064702.2017.1342858
564 Poulsen, N.A., Bertelsen, H.P., Jensen, H.B., Gustavsson, F., Glantz, M., Månsson, H.L., Andrén,
565 A., Paulsson, M., Bendixen, C., Buitenhuis, A.J., Larsen, L.B., 2013. The occurrence of
566 noncoagulating milk and the association of bovine milk coagulation properties with genetic
567 variants of the caseins in 3 Scandinavian dairy breeds. J. Dairy Sci. 96, 4830-4842.
568 https://doi.org/10.3168/jds.2012-6422
569 Russo, V., Davoli, R., Bosi, P., Bruzzone, P., 1984. Varianti genetiche della beta-caseina a pH acido
570 nelle razze bovine Bruna, Frisona Italiana, Modenese e Reggiana. Zoot. Nutr. Anim. 20, 47-
571 53.
572 Russo, V., Mariani, P., 1972. Polimorfismo genetico delle caseine e della β-lattoglobulina nella
573 razza bovina Modenese. Riv. Zoot. 45, 43-51.
574 Salmasi, M., Moretti, R., Chessa, S., Sartore, S., Mimosi, A., Cornale, P., 2020. Genetic variability
575 of milk proteins in two cattle breeds of Piedmont region and the potential effects on milk
576 quality. Ital. J. Anim. Sci. 19, 1483-1489. https://doi.org/10.1080/1828051X.2020.1850217
577 Sanchez, M.P., Fritz, S., Patry, C., Delacroix-Buchet, A., Boichard, D., 2020. Frequencies of milk
578 protein variants and haplotypes estimated from genotypes of more than 1 million bulls and
579 cows of 12 French cattle breeds. J. Dairy Sci. 103, 9124-9141.
580 https://doi.org/10.3168/jds.2020-18492
581 Schiavo, G., Bovo, S., Ribani, A., Moscatelli, G., Bonacini, M., Prandi, M., Mancin, E., Mantovani,
582 R., Dall'Olio, S., Fontanesi, L., 2021. Comparative analysis of inbreeding parameters and runs
583 of homozygosity islands in 2 Italian autochthonous cattle breeds mainly raised in the
584 Parmigiano-Reggiano cheese production region. J. Dairy Sci. 23, S0022-0302(21)01064-X.
585 https://doi.org/10.3168/jds.2021-20915

586 Sebastiani, C., Arcangeli, C., Ciullo, M., Torricelli, M., Cinti, G., Fisichella, S., Biagetti, M., 2020.
587 Frequencies evaluation of β-casein gene polymorphisms in dairy cows reared in central Italy.
588 Animals (Basel) 10, 252. https://doi.org/10.3390/ani10020252
589 Schaar, J., Hansson, B., Pettersson, H. E., 1985. Effects of genetic variants of κ-casein and β-
590 lactoglobulin on cheesemaking. J. Dairy Res. 52, 429-437.
591 https://doi.org/10.1017/S002202990002433X
592 Stephens, M., Donnelly, P., 2003. A comparison of bayesian methods for haplotype reconstruction
593 from population genotype data. Am. J. Hum. Genet. 73, 1162-1169.
594 https://doi.org/10.1086/379378
595 Summer, A., Di Frangia, F., Ajmone Marsan, P., De Noni, I., Malacarne, M., 2020. Occurrence,
596 biological properties and potential effects on human health of β-casomorphin 7: current
597 knowledge and concerns. Crit. Rev. Food Sci. 60, 3705-3723.
598 https://doi.org/10.1080/10408398.2019.1707157
599 Thiruvengadam, M., Venkidasamy, B., Thirupathi, P., Chung, I.M., Subramanian, U., 2021. β-
600 Casomorphin: A complete health perspective. Food Chem. 337, 127765.
601 https://doi.org/10.1016/j.foodchem.2020.127765
602

Figure 1. Allele frequency trends at the three investigated genes (a: *CSN2*; b: *CSN3*; c: *PAEP*) over time in the Reggiana breed. The y axis reports the allele frequency of the corresponding allele. The indicated years in the x axis refer to the years of publication of the studies from which the information was retrieved (Mariani and Russo, 1971; Russo et al., 1984; Caroli et al., 2004; Fontanesi et al., 2015) and this study (2024). Only the most important alleles were included. For some alleles, information in a few time points was not available. All details are reported in Supplementary Table S4.



611 **Table 1.** Details of the single nucleotide polymorphisms (SNPs) analysed in the SNP array and correspondence with the predicted amino acid
612 substitutions for the encoded proteins by the *CSN2*, *CSN3* and *PAEP* genes.

Gene/encoded protein Chip SNP ID[a]	dbSNP ID	Position[b]	SNP allele; predicted amino acid substitution[c]	Disentangled milk protein variants/variant groups from the tested SNPs based on the genotyping strategy [d]
CSN2/β-CN				
BCN_6687	-	85,452,713	g.6687G>A; p.Glu36Lys	E (p.36Lys) *vs* all other variants (p.36Glu)
BCN_6690	rs721259074	85,452,710	g.6690G>A; p.Glu37Lys	C (p.37Lys) *vs* all other variants (p.37Glu)
BCN A2	rs43703011	85,451,298	g.8101A>C; p.His67Pro	A2 (p.67Pro) *vs* A1 (p.67His) [e]
BCN_8163	-	85,451,236	g.8163C>A; p.Leu88Ile	H1 (p.88Ile) *vs* all other variants (p.88Leu)
BCN_8178	rs109299401	85,451,221	g.8178A>C; p.Met93Leu	H2+I (p.93Leu) *vs* all other variants (p.93Met)
BCN_8219	rs43703012	85,451,180	g.8219C>A; p.His106Gln	A3 (p.106Gln) *vs* all other variants (p.106His)
BCN_8267	rs43703013	85,451,132	g.8267C>G; p.Ser122Arg	B (p.122Arg) *vs* all other variants (p.122Serl)
BCN_8356	rs433954503	85,451,043	g.8356C>T; p.Pro152Leu	F (p.152Leu) *vs* all other variants (p.152Pro)
BCN_8491	rs715383373	85,450,908	g.8491T>C; p.Val197Ala, not sequenced at protein level	L (p.197Ala) *vs* all other variants (p.197Val)
CSN3/k-CN				
GNSC319	rs43703016	85,656,772	g.13104A>C; p.Asp148Ala	A group (p.Asp148: A+F2+G1+H+I) *vs* B group (p.148Ala: B+G2+J) [f]
GNSC355	rs43703017	85,656,792	g.13124A>G; p. Ser155Gly	E (p.155Gly) *vs* all other variants (p.155Ser)
KappaCasein12951	rs716557965	85,656,619	g.12951G>A; p.Arg97His	C + D (p.97His) *vs* all other variants (p.97Arg) [g]
PAEP/BLG				

BetaLact_2	rs109625649	103,259,232	g.5263C>T; p.Ala118Val	A group (p.118Val: A+H) *vs* B group (p.118Ala: B+C+D+I+J+W)[d]

[a] ID of the SNP as indicated in the GeneSeek® Genomic Profiler™ Bovine 150K Array.
[b] Nucleotide position on *Bos taurus* chromosome 6 (BTA6) for *CSN2* and *CSN3* genes and on BTA11 for *PAEP* gene, as reported in the ARS-UCD2.0 genome assembly.
[c] Missense variants are reported considering the position in the following sequences: *CSN2*, GenBank accession number X14711, corresponding to the nucleotide sequence of allele *CSN2**A1; *CSN3*, GenBank accession number AY380228, corresponding to the nucleotide sequence of allele *CSN3**A; *PAEP*, GenBank accession number X14710, corresponding to the nucleotide sequence of allele *PAEP**B. Amino acid substitution in the mature protein encoded by the corresponding genes.
[d] Only variants detected in *Bos Taurus* were considered.
[e] The SNP test discriminates *CSN2**A1 "like" alleles (namely A1+B+ C+F+G+L alleles; indicated as variants if referred to the β-CN protein) and *CSN2**A2 "like" alleles (A2+A3+D+E+H1+H2+I alleles; indicated as variants if referred to the β-CN protein). Based on the genotyping strategy described in material and methods, here it was possible to distinguish A1 and A2 alleles.
[f] A group of *CSN3* gene includes also D and E alleles, B group includes the C allele. Specific tests for these alleles (E, C + D) were available in the chip.
[g] *CSN3**C and *CSN3**D alleles can be further identified considering the g.13104A>C SNP (g.13104C for the *CSN3**C allele and g.13104A for the *CSN3**D allele).

Table 2. Allele frequencies of protein variants at the *CSN*2, *CSN*3 and *PAEP* genes in the Reggiana and Modenese cattle breeds.

Gene/protein	Alleles[a]	Allele frequencies	
		Reggiana (no. 2,734)	Modenese (no. 564)
CSN2/β-CN	A (A1+A2 alleles)	0.760	0.600
	A1[b]	0.174	0.227
	A2[b]	0.586	0.373
	A3	0.001	0.000
	B	0.183	0.281
	C	0.044	0.119
	E	0.000	0.000
	F	0.011	0.000
	H1	0.000	0.000
	H2+I	0.001	0.000
	L	0.000	0.000
	A1 "like" alleles[c]	0.412	0.627
	A2 "like" alleles[c]	0.588	0.373
CSN3/k-CN	A (A+F2+G1+H+I alleles)	0.418	0.297
	B (B+G2+J alleles)	0.555	0.699
	C	0.000	0.000
	D	0.000	0.000
	E	0.027	0.003
PAEP/BLG	A (A+H alleles)	0.610	0.321
	B (B+C+D+I+J+W alleles)	0.390	0.679

[a] Alleles or allele groups identified based on the genotyping strategy and available information defined in Table 1.

[b] Frequencies of the A1 and A2 alleles of *CSN2* gene were estimated based on the relative proportion of what was reported for the genotyped subset of Reggiana (no. 600) and Modenese (no. 300) cattle (see Table S1 for the frequencies reported in the subsets of genotyped cattle).

[c] Determined in the present work based on allele frequencies of the *CSN2* gene. A1 "like" alleles = A1+B+C+F alleles; A2 "like" alleles = A2+A3+H2+I alleles.

637 **Table 3.** Frequency of the *CSN2-CSN3* haplotypes in the Reggiana and Modenese breeds.

CSN2-CSN3 haplotypes[a]	Reggiana (no. = 600)	Modenese (no.= 300)
A1-A	0.075	0.098
A1-B	0.084	0.139
A2-A	0.218	0.095
A2-B	0.407	0.298
B-A	0.091	0.112
B-B	0.048	0.136
C-B	0.037	0.115
Other haplotypes	0.040	0.007

638 [a] Only haplotypes with a minimum frequency of 0.05 in at least one breed are reported.

639

640