

Geographic distribution of haplotype diversity at the bovine casein locus

Oliver C. JANN^a, Eveline M. IBEAGHA-AWEMU^a, Ceyhan ÖZBEYAZ^b,
Pilar ZARAGOZA^c, John L. WILLIAMS^d, Paolo AJMONE-MARSAN^e,
Johannes A. LENSTRA^f, Katy MOAZAMI-GOUDARZI^g, Georg ERHARDT^{a*}

^a Department for Animal Breeding and Genetics, Justus-Liebig University of Giessen, Germany

^b Department of Zootechnics, Faculty of Veterinary Medicine, University of Ankara, Turkey

^c Laboratorio de Genetica Bioquimica y Grupos Sanguineos, Facultad Veterinaria, Zaragoza, Spain

^d Roslin Institute (Edinburgh), Midlothian, Scotland, EH25 9PS, UK

^e Institute of Zootechnics, Catholic University of Sacred Heart, Piacenza, Italy

^f Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

^g INRA-LGBC, Domaine de Vilvert, 78352 Jouy-en-Josas, France

(Received 1 October 2002; accepted 10 September 2003)

Abstract – The genetic diversity of the casein locus in cattle was studied on the basis of haplotype analysis. Consideration of recently described genetic variants of the casein genes which to date have not been the subject of diversity studies, allowed the identification of new haplotypes. Genotyping of 30 cattle breeds from four continents revealed a geographically associated distribution of haplotypes, mainly defined by frequencies of alleles at *CSN1S1* and *CSN3*. The genetic diversity within taurine breeds in Europe was found to decrease significantly from the south to the north and from the east to the west. Such geographic patterns of cattle genetic variation at the casein locus may be a result of the domestication process of modern cattle as well as geographically differentiated natural or artificial selection. The comparison of African *Bos taurus* and *Bos indicus* breeds allowed the identification of several *Bos indicus* specific haplotypes (*CSN1S1**C-*CSN2**A²-*CSN3**A₁/*CSN3**H) that are not found in pure taurine breeds. The occurrence of such haplotypes in southern European breeds also suggests that an introgression of indicine genes into taurine breeds could have contributed to the distribution of the genetic variation observed.

casein / haplotype / *Bos taurus* / *Bos indicus* / phylogeny

1. INTRODUCTION

The bovine casein locus, mapped on *BTA6q31-33* [43], contains four milk protein genes which are closely linked, and in the order α_{s1} -casein (*CSN1S1*),

* Corresponding author: Georg.Erhardt@agrar.uni-giessen.de

β -casein (*CSN2*), α_{s2} -casein (*CSNIS2*), and κ -casein (*CSN3*). The genes are organised in a cluster of approximately 250 kB [13, 41] and share common transcription regulating elements [41]. The locus is considered to influence milk production traits [8, 9, 17, 22, 46] and antibacterial activities of derived peptides [29] may also affect the biological fitness of the offspring. Moreover, casein genes harbour a number of variants with suggested effects concerning traits such as the manufacturing properties of milk [1, 35]. Therefore casein genes could be subject to natural and artificial selection [47]. Polymorphisms in the casein genes allow the determination of casein haplotypes, which can be used for studies concerning quantitative traits [14, 22, 45] or phylogeny [7, 28] since they provide more information than the individual genes [21]. Novel casein variants at *CSN2* [18] and *CSN3* [16, 37, 38] have been described recently, but up to now it has not been clear how these are linked within the haplotypes.

The population structure of cattle (*Bos taurus*) reflects its phylogeny. After the domestication during the Neolithic transition in the Near East, human migrants introduced plants and animals from the domestication centre to Europe [2] and also created the genetic basis of the present cattle breeds [3, 32, 44]. According to the demic expansion model, genetic diversity is expected to be higher at the centre of origin and to decrease with distance [5, 42]. The genetic diversity of cattle measured by biochemical or microsatellite markers follows this pattern with allele frequency gradients following the expansion routes [4, 30, 32]. These studies also suggest a higher genetic diversity of south eastern European breeds compared with those of north western Europe. Additionally, separate domestication and subsequent introgressions of indicine genes into taurine populations in Africa [27] and the Near East [25] produced higher genetic diversity within the hybridisation zones.

The objective of this study was to investigate the diversity of the casein locus in the context of the origin and phylogeny of taurine cattle, including variants, which until today have not been the subject of phylogenetic studies.

2. MATERIALS AND METHODS

2.1. Sampling and DNA-extraction

A total of 1396 blood and DNA samples were collected from 30 cattle breeds of taurine and indicine origin (8–77 unrelated animals per breed) (Tab. I). From most breeds, a minimum of 30 animals were analysed. The exceptions were Slovenian-syrmian (8 samples), a population with an effective population size of less than 10 animals [12], Belgian Blue (mixed purpose,

18 samples), and N'Dama (26 samples). European and Anatolian breeds were selected to represent most of the likely genetic variation of European *Bos taurus* and according to their geographic origin as specified by the longitude (LO) and latitude (LT) of the sampling area (Tab. I). DNA was extracted from leukocytes by standard protocols [33].

2.2. Genotyping of casein polymorphisms

The α_{s1} -casein gene was typed for a *Mae*III polymorphism in the promoter region (*CSN1S1prom*) by PCR-RFLP according to the protocol of [20] and for a polymorphism in exon 17 (*CSN1S1*) with PCR-SSCP which differentiates *CSN1S1*B* from *CSN1S1*C* [19]. Within the α_{s2} -casein gene, the nucleotide exchange differentiating *CSN1S2*A* and *D* was analysed by ACRS [36].

The β -casein (*CSN2*) and κ -casein (*CSN3*) genes were genotyped by PCR-SSCP which differentiates alleles which cannot be identified by isoelectric focusing at the protein level. The techniques used differentiate the *CSN2* alleles *A*¹, *A*², *A*³, *B*, *C*, and *I* [6], and the *CSN3* alleles *A*, *A*_I, *B*, *C*, *E*, *F*, *G*, *H*, and *I* [38], respectively.

2.3. Statistical analyses

2.3.1. Estimation of allele frequencies and test for Hardy-Weinberg equilibrium

Allele frequencies and deviation from the Hardy-Weinberg equilibrium were estimated using GENEPOP V3.1 software [40]. Deviation from the Hardy-Weinberg equilibrium was analysed using a Markov chain method with 1000 iterations.

2.3.2. Effective number of alleles and effective number of haplotypes

For each locus of each breed, the effective number of alleles was calculated using POPGENE V1.31 software [49]. The effective number of haplotypes (N_{hap}) was calculated by the same software, where the effective number of haplotypes is defined as the reciprocal of the expected homozygosity derived from the haplotype frequencies.

Table 1. Collected cattle breeds, number of animals (N), breed origin (Origin), breed group assignment (asm*), longitude (LO), and latitude (LT) of sampling area. The code represents the abbreviations used. Allele frequencies at *CSN1S1prom* (*Prom*), *CSN1S1*, *CSN2*, *CSN1S2*, and *CSN3* are indicated. Indicine and African bovine breeds were used as comparative groups and were not included in the geographic analysis (without LO/LT values).

Code	Breed	N	Origin (asm*)	LO		LT		<i>Prom</i>			<i>CSN1S1</i>			<i>CSN2</i>			<i>CSN1S2</i>			<i>CSN3</i>			
				LO	LT	B	C	B	C	A ¹	A ²	A ³	B	C	I	A	D	A	A _I	B	C	E	H
AA	Aberdeen Angus	43	UK (NC)	-2	58	0.73	0.27	0.78	0.22	0.17	0.82	0.11	0.04	0.01	1.00	0.74	0.21	0.05	0.05	0.05	0.05	0.05	0.05
AB	Anatolian Black	50	Turkey (SE)	35	38	0.70	0.30	0.57	0.43	0.04	0.79	0.11	0.04	0.01	1.00	0.24	0.03	0.41	0.33	0.33	0.33	0.33	0.33
AM	Asturian Mountain	50	Spain (-)	-7	42	0.67	0.33	0.61	0.39	0.20	0.71	0.02	0.04	0.04	0.95	0.05	0.38	0.07	0.07	0.07	0.07	0.07	0.07
AN	Angler	50	Germany (NC)	9	53	1.00	1.00	1.00	1.00	0.48	0.40	0.10	0.02	0.02	1.00	0.59	0.40	0.01	0.01	0.01	0.01	0.01	0.01
AV	Asturian Valley	50	Spain (-)	-7	43	0.92	0.08	0.78	0.22	0.26	0.65	0.053	0.04	0.04	0.95	0.48	0.50	0.01	0.01	0.01	0.01	0.01	0.01
AY	Ayrshire	49	UK (NC)	-5	56	0.97	0.03	1.00	0.00	0.58	0.42	0.02	0.02	0.02	0.93	0.07	0.13	0.32	0.32	0.32	0.32	0.32	0.32
BBb	Belgian Blue, beef	30	Belgium (NC)	5	51	0.83	0.17	0.95	0.05	0.45	0.41	0.02	0.09	0.02	1.00	0.82	0.19	0.05	0.05	0.05	0.05	0.05	0.05
BBm	Belgian Blue, mixed	18	Belgium (NC)	5	51	0.97	0.03	0.97	0.03	0.35	0.41	0.03	0.12	0.09	1.00	0.72	0.28	0.05	0.05	0.05	0.05	0.05	0.05
BF	British Friesian	51	UK (NC)	0	52	0.96	0.04	0.96	0.04	0.64	0.26	0.05	0.05	0.05	1.00	0.83	0.17	0.05	0.05	0.05	0.05	0.05	0.05
BH	Brahman	50	Paraguay (BI)	14	50	0.80	0.20	0.10	0.90	0.18	0.80	0.02	0.02	0.04	1.00	0.29	0.33	0.04	0.04	0.04	0.04	0.04	0.04
BR	Bohemian Red	50	Czech Rep (NC)	14	50	0.83	0.17	0.90	0.10	0.63	0.31	0.02	0.02	0.04	0.98	0.53	0.39	0.08	0.08	0.08	0.08	0.08	0.08
CH	Charolais	56	France (NC)	3	46	0.89	0.11	0.89	0.11	0.18	0.64	0.15	0.03	0.01	1.00	0.42	0.58	0.08	0.08	0.08	0.08	0.08	0.08
CI	Chianina	50	Italy (SE)	12	43	0.70	0.30	0.85	0.15	0.59	0.35	0.01	0.04	0.04	1.00	0.24	0.70	0.06	0.06	0.06	0.06	0.06	0.06
CN	Castia Navarra	50	Spain (SE)	0	42	0.63	0.38	0.83	0.17	0.32	0.65	0.03	0.03	0.03	1.00	0.20	0.02	0.58	0.58	0.58	0.58	0.58	0.58
GB	Banyo Guddali	77	Cameroon (BI)	11	50	0.77	0.23	0.90	0.10	0.53	0.47	0.06	0.06	0.06	1.00	0.24	0.40	0.40	0.40	0.40	0.40	0.40	0.40
GV	German Yellow	36	Germany (NC)	11	50	0.77	0.23	0.90	0.10	0.53	0.47	0.06	0.06	0.06	0.91	0.09	0.60	0.40	0.40	0.40	0.40	0.40	0.40
HE	Hereford	50	UK (NC)	-3	52	0.96	0.04	0.85	0.15	0.49	0.37	0.13	0.13	0.01	1.00	0.77	0.23	0.23	0.23	0.23	0.23	0.23	0.23
IS	Istrian	56	Croatia (SE)	15	45	0.80	0.20	0.77	0.23	0.37	0.51	0.01	0.01	0.01	0.97	0.43	0.46	0.11	0.11	0.11	0.11	0.11	0.11
JE	Jersey	46	UK (NC)	-3	49	0.61	0.39	0.65	0.35	0.17	0.60	0.18	0.18	0.06	1.00	0.29	0.71	0.11	0.11	0.11	0.11	0.11	0.11
MA	Maremmana	52	Italy (SE)	12	42	0.82	0.18	0.83	0.18	0.56	0.36	0.07	0.07	0.01	0.99	0.01	0.84	0.08	0.08	0.08	0.08	0.08	0.08
ME	Menorquina	50	Spain (SE)	4	40	0.73	0.27	0.85	0.15	0.69	0.29	0.02	0.02	0.02	0.93	0.07	0.25	0.10	0.10	0.10	0.10	0.10	0.10
ND	N'Dama	26	Nigeria (SE)	7	45	0.81	0.19	0.96	0.05	0.65	0.31	0.04	0.04	0.04	1.00	0.31	0.69	0.10	0.10	0.10	0.10	0.10	0.10
NE	Nellore	50	Brassil (BI)	7	45	0.58	0.42	1.00	0.00	0.12	0.85	0.04	0.04	0.04	1.00	0.11	0.17	0.11	0.11	0.11	0.11	0.11	0.11
PI	Piemontese	50	Italy (-)	7	45	0.80	0.20	0.85	0.16	0.24	0.42	0.24	0.24	0.05	1.00	0.45	0.47	0.08	0.08	0.08	0.08	0.08	0.08
PR	Polish Red	50	Poland (NC)	17	51	0.67	0.33	0.84	0.16	0.55	0.40	0.06	0.06	0.14	0.99	0.01	0.33	0.02	0.02	0.02	0.02	0.02	0.02
PRI	Pezzata Rossa	51	Italy (NC)	12	46	0.95	0.05	0.89	0.11	0.17	0.51	0.18	0.18	0.01	1.00	0.55	0.45	0.05	0.05	0.05	0.05	0.05	0.05
SG	Santa Gertrudis	50	Paraguay (BI)	17	46	0.78	0.22	0.58	0.42	0.37	0.62	0.02	0.02	0.02	1.00	0.35	0.16	0.43	0.43	0.43	0.43	0.43	0.43
SS	Slavonian-syrmian	8	Croatia (SE)	17	46	0.86	0.14	0.79	0.21	0.29	0.36	0.07	0.29	0.29	1.00	0.63	0.13	0.25	0.25	0.25	0.25	0.25	0.25
TL	Turkish Grey Steppe	50	Turkey (SE)	30	40	0.68	0.32	0.59	0.41	0.24	0.50	0.22	0.22	0.05	1.00	0.41	0.24	0.24	0.24	0.24	0.24	0.24	0.24
TL	Fighting Bull	47	Spain (SE)	-6	37	0.81	0.19	0.92	0.08	0.48	0.50	0.02	0.02	0.02	1.00	0.37	0.63	0.02	0.02	0.02	0.02	0.02	0.02

* NC: northern and central European; SE: southern European and African taurine; BI: *Bos indicus*; -: no clear assignment.

2.3.3. Haplotype frequencies

Haplotype frequencies were estimated under the assumption of allelic association on the basis of all genotype combinations found using EH software [48]. The program uses an iterative Maximum-Likelihood algorithm and compares haplotype frequencies under the assumption of allelic association (calculated value) with those under the assumption of independence (expected value). In addition it gives χ^2 values for this comparison, which were used to calculate *P*-values for the hypothesis that the calculated values differ from the expected values.

2.3.4. Analysis of principal components, correlations and regressions

The analysis of the principal components, correlations and *P*-values, regressions, and variances of allele frequencies, intra-breed diversity and geographic data were performed using SPSS 8.0.0 Software (SPSS Inc., Chicago, USA). For regression analysis of frequency or diversity data with the geographic origin of the breeds, only European and Anatolian *Bos taurus* breeds were used.

3. RESULTS

3.1. Allele frequencies at the casein loci and test for Hardy-Weinberg equilibrium

As indicated in Table I, there were great differences in the occurrence and frequencies of the different alleles at the casein loci between breeds.

Twelve out of 150 tests for Hardy-Weinberg equilibrium (for each gene and breed separately) rejected the null hypothesis of Hardy-Weinberg equilibrium at a 5% probability level. Most of these 12 deviations were found at *CSN3* in Brahman, Banyo Gudali, Istrian, Piemontese, and Pezzata Rossa. Pezzata Rossa, Piemontese, and Nellore also deviated significantly from Hardy-Weinberg equilibrium when all five loci were pooled together.

3.2. Casein haplotype frequencies and linkage disequilibrium

The 19 alleles at the five linked loci were combined in 83 haplotypes. Twenty-one of those were estimated with frequencies over 0.10 in at least one breed (Tab. II). In the 30 breeds analysed, the most frequent haplotype was

*CSNIS1prom*B-CSNIS1*B-CSN2*A²-CSNIS2*A-CSN3*A* with a mean frequency of 0.17, followed by *BBA¹AA* with a mean frequency of 0.15. Neither of these haplotypes were present in Anatolian Black (AB) and Nellore (NE). The related haplotypes *BBA¹AB* and *BBA²AB* were also widely distributed, being present in 26 and 22 breeds respectively. Various haplotypes were limited to specific breed groups *e.g.* *BCA²AA_I* and *BCA²AH* in Brahman (BH) and Nellore (NE). The latter appears as the predominant haplotype in these breeds, but was also found in Banyo Gudali (GB), Istrian (IS), Polish Red (PR), and Turkish Grey Steppe (TG). Also *BCA²AB* occurs at a high frequency only in the hybrid *Bos indicus-Bos taurus* breeds Anatolian Black (AB) and Santa Gertrudis (SG). The *BBCAH*, *BCCA_H*, and *BBA¹AE* haplotypes are completely or almost completely breed-specific, the first two in the Slovenian-syrmian (SS) and the third is a predominant haplotype in the Ayrshire (AY).

The distribution of the casein haplotypes shows a clear dependence on the geographic origin of the breeds (Tab. II, Fig. 1). The haplotypes *BBA²AA* and *BBA¹AA* were found predominantly in north western and central (NC) European cattle breeds; haplotypes *BBA¹AB* and *BBA²AB* are predominant in southern European and African taurine breeds (SE), while in *Bos indicus* breeds (BI) the haplotypes *BCA²AA_I*, *CCA²AA_I*, *BCA²AH*, and *CCA²AH* occur as specific haplotypes or at a high frequency. Such haplotypes were assigned as the basis haplotypes to the corresponding breed groups. In southern Europe many breeds show predominance or a high frequency of further haplotypes which cannot be related to specific breed groups and which may have originated from recent mutations or recombination within haplotypes. In four British (Aberdeen Angus, Ayrshire, Hereford, Jersey) and one African zebu breed (Banyo Gudali), significant ($P < 0.05$) differences between the calculated and expected haplotype frequencies were observed and in two further breeds (Charolais, Santa Gertrudis), marginal differences ($P < 0.1$) were seen.

3.3. Variability within breeds

The effective number of haplotypes (N_{hap}) as a measurement of intra-breed diversity is indicated in Table II. Piemontese (PI) and Turkish Grey Steppe (TG) had the highest N_{hap} values, while the lowest N_{hap} was found in the British Friesian (BF).

The effective number of haplotypes (N_{hap}) was significantly correlated ($P = 0.014$) with the latitude (LT) of the corresponding sampling area. Regression analysis revealed a fit to the linear equation of $N_{\text{hap}} = 13.9823 - 0.1700 \cdot \text{LT}$. A correlation between N_{hap} with the longitude (LO) of breed origin was also

Table II. Frequencies of the predominant haplotypes (frequency > 0.10 at least in one breed) as estimated from casein allele frequencies in 30 breeds. The first letter represents *CSN1S1/prom*, the second *CSN1S1*, the third *CSN2*, the fourth *CSN1S2* and the last *CSN3*. Assignment of haplotypes to breed groups due to predominance or specific occurrence is indicated (*Asm**) and the effective number of haplotypes (N_{hap}), including also rare haplotypes, are indicated.

Haplotype	Asm*	Breed																											Mean							
		AA	AB	AM	AN	AV	AY	BBb	BBm	BF	BH	BR	CH	CI	CN	GB	GV	HE	IS	JE	MA	ME	ND	NE	PI	PR	PRI	SG		SS	TG	TL				
<i>B B A² A A</i>	NC	0.51	0.28	0.17	0.26	0.34	0.33	0.27	0.26	0.15	0.25	0.02	0.13	0.04	0.12	0.23	0.09	0.16	0.06	0.13	0.08	0.14	0.35	0.06	0.33	0.19	0.13	0.17								
<i>B B A¹ A A</i>	NC	0.10	0.33	0.15	0.18	0.14	0.18	0.60	0.02	0.26	0.05	0.15	0.05	0.20	0.40	0.14	0.04	0.12	0.08	0.16	0.25	0.11	0.26	0.25	0.15	0.18	0.15									
<i>B B A¹ A B</i>	SE	0.03	0.09	0.12	0.06	0.05	0.05	0.23	0.08	0.20	0.13	0.17	0.09	0.18	0.28	0.03	0.11	0.10	0.38	0.41	0.46	0.07	0.15	0.06	0.14	0.04	0.35	0.13								
<i>B B A² A B</i>	SE	0.03	0.08	0.26	0.23	0.07	0.04	0.02	0.28	0.18	0.19	0.15	0.04	0.01	0.28	0.04	0.23	0.04	0.11	0.04	0.12	0.07	0.11	0.04	0.12	0.07	0.20	0.09								
<i>C C A² A B</i>		0.13	0.04	0.08			0.07	0.04	0.04					0.23	0.04	0.10			0.08	0.11	0.09	0.02					0.04									
<i>B C A² A H</i>	BI						0.37			0.11		0.07							0.37		0.02					0.12	0.04									
<i>B C A² A A</i>		0.03	0.13	0.15	0.03		0.06	0.05	0.02	0.03	0.06	0.07	0.11	0.02	0.03	0.01	0.05	0.05	0.03	0.01	0.03	0.03			0.03	0.03	0.03									
<i>B B A A B</i>			0.02				0.02	0.06					0.14	0.17	0.07	0.02	0.11	0.03	0.11	0.10					0.16	0.10	0.03									
<i>B C A² A B</i>		0.26	0.09	0.07			0.02	0.02	0.05				0.03	0.03	0.07		0.03	0.01	0.03	0.01	0.16				0.16	0.03	0.03									
<i>C B A² A B</i>		0.05	0.03				0.02	0.01	0.03	0.19	0.08	0.03	0.04	0.02	0.13	0.02	0.03	0.12	0.05	0.08					0.03	0.08	0.03									
<i>B B B A A</i>		0.03	0.07	0.02	0.09	0.14	0.09							0.02	0.02	0.02	0.03	0.12	0.05	0.02					0.02	0.01	0.02									
<i>C B A¹ A B</i>		0.01			0.02	0.15	0.04	0.04	0.17	0.10	0.05	0.04	0.08	0.03	0.04	0.08	0.03	0.04	0.04	0.08	0.03	0.04			0.16	0.02	0.02									
<i>C B A¹ A A</i>			0.01	0.31			0.08						0.16	0.02	0.02	0.03	0.14	0.02	0.03	0.03				0.16	0.02	0.02										
<i>B B A¹ A E</i>			0.03		0.03	0.03	0.06	0.06	0.14	0.02	0.02	0.01	0.03	0.01	0.03	0.01	0.14	0.02	0.03	0.01	0.05			0.04	0.05	0.01	0.01									
<i>C C A² A H</i>	BI		0.03				0.15	0.01	0.02	0.02			0.02			0.04	0.05	0.05	0.05	0.01	0.05			0.04	0.05	0.01	0.01									
<i>B B I A B</i>			0.03		0.06		0.20			0.15	0.01	0.02				0.02			0.04	0.05				0.04	0.05	0.01	0.01									
<i>B C A¹ A A</i>			0.03		0.06		0.20			0.15	0.01	0.02				0.02			0.04	0.05				0.04	0.05	0.01	0.01									
<i>B C A² A A_I</i>	BI						0.03			0.03						0.05			0.05	0.01	0.15			0.05	0.15	0.01	0.01									
<i>C C A² A A_I</i>	BI						0.03			0.03						0.06			0.06	0.15			0.06	0.15	0.01	0.01	0.01									
<i>B B A² A H</i>			0.12				0.03			0.03			0.04			0.06			0.06	0.06			0.06	0.15	0.01	0.01	0.01									
<i>B B C A H</i>							0.12			0.03			0.04			0.06			0.06	0.06			0.06	0.15	0.01	0.01	0.01									
N_{hap}		3.4	8.0	7.4	4.5	6.2	3.9	5.8	5.0	2.3	4.7	6.6	5.6	6.9	8.4	7.4	6.3	4.0	7.6	7.6	4.6	5.0	3.9	5.4	10.4	6.9	5.7	6.8	4.5	9.8	4.5					

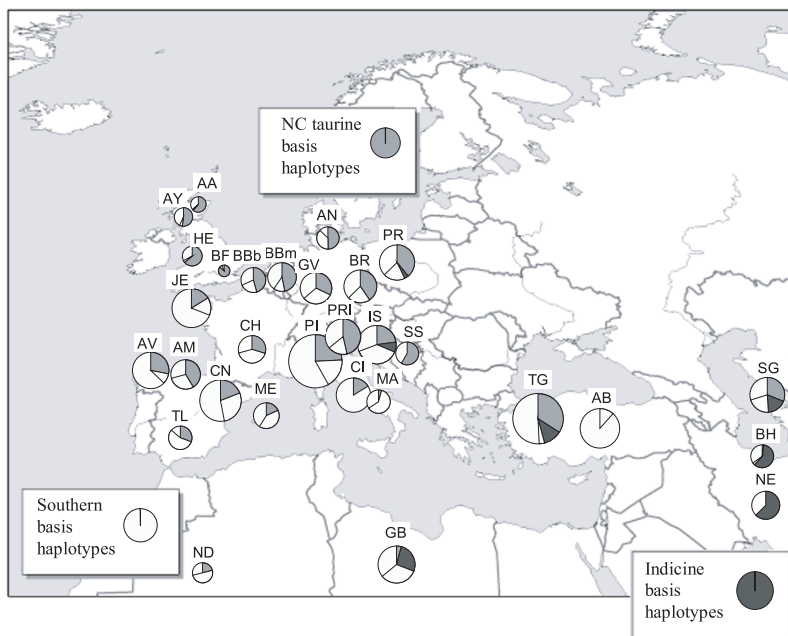


Figure 1. Geographic distribution of predominant casein haplotypes. The dimension of circles is proportional to the intra-breed diversity, measured by the effective number of haplotypes. Haplotypes assigned to north central (NC) European cattle breeds are represented by grey, haplotypes with major frequency in south European and African taurine breeds (SE) by white, haplotypes originated in *Bos indicus* breeds (BI) by black (see Tab. II for specification). Haplotypes based on mutation or recombination events between these ancestor haplotypes are represented by stippled grey.

found to be significant ($P = 0.040$) with a linear regression of $N_{\text{hap}} = 5.5184 + 0.07464 \cdot \text{LO}$.

3.4. Principal components of haplotype distribution

The first principal component (PC1), accounts for 27.84% of the complete variation of haplotype frequencies and the second component (PC2) accounts for 20.02%.

Within the plot of the first two components in the principal component analysis (PCA) (Fig. 2), three extreme positions can be distinguished: the pure *Bos indicus* breeds Brahman (BH) and Nellore (NE) with high values for the first two components, British Friesian (BF) with low values and N'Dama (ND), Maremmana (MA), Menorquina (ME), Fighting Bull (TL), and Chianina (CI), which share major haplotypes in similar frequencies and have high values for

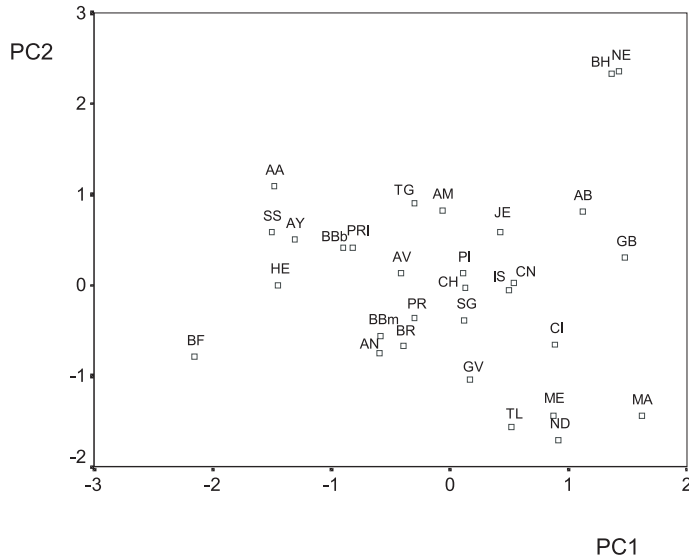


Figure 2. Plotting of the first two principal components (PC1 and PC2) of the casein haplotype frequency distribution in the analysed cattle breeds. PC1 accounts for 27.84%, PC2 for 20.02% of the total variation.

the first and low values for the second component. Two intermediate clusters are formed by the north central European breeds Bohemian Red (BR), Angler (AN), Polish Red (PR), and Belgian Blue mixed purpose (BBm) and by the British breeds Aberdeen Angus (AA), Ayrshire (AY), and Hereford (HE), respectively. The Slovenian-syrmian (SS) appears within the latter group. Anatolian Black (AB) and Banyo Gudali (GB) are positioned between the *Bos indicus* cluster and other breeds. Further breeds are dispersed between these clusters.

The first principal component (PC1) was found to be dependent on the geographic origin of the samples. A highly significant linear regression ($P < 0.00$) was found as $PC1 = 5.2935 - 0.1179 * LT$. A logarithmic equation also showed a highly significant fit ($P < 0.00$) as $PC1 = 20.6947 - 5.4486 * \ln(LT)$. No significant correlation between PC1 and the longitude of breed origin was found. Further components from the PCA were not correlated with the geographic data.

4. DISCUSSION

The DNA-based genotyping allowed those alleles to be identified that have not been included in diversity studies up to now, and which cannot

be separated by protein phenotyping: $CSN3^*A_I$, H , and I , cannot be distinguished from $CSN3^*A$ and likewise $CSN2^*I$ cannot be separated from $CSN2^*A^2$ by electrophoresis of milk samples. Variants in the promoter region of $CSN1S1^{prom}*B$ and C have not been included in previous phylogenetic studies. Up to now $CSN3^*A_I$, H , and I have only been described in *Bos indicus* [38], but in this study they were also found at a lower frequency in taurine breeds. $CSN3^*H$ is present in various southern or eastern European breeds, occurs with a relatively high frequency in Turkish cattle breeds and is predominant in *Bos indicus* breeds. These observations suggest zebu introgressions in southern and eastern European cattle and confirms the results obtained by studies using microsatellites [25] and mitochondria DNA sequences [10].

Haplotype frequencies could not be enumerated by direct gene counting, because multiple heterozygous individuals cannot be resolved when the haplotypic phase is unknown. Therefore the application of iterative methods is necessary to estimate the distribution of haplotypes behind the recognisable genotype combinations found [48]. This approach may result in a bias, especially for rare haplotypes due to a limited sample size, however, this is the only possible approach to estimate haplotype frequencies of unrelated animals. The assumption of Hardy-Weinberg equilibrium for the distribution of haplotypes used by the algorithm in the EH software is problematic in some breeds which were found to deviate from the Hardy-Weinberg equilibrium. This limitation should not affect the final results of the study appreciably because the extent of the deviation was relatively small and restricted to a few breeds.

The observation that casein haplotype frequencies are geographically distributed is in accordance to the findings of former studies based on protein polymorphism [7, 23, 28]. Mahé *et al.* [28] described the predominance of a haplotype on the basis of three casein genes $CSN1S1^*C$ - $CSN2^*A^2$ - $CSN3^*A$ in zebu breeds. However, the electrophoretic methods they used does not allow the discrimination between $CSN3^*H$ and $CSN3^*A_I$ from $CSN3^*A$. Consequently, the occurrence of haplotypes CA^2A_I and CA^2H , which are within BCA^2AA_I , CCA^2AA_I , BCA^2AH , and CCA^2AH , is in agreement with these findings and indicates the introgression of *Bos indicus* in southern and eastern European cattle breeds. These breeds also show an increased gene diversity and haplotypes, which apparently originate from recombination events between taurine and indicine haplotypes *e.g.* BBA^2AH and $BBCAH$. Similarly mt-DNA-analyses [10] and casein haplotype typing [7] indicate the influences of African cattle on the breeds of the Iberian Peninsula, which is confirmed by the predominant appearance of common haplotypes (“southern haplotypes”) in African and in southern European cattle. In contrast to the southern breeds,

Lien *et al.* [23] observed BA^2A ($CSN1S1$ - $CSN2$ - $CSN3$) and CA^2B widely distributed in northern European breeds. The CA^2B was found mainly in autochthonous Nordic breeds and BA^2A was predominant in the highly selected commercial dairy breeds like Finnish Ayrshire or Holstein-Friesian. Other studies [24,34] have found positive effects of $CSN1S1*B$ on the milk yields of dairy cows and [45] describe a positive effect of BA^2A on milk yield in one of four examined grandsire families in the Finnish Ayrshire. Our results show that geographic patterns of haplotype distributions follow frequency trends of alleles at $CSN1S1$ ($CSN1S1*B$: $P \leq 0.010$) and $CSN3$ ($CSN3*A$: $P \leq 0.00$) along the latitude with highest frequencies for the yield related alleles $CSN1S1*B$ and $CSN3*A$ in north western Europe. The $CSN1S1$ alleles seem to be an indicator of artificial selection on the distribution of haplotypes. Most breeds selected for milk yield (*e.g.* Angler, Holstein-Friesian, Ayrshire) originated in north western Europe and are near fixation for $CSN1S1*B$. In comparison central European Highland breeds are mostly dual-purpose breeds. The dual purpose breeds analysed in this study (German Yellow, Pezzata Rossa), and in other studies [11, 15] have a slightly higher frequency of $CSN1S1*C$. Unselected breeds and those used in extensive production systems analysed in this study originate in southern Europe (Casta Navarra, Fighting Bull, Istrian and Slovenian-syrmian) or Turkey (Anatolian Black and Turkish Grey Steppe) and showed much higher frequencies of $CSN1S1*C$. The κ -casein gene is suggested to be strongly effected by selection [47] which may explain the observed deviation of this locus from Hardy-Weinberg equilibrium. The observed frequency trend of $CSN3$ did not seem to be influenced by the way the breed has been selected, indicating that its distribution could be caused by natural rather than by artificial selection.

Medjugorac [31, 32] discussed the frequency gradients at milk protein and further biochemical loci, in the context of Neolithic expansion of cattle following domestication in the Near East. Introgression of migrating domesticated cattle into wild aurochs populations could have caused a “diffuse gene gradient” from the domestication centre.

A similar argument could be used to explain the differences in genetic diversity in Europe, with high diversity in south eastern European breeds in comparison with breeds of northern European origin [26, 30, 44]. This is in agreement with our results of a genetic diversity gradient at the casein genes. It must be mentioned that these gradients are partly caused by an extremely low variability of British Friesian and Aberdeen Angus which might be rather breed than region specific: this could be caused by strong selection pressure on milk production traits in British Friesian and a low effective population size in

Aberdeen Angus [12]. If both these breeds are removed from the analysis, correlations of LT and LO with genetic variation show only marginal significance ($P = 0.076$ for both) and should therefore be assessed with care. It appears possible that the diversity gradients were also formed by differentiated selection or by an introgression of the genetically diverse *Bos indicus*, increasing the variability in a southern European hybridisation zone [7, 10, 25]. Finally, surprising low linkage disequilibrium (LD) was observed between the casein genes, resulting in few breeds with significant differences between the calculated haplotype frequencies and the haplotype frequencies expected under an assumption of allelic independence. This may be a consequence of the small datasets used, because applied methods require large amounts of data to prove a significant difference between the calculated and expected haplotype frequencies [48]. However, it can be concluded that despite the close physical linkage of the genes, recombination seems to be relatively frequent within the casein locus. This confirms the observations already described by [19] and [39].

It can be concluded that the geographic origin of breeds is not independent from selection effects on the haplotype distribution and the genetic variability. In addition the effects of the Neolithic expansion and migration of modern cattle breeds can be found in the correlations between geographic data and the allele frequencies or genetic diversity. An analysis of further breed group-specific markers and a comparative study using neutral markers may elucidate the relative role of selection and cattle migration events on the genetic diversity and distribution of haplotypes at the bovine casein locus.

ACKNOWLEDGEMENTS

The authors thank the German breeder associations; J. Citek, Ceske Budejovice, Czech Republic; I. Medjugorac, Munich, Germany; M. Premzl, Zagreb, Croatia and R. Zieminski, Wroclaw, Poland, for providing samples and H. Brandt, Germany, for a critical review of the used statistical methods. The study was partly financed by the European Community RESGEN-CT98-118 project and by a DAAD scholarship to E.M. Ibeagha.

REFERENCES

- [1] Aleandri R.B., Schneider J.C., Effects of milk protein polymorphisms on milk components and cheese-producing ability, *J. Dairy Sci.* 72 (1990) 241–255.
- [2] Ammerman A., Cavalli-Sforza L.L., *The Neolithic Transition and the Genetics of Populations in Europe*, Princeton University Press, Princeton, 1984.

- [3] Bailey J.F., Richards M.B., Macaulay V.A., Colson I.B., James I.T., Bradley D.G., Hedges R.E., Sykes B.C., Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species, *Proc. R. Soc. Lond. B. Biol. Sci.* 263 (1996) 1467–1473.
- [4] Baker C.M., Manwell C., Chemical classification of cattle, 1. Breed groups, *Anim. Blood Groups Biochem. Genet.* 11 (1980) 127–150.
- [5] Barbujani G., Autocorrelation of gene frequencies under isolation by distance, *Genetics* 117 (1987) 777–782.
- [6] Barroso A., Dunner S., Canon J., A multiplex PCR-SSCP test to genotype bovine beta-casein alleles A^1 , A^2 , A^3 , B , and C , *Anim. Genet.* 30 (1999) 322–323.
- [7] Beja-Pereira A., Erhardt G., Matos C., Gama L., Ferrand N., Evidence for a geographical cline of casein haplotypes in Portuguese cattle breeds, *Anim. Genet.* 33 (2002) 295–300.
- [8] Bovenhuis H., Weller J.I., Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers, *Genetics* 137 (1994) 267–280.
- [9] Bovenhuis H., van Arendonk J.A., Korver S., Associations between milk protein polymorphisms and milk production traits, *J. Dairy Sci.* 75 (1992) 2549–2559.
- [10] Cymbron T., Loftus R.T., Malheiro M.I., Bradley D.G., Mitochondrial sequence variation suggests an African influence in Portuguese cattle, *Proc. R. Soc. Lond. B. Biol. Sci.* 266 (1999) 597–603.
- [11] Erhardt G., Allele frequencies of milk proteins in German cattle breeds demonstration of α_{s2} -casein variants by isoelectric focusing, *Arch. Tierz. Dummerstorf* 36 (1993) 145–152.
- [12] FAO, Domestic animal diversity information system (DAD-IS 2.0), 2002, <http://ded.fao.org>.
- [13] Ferretti L., Leone P., Sgaramella V., Long range restriction analysis of the bovine casein genes, *Nucleic Acids Res.* 18 (1990) 6829–6833.
- [14] Freyer G., Liu Z., Erhardt G., Panicke L., Schätzung von QTL-Effekten für Milchleistungseigenschaften und Untersuchung der genetischen Kopplung mit Caseinmarkern unter Anwendung der Multipointanalyse, *Arch. Tierz. Dummerstorf* 39 (1996) 369–385.
- [15] Grosclaude F., Le polymorphisme génétique des principales lactoprotéines bovines, *INRA Prod. Anim.* 1 (1988) 5–17.
- [16] Ikonen T., Ruottinen O., Erhardt G., Ojala M., Allele frequencies of the major milk proteins in the Finnish Ayrshire and detection of a new kappa-casein variant, *Anim. Genet.* 27 (1996) 179–181.
- [17] Ikonen T., Bovenhuis H., Ojala M., Ruottinen O., Georges M., Associations between casein haplotypes and first lactation milk production traits in Finnish Ayrshire cows, *J. Dairy Sci.* 84 (2001) 507–514.
- [18] Jann O.C., Caroli A., Erhardt G., A new variant in exon VII of bovine β -casein gene (CSN2) and its distribution among European cattle breeds, *J. Anim. Breed. Genet.* 119 (2002) 65–68.

- [19] Jann O.C., Brandt H., Williams J.L., Ajmone-Marsan P., Zaragoza P., Özbeyaz C., Erhardt G., Intragenic haplotypes at the bovine *CSN1S1* locus, Arch. Tierz. Dummerstorf 45 (2002) 13–22.
- [20] Koczan D., Hobom G., Seyfert H.M., Characterization of the bovine α_{s1} -casein gene C-allele, based on a *MaeIII* polymorphism, Anim. Genet. 24 (1993) 74.
- [21] Lien S., Rogne S., Bovine casein haplotypes: number, frequencies and applicability as genetic markers, Anim. Genet. 24 (1993) 373–376.
- [22] Lien S., Gomez-Raya L., Steine T., Fimland E., Rogne S., Associations between casein haplotypes and milk yield traits, J. Dairy Sci. 78 (1995) 2047–2056.
- [23] Lien S., Kantanen J., Olsaker I., Holm L.E., Eythorsdottir E., Sandberg K., Dalsgard B., Adalsteinsson S., Comparison of milk protein allele frequencies in Nordic cattle breeds, Anim. Genet. 30 (1999) 85–91.
- [24] Lin C.Y., McAllister A.J., Ng-Kwai-Hang K.F., Hayes J.F., Effects of milk protein loci on first lactation production in dairy cattle, J. Dairy Sci. 69 (1986) 704–712.
- [25] Loftus R.T., Ertugrul O., Harba A.H., El-Barody M.A., MacHugh D.E., Park S.D., Bradley D.G., A microsatellite survey of cattle from a centre of origin: the Near East, Mol. Ecol. 8 (1999) 2015–2022.
- [26] MacHugh D.E., Bradley D.G., Sharp P.M., Cunningham P., Microsatellite DNA variation within and among European cattle breeds, Proc. R. Soc. Lond. B. Biol. Sci. (1994) 25–31.
- [27] MacHugh D.E., Shriver M.D., Loftus R.T., Cunningham P., Bradley D.G., Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*), Genetics 146 (1997) 1071–1086.
- [28] Mahé M.F., Queval R., Bado A., Zafindrajaona P.S., Grosclaude F., Genetic polymorphism of milk proteins in African *Bos taurus* and *Bos indicus* populations. Characterization of variants α_{s1} -Cn H and κ -Cn J, Genet. Sel. Evol. 31 (1999) 239–253.
- [29] Malkoski M., Dashper S.G., O'Brien-Simpson N.M., Talbo G.H., Macris M., Cross K.J., Reynolds E.C., Kappacin, a novel antibacterial peptide from bovine milk, Antimicrob. Agents Chemother. 45 (2001) 2309–2315.
- [30] Medjugorac A., Molekulargenetische Charakterisierung europäischer und vorderasiatischer Rinderrassen, Thesis, Humboldt-Universität zu Berlin, Berlin, 1999.
- [31] Medjugorac I., Genetischer Polymorphismus in Rinderrassen des Balkan und Phylogenie europäischer Rinder, Thesis, Technische Universität München, München, 1995.
- [32] Medjugorac I., Kustermann W., Lazar P., Russ I., Pirchner F., Marker-derived phylogeny of European cattle supports demic expansion of agriculture, Anim. Genet. 25 Suppl. 1 (1994) 19–27.
- [33] Montgomery G.W., Sise J.A., Extraction of DNA from sheep white blood cells, N.Z. J. Agric. Res. 33 (1990) 437–441.

- [34] Ng-Kwai-Hang K.F., Monardes H.G., Hayes J.F., Association between genetic polymorphism of milk proteins and production traits during three lactations, *J. Dairy Sci.* 73 (1990) 3414–3420.
- [35] Pagnacco G., Caroli A., Effect of casein and β -lactoglobulin genotypes on renneting properties of milk, *J. Dairy Res.* 54 (1987) 479–485.
- [36] Prinzenberg E.M., Entwicklung von Gendiagnoseverfahren für seltene Milchproteinvarianten beim Rind unter Berücksichtigung des Vorkommens bei vom Aussterben bedrohten Rassen, Thesis, Justus-Liebig Universität Gießen, Gießen, 1998.
- [37] Prinzenberg E.M., Hiendleder S., Ikonen T., Erhardt G., Molecular genetic characterization of new bovine kappa-casein alleles CSN3F and CSN3G and genotyping by PCR-RFLP, *Anim. Genet.* 27 (1996) 347–349.
- [38] Prinzenberg E.M., Krause I., Erhardt G., SSCP analysis at the bovine CSN3 locus discriminates six alleles corresponding to known protein variants (*A, B, C, E, F, G*) and three new DNA polymorphisms (*H, I, A₁*), *Anim. Biotechnol.* 10 (1999) 49–62.
- [39] Prinzenberg E.M., Weimann C., Brandt H., Bennewitz J., Kalm E., Schwerin M., Erhardt G., Polymorphism of the bovine *CSN1S1* promoter: linkage mapping, intragenic haplotypes and effects on milk production traits, *J. Dairy Sci.* 86 (2003) 2696–2705.
- [40] Raymond M., Rousset F., GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism, *J. Hered.* 86 (1995) 248–249.
- [41] Rjinkels M., Kooiman P.M., Deboer H.A., Pieper F.R., Organization of the bovine casein gene locus, *Mamm. Genome* 8 (1997) 148–152.
- [42] Slatkin M., Maddison W.P., Detecting isolation by distance using phylogenies of genes, *Genetics* 126 (1990) 249–260.
- [43] Threadgill D.W., Womack J.E., Genomic analysis of the major bovine milk protein genes, *Nucleic Acids Res.* 18 (1990) 6935–6942.
- [44] Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C., Bradley D.G., Genetic evidence for Near-Eastern origins of European cattle, *Nature* 410 (2001) 1088–1091.
- [45] Velmala R., Vilkki J., Elo K., Maki-Tanila A., Casein haplotypes and their association with milk production traits in the Finnish Ayrshire cattle, *Anim. Genet.* 26 (1995) 419–425.
- [46] Velmala R.J., Vilkki H.J., Elo K.T., de Koning D.J., Maki-Tanila A.V., A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle, *Anim. Genet.* 30 (1999) 136–143.
- [47] Ward T.J., Honeycutt R.L., Derr J.N., Nucleotide sequence evolution at the kappa-casein locus: evidence for positive selection within the family Bovidae, *Genetics* 147 (1997) 1863–1872.
- [48] Xie X., Ott J., Testing linkage disequilibrium between a disease gene and marker loci, *Am. J. Hum. Genet.* 53 (1993) 1107.
- [49] Yeh C.Y., Yang R., POPGENE VERSION 1.31, 1999.