Journal of Animal and Feed Sciences, 13, 2004, 51–64

Single nucleotide polimorphism database of candidate genes associated with cow milk protein biosynthesis*

S. Kamiński[1,3], A. Ruść[1] and T. Malewski[2]

[1]University of Warmia and Mazury in Olsztyn, Department of Animal Genetics
Oczapowskiego 5, 10-718 Olsztyn, Poland
[2]Institute of Animal Genetics and Breeding, Polish Academy of Sciences
Jastrzębiec, 05-552 Wólka Kosowska, Poland

(Received 9 April 2003; revised version 16 October 2003; accepted 14 January 2004)

ABSTRACT

A growing number of mutations within milk protein genes and genes associated milk protein biosynthesis are not classified and described in ways which facilitate the design and interpretation of experiments with the use of multiplex PCR or other high throughout screening techniques. The aim of the study was to process and catalog all available information on single nucleotide polymorphisms (SNPs) located within genes directly, indirectly or potentially associated with bovine milk protein biosynthesis. All records were divided into 3 groups of polymorphic sequences: milk protein genes, genes associated with milk protein genes regulation and genes potentially associated with milk protein biosynthesis.

A database was constructed containing 339 SNPs within 49 genes. Among the 339 SNPs, 316 single nucleotide substitutions, 8 deletions, 5 repeats, 7 indels and 3 insertions were identified. All collected SNPs were described in such a way as to enable the automatic downloading of GeneBank records to specialized software and simultaneous design of PCR primers and allele specific probes used in microarray technology.

It is believed that collection of SNPs presented in this study will serve as a reliable resource for studies on the genetic determination of milk protein biosynthesis variation and after wide population screening, also for paternity testing and evolutionary studies in dairy cattle

KEY WORDS: SNP, database, milk protein, biosynthesis, cow

* Supported by the State Committee for Scientific Research, Grant No. 036/P06/2000/05
[3] Corresponding author: e-mail: stachel@uwm.edu.pl

INTRODUCTION

Genetic determinants of protein content in ruminant milk have been the subject of many studies for almost 50 years. They were initiated by the discovery of bovine beta-lactoglobulin polymorphism by Aschaffenburg and Drewry (1955). Over the next 30 years most genetic variants of milk protein have been characterized. They were classified into two groups: caseins (alfa S1 - CSN1S1, alfa S2 - CSN1S2, beta - CSN2 and kappa - CSN3) and whey proteins (beta-lactoglobulin - LGB i alfa-lactalbumin - LALBA) (reviewed by Eigel et al., 1984). For many years, this polymorphism was identified at the protein level by observing different electrophoretic mobilities of milk protein in starch, agarose or polyacrylamid gels. Different populations of dairy cattle were screened to determine the significance of milk protein variants for milk content and yield. Research conducted by numerous groups concluded that polymorphism of kappa-casein and beta-lactoglobulin is strongly associated with the chemical content and technological properties of milk (Jakob and Puhan, 1992; Mao et al., 1992; Walawski et al., 1994). The sequencing of milk protein genes initiated by Gorodetsky et al. (1983) and Steward et al. (1984) enabled the development of methods allowing for the genotyping of bulls (Leveziel et al., 1988; Rando et al., 1988; Lien et al., 1990). Genotyping of CSN3 *locus* was even introduced to breeding programs of A.I. bulls by several commercial companies in the early 1990s.

In the last decade, special attention has been paid to polymorphism within the regulatory sequences of milk protein genes as a possible source of quantitative differences in milk protein biosynthesis (Bleck and Bremel 1993a; Schild et al., 1994; Wagner et al., 1994; Voelker et al., 1997; Kamiński and Zabolewicz, 1998; Kamiński, 1999) and associations between BLG, CSN3, LALBA and milk performance traits were found (Bleck and Bremel 1993b; Ehrman et al., 1997; Kamiński, 2000; Kamiński and Zabolewicz, 2000).

Today, milk protein genes are one of the best studied genes in livestock. Moreover, the number of other SNPs related to milk protein biosynthesis is constantly growing. The most promising ones were found within the following genes: prolactin - PRL (Sasavage et al., 1982; Hart et al., 1993), the signal transducer and activator of transcription - STAT5 (Antoniou et al., 1998; Flisikowski and Zwierzchowski, 2002), growth hormone - GH (Lageziel et al., 1996), growth hormone receptor - GHR (Falaki et al., 1996; Blott et al., 2003) and the ornitine decarboxylase gene (Yao et al., 1998). Anonymous new *loci* associated with milk protein content were proposed in QTL experiments (Georges et al., 1995; Ashwell et al., 1997; Vilkki et al., 1997; Mosig et al., 2001; Boichard et al., 2003). All these reports suggest that milk protein content is a polygenic trait determined by variants located not only within milk protein genes and their promoters but also within other genes involved in milk protein biosynthesis. It is thought that the

simultaneous genotyping of as many informative SNPs as possible will lead to a better understanding of genetic background of milk protein content. Currently the best method for typing SNPs determining complex traits is DNA microarray (review by Syvänen, 2001, and Kamiński, 2002). This technology, however, requires precise DNA sequence information, mainly on the type and location of SNP.

The general aim of this work was to construct a database of all available polymorphic sequences directly, indirectly or potentially associated with cow milk protein biosynthesis.

MATERIAL AND METHODS

SNP definition

Single nucleotide polymorphisms (SNPs) are single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal individuals in some population(s), wherein the least frequent allele has an abundance of 1% or greater (Brookes, 1999). In practice, the term SNP is typically used more loosely and encompasses many different types of subtle sequence variations (including small deletions and insertions) with the frequency of rare allele being less than 1%. To maintain the clarity of this work, the latter SNP definition has been employed.

Database structure

All records of the database were organized in a table (Table 1) consisting of the following columns: position in cytogenetic map (cattle chromosome), *locus* symbol, bovine gene name, sequence description (length, type - DNA or RNA, GenBank acc. no), SNP description (position, location within gene structure, functional significance, and reference).

The mapping position and *locus* symbols were based on the ARK database (www.thearkdb.org) and on Band et al. (2000).

All records were divided into 3 groups of polymorphic sequences: milk protein genes, genes associated with milk protein genes regulation, genes potentially associated with milk protein biosynthesis.

Sources of sequence information

The primary source of records was the GenBank database (NCBI, www.ncbi.nlm.nih.gov) in which 499 records (gene or nucleotide sequence) were found by searching for “Bos AND taurus AND variation”. Records named

"genomic sequence containing highly polymorphic single nucleotide sites" (specific for beef cattle) and "Bos taurus genomic sequence" (unknown function) were rejected from further data mining. Additional resources were also used: bovine mapping genome database (www.thearkdb.org, http://locus.jouy.inra.fr), SNPZoo database (http://snpzoo.de), patent database (http://www.epo.co.at, http://www1.uspto.gov/), database of genes and ESTs expressed in bovine mammary gland (Looft et al., 2001; Malewski and Zwierzchowski, 2002) and human-cattle comparative mapping (Band et al., 2000). Another source of information was the world-wide bibliographic databases (Life Science, CAB, Medline) processed by Reference Manager software (ISI Research Soft, 1999). Column "References" contain references mostly to the documented functional effects of SNP as well as allele frequency data (marked by FD). All of these resources were first previewed and evaluated to ensure they contained at least three elements: GenBank acc. no, position of SNP and minimum length of sequence (250-500 bp DNA or RNA). For some portion of records, individual searching was conducted. For example, if only a variant on protein level was known, the appropriate DNA (RNA) sequence in GenBank database was found and the SNP was marked. Conversely, in some instances, SNPs marked in the GenBank sequence were translated at the protein level or annotated by additional information gained from papers.

RESULTS

A database was constructed containing 339 SNPs within 49 genes (Table 1). Among the 339 SNPs, 316 single nucleotide substitutions, 8 deletions, 5 repeats, 7 indels and 3 insertions were collected. Most SNPs were located in non-coding regions of the genome (mainly within 5' flanking regions) and had no direct known impact on the phenotype of an individual. These SNPs may influence the yield of gene expression and can also be used as markers for unknown adjacent genomic regions.

The most important feature of the SNP database is the precise information of the nature and location of certain SNP. Each SNP is described in the same way, for example, the first SNP in Table 1 (A11115T) means that in the sequence recorded under GenBank acc. no X59856, in position 11115, A is replaced by T. Sometimes the nature of SNP is more complicated and had to be written in a more descriptive way, for example: 2561..2624 (GT)n means that in position between 2561 and 2624 is a GT repeat polymorphism with a different number of GT repeats. This uniform method of SNP description enables the automatic downloading of GeneBank records to specialized software and simultaneous design of PCR primers and allele specific probes used in microarray technology.

Standardization of different data in the same way revealed that numerous papers or GenBank records contain insufficient, conflicting or even error-prone

TABLE 1

SNPs database

Cattle chrom.	Locus symbol	Bovine gene name	Sequence description: length (bp)	type	GenBank Acc. No	SNP description: position	location within gene structure	functional significance	References
I. Milk protein genes									
6q31-33	CSN1S1	casein, alpha s1	22069	DNA	X59856	A 1115 T	5’		FD: Schild and Geldermann, 1996
						C 1255 T	5’		
						A 1311 G	5’		
						C 1335 T	5’		
						C 1357 T	5’	AP2 binding site (comp.)	
						1403 Tdel	5’		
						A 1409 T	5’	MGF binding site (comp.)	
						A 1510 C or T	5’		
						A 1595 G	5’		
						A 1610 G	5’		
						A 1801 C	5’	OCT1 binding site (comp.)	
						1901..1902 TTdel	5’		
						A 2056 G	5’		
						T 1398 C	5’	SP1 and GCF binding site (comp.)	Martin et al., 2002
						C 1390 T	5’		Martin et al., 2002
						T 6561 A	intron 4	exon 4 skipping	Mohr et al., 1994; FD: Erhardt, 1993
						A 1957 G	5’		FD: David and Deutch, 1992; Koczan 1993
						C 2159 T	exon 1	sil	FD: Schild and Geldermann, 1996
						G 9009 A	exon 8	mis, A 53 W	FD: Erhardt, 1993
						A 10527 C	exon 9	mis, K 59 Q	
						C 10549 T	exon 9	mis, S 66 L	
						19312^19313	exon 19	ins 371bp	Mariani et al., 2001; FD: Rando et al., 1998
						A 17807 G	exon 17	mis, E 192 G; protein variants: B or C, respectively	FD: David and Deutch, 1992
6q31-q33	CSN2	beta-casein	10338	DNA	M55158	A 798 G	5’		FD: Schild and Geldermann, 1996
						A 874 T	5’		
						1203 Tdel	5’		
						C 1249 T	5’		
						C 1371 T	5’	SP1 and GCF binding site (comp.)	
						C 1459 T	5’		
						C 8101 A	exon 7	mis, P 67 H; protein variants: A2 or both A1 and B, respectively (If P inhibitor of prolyl endo-peptidase is created, if H casomorphin7 and 11 are created.); Ig E and Ig G binding epitopes	Chatchatee et al., 2001; Gobbetti et al., 2002; Dziuba et al., 1999; Barroso et al., 1999; FD: Damiani et al., 1992
						G 8267 C	exon 7	mis, R 122 S; protein variants:B or A2, respectively	FD: Damiani et al., 1992
						A 8219 C	exon 7	mis, Q 106 H; protein variants:A3 or A2, respectively	FD: Damiani et al., 1992
						G 1614 C	5’	PMF binding site (comp.)	Martin et al., 2002; FD: Schild and Geldermann, 1996
			116	RNA	M64756	C 56 T	CDS		
6q31-33	CSN1S2	alphaS2-casein	21246	DNA	M94327	C 1975 T	5’		FD: Schild and Geldermann, 1996
						A 2109 G	5’	OCT1 binding site (comp.)	
						A 2193 G	5’		
						A 2503 C	5’		
						C 2516 T	exon 1	sil	
						C 2566 T	intron 1	ISGF2 binding site (comp.)	
						G 8879 T	exon 8	exon 8 skipping	FD: Bouniol et. al., 1993
						T 2324 C	promoter	AP1 and CREB binding site (exp.)	Szymanowska, 2001 (dissertation)
						AC 1407/1408 CT	promoter	TR and NFkB binding site (exp.)	
						C 1424 T	promoter		
6q31-q33	CSN3	kappa-casein (QTL marker by Mosig et al., 2001; Lipkin et al., 1998; Tchourzyna et al., 2002)	299	DNA	AF041482	A 130 G	exon 4	mis, S 155 G; protein variants A or E, respectively; Ig E binding epitopes	Chatchatee et al., 2001
			7595	DNA	X14908	C 5309 T	exon 4	mis, T 136 I; protein variants A or B, respectively	FD: Kamiński, 2000
						A 5345 C	exon 4	mis, D 148 A; protein variants A or B, respectively; Ig E binding epitopes	Chatchatee et al., 2001; FD: Kamiński, 2000
						A 5406 G	exon 4	sil	
						A 5413 T	exon 4	sil	
						T 7382 C	exon 5	sil	
						T 7484 C	3’		
						T 7486 C	3’		
						T 7504 C	3’		
						G 7554 A	3’		
						7570^7571 C	3’	ins	
			2945	DNA	AY185364	G 1552 T	intron 2		
			2250	DNA	AF097400	G 106 T	promoter		FD: Medrano, 1999
						A 491 T	promoter		
						C 1113 T	promoter	MGF/STAT5A binding site (comp.)	FD: Schild et al., 1994
						1142^1143 Tdel	promoter		
						G 1169 C	promoter		
						C 1184 T	promoter		
						G 1196 A	promoter		
						C 1257 T	promoter		
						A 1459 G	promoter		
						A 1471 T	promoter		
						1503^1504 ATdel	promoter		
						A 1530 G	promoter		
						G 1626 T	promoter		
						C 1714 T	promoter		
						C 1756 T	promoter	AP-2 binding site (comp.)	FD: Kamiński, 2000
						C 1890 T	promoter		FD: Schild et al., 1994
						A 2081 T	promoter		
			551	DNA	AF105260	C 415 T	exon 4	mis, W 135 I	
					AF121023	T 321 G	exon 4	mis, S 104 A; Ig E binding epitopes	Chatchatee et al., 2001; FD: Prinzenberg et al., 1999
					AF092513	A 461 G	exon 4	sil	
			850	mRNA	X00565	A 540 C	exon 5 (UTR)		Debeljak et al., 2000
						A 576 G	exon 5		
						C 680 T	exon 5		
						C 691 T	exon 5		
						A 692 G	exon 5		
						G 735 A	exon 5		
						C 764 T	exon 5		
						T 765 C	exon 5		
						C 776 T	exon 5		
11q28	LGB	lactoglobulin, beta	7877	DNA	X14710	C 3065 G	exon 2	mis, E 45 Q	
						A 1411 G	5’		FD: Wagner et al., 1994
						G 1485 A			
						G 1514 C	5’		
						C 1542 T	5’		
						C 1575 A	5’	TR binding site (comp.)	
						C 1713 G	5’		
						G 1740 C	5’	AP-2 binding site (exp.)	Lum et al., 1997; FD: Wagner et al., 1994
						T 1751 G	5’		FD: Wagner et al., 1994
						G 1753 A	5’	MAF binding site (comp.)	
						T 1809 C	5’		
						C 1955 T	5’		
						G 1967 C	5’		
						G 2073 C	5’		
						G 2154 A	5’	MAF binding site (comp.)	
						C 2202 T	5’-UTR of exon 1		
						A 2206 G			
						A 3984 G	exon 3	mis, D 64 G, protein variants A or B, respectively	FD: Alexander et al., 1993
						T 5263 C	exon 4	mis, V 118 A; protein variants A or B, respectively	
						C 2171 A	exon 1	alternative transcription start	
						G 5223 T	exon 4	mis, V 105 F; IgE epitopes binding	Heinzmann et al., 1999; Miller et al., 1999
						A 5233 G	exon4	mis, E 108 G; IgE epitopes binding	
						C 5962 T	exon 5	mis, P 126 L; IgE and IgG epitopes binding	
5q21	LALBA	alpha-lactalbumin	2029	DNA	U63109, U63110	A 263 G	5’		FD: Voelker et al., 1997
					U63109	G 261 C	5’		FD: Kazmer et al., 2001
						G 1966 A	5’		FD: Bleck and Bremel, 1993a
II. Genes associated with milk protein genes regulation									
23q23	PRL	prolactin	2209	DNA	X16641	T 1286 C	5’ promoter	AP1 binding site (exp.)	FD: Klauzińska, 2002 (dissertation)
						1341 del TGTG	5’ promoter	GR binding site (exp.)	
						A 1214 C	5’ promoter		
						A 1332 G	5’ promoter		
						A 2084 C	5’ promoter		
						A 2104 T	5’ promoter		
			9388	DNA	AF426315	G 8377 C	exon 4	sil	Sasavage et al., 1982
						A 8398 G	exon 4		
						T 8362 C	exon 4		
						T 9131 A	exon 5	sil	Sasavage et al., 1982; FD: Kamiński et al. (to be published)
						T 9173 A	exon 5		Sasavage et al., 1982
						T 9188 G	exon 5		
						T 9218 C	exon 5		Sasavage et al., 1982; FD: Kamiński et al. (to be published)
			9389	DNA	AF426315	C 3632 T	intron 1		FD: Kamiński et al. (to be published)
						C 3646 T	intron 1		
						A 9061 T	intron 4		
						T 9082 A	intron 4		
						A 9109 G	intron 4		
						C 8307 T	intron 3		
						C 8314 T	intron 3		
						C 8494 T	intron 4		
19q17	STAT5A	signal transducer and activator of transcription 5A	15947	DNA	AJ237937	10872 (TG)n	intron 12	TG rpt	McCracken, 1997
						A 11955 C or G	exon 14	mis, R 593 G	Antoniou et al., 1998
						A 12113 T	exon 15	mis, L 617 M	Antoniou et al., 1998
						A 12069 T	exon 15	mis, Q 602 L	Antoniou et al., 1998
						A 9501 G	intron 9		FD: Kamiński et al. (to be published)
19	STAT5B	signal transducer and activator of transcription 5B	8389	DNA	AJ005638	2561..2624 (GT)n	intron 11	GT rpt	Shillingford and Hennighausen, 2001; Antoniou et al., 1998
20q17	PRLR	prolactin receptor	1114	DNA	AF042780	A 158 C	intron 9		FD: Kamiński et al. (to be published)
III. Genes potentially associated with milk protein biosynthesis									
1q31-36	UMPS	uridine 5’-monophosphate synthase; (UMPS)	1869	mRNA	X65125	C 1247 T	exon 5	mis, R 405 STOP	
1	ITGB2	integrin, beta 2, CD18 subunit (BLAD)	1640	DNA	Y12672	A 1200 G	exon 2	mis, D 128 G	
			2817	mRNA	M81233	C 880 T	CDS	sil	
1q21-22	POU1F1	growth-hormone factor 1 (bGHF-1), (Pit1), pituitary transcription factor 1	1301	DNA	Y15995	G 1256 A	exon 6	sil	
4q32	OBESE	Leptin, [obesity] [murine obesity homoloque]	1292	DNA	AJ236854	T 103 C	intron 1		
			496	DNA	AJ132764	C 140 T	exon 3		
						C 297 T	exon 3		
						T 300 C	exon 3		
						T 312 C	exon 3		
						C 396 T	exon 3		
			1292	DNA	AJ236854	C 126 G	intron 1		
						C 143 T	intron 1		
						T 305 C	exon 2	mis, A in V	FD: Liefers et al., 2002
						C 536 T	intron 2		
						C 538 A	intron 2		
						A 600 G	intron 2		
						G 644 C	intron 2		
						C 726 T	intron 2		
						C 744 T	intron 2		
						T 852 C	intron 2		
						C 860 T	intron 2		
						G 867 A	intron 2		
						G 964 A	intron 2		
						C 1185 T	intron 2		
5	IGF1	insulin-like growth factor-I, [somatomedin C], regulator of folliculogenesis	1942	DNA	AF210383	C 1407 T	5’		Accorsi et al., 2002
			344	DNA	AF210386	27..30 TTTG del	intron 4		Lien et al., 2000
						C 270 T	intron 5		

Cattle chrom.	Locus symbol	Bovine gene name	Sequence description			SNP description			References
			length (bp)	type	GenBank Acc. No	position	location within gene structure	functional significance	
5	COXP1	cytochrome c oxidase, EC=1.9.3.1	571	mRNA	Y14076	A 151 G	5’		
						T 168 C	5’		
						G 198 A	5’		
						G 212 A	5’		
						T 217 C	5’		
						A 223 G	CDS		
						G 251 A	CDS		
						A 436 G	CDS		
						A 489 C	CDS		
						T 495 A	CDS		
						cag 498...500 aca	CDS		
6	MDBK	epithelial cell inflammatory protein-1	1010	DNA	AF061522	A 228 G			
						C 255 T			
						C 288 T			
						C 392 T			
						T 479 C			
						C 603 T			
						A 604 T			
						A 613 T			
						G 667 A			
						G 695 T			
						A 754 T			
6	IL8	interleukin 8	1314	DNA	AF061521	A 136 C			
						G 140 A			
						G 183 A			
						G 1028 A	3’		
						C 1101 T	3’		
						T 1204 C	3’		
						A 1227 G	3’		
8	CTSB	cathepsin B, EC= 3.4.22.1.	718	DNA	AF230197	T 164 G			
						G 207 A			
						G 212 A			
						A 420 N	intron		
						G 289 A			
						A 294 G			
						A 563 G			
10	RNS1	pancreatic ribonuclease, EC= 3.1.27.5.	2705	DNA	X07283	T 1428 C	intron		Sonstegard et al., 2000
11	CPSF	CPSF (cleavage and polyadenylation specificity factor) 73 kDa	2351	mRNA	X95906	T 51 A	5’		
						C 81 T	5’		
						C 822 T	CDS		
						T 831 C	CDS		
						C 882 T	CDS		
						T 966 C	CDS		
						C 1344 T	CDS		
						T 1350 C	CDS		
						G 1425 A	CDS		
11	ODC1	ornithine decarboxylase, EC= 4.1.1.17	9462	DNA	U36394	G 2512 T	intron 1		Yao et al., 1998
						G 7147 A	exon 9	sil	
13	GHRH	growth hormone-releasing hormone receptor (GHRH-R)	456	DNA	AH009425	C 399 G	intron 6		
13q17	PRNP	prion protein	78056	DNA	AJ298878	A 65812 G	exon 3		
						C 66154 T	exon 3		
						C 66877 T	exon 3		
						C 68652 T	exon 3		
						A 69085 G	exon 3		
						12 bp del (GGGGGCC GCGGC) 49729.. 49730 CT	intron 2		Hills et al., 2001
			4244	mRNA	AB001468	G 2182 A			
			6923	DNA	AF163764	GCC 6864.. 6866 AAG	exon 2		
14	KIEL E8	EST158/REV, bovine mammary gland EST	520	mRNA	AI461432	A 293 G			Karall-Albrecht, 1999 (dissertation); Looft et al., 2001
						A 337 C			
14q12	DGAT1	Acyl-CoA:1,2-diacylglycerol O-transferase	14117	DNA	AJ318490	C 3343 G			
						T 3399 G			
						G 7232 G	intron 1		
						A 8567 G	intron 2		
						G 8607 A	intron 2		
						T 9284 C	intron 2		
						A 10147 C	intron 6		
						G 10433 A	exon 8	mis, K 232 A	Kaupe, 2002
						C 10434 A	exon 8	mis, K 232 A	
						C 10486 T	exon 8		
						G 11030 A	intron 12		
						C 11048 T	intron 12		
						C 11521G	exon15		
						T 11993 C	3’		
						A 12005 C	3’		
						T 12036 C	3’		
						A 12056 G	3’		
						G 12136 A	3’		
						G 13309 C	3’		Winter et al., 2002
15	FDX1	adrenodoxin, EC= 1.18.1.2; or ferredoxin--NADP(+) reductase	1751	DNA	D00468	C 469 G	exon 1’		
						G 1329 T	exon 1		
						G 1344 C	exon 1		
						C 1481 A	exon 1		
						C 1543 T	exon 1		
						G 1657 C	intron 1		
16q13	PIGR	polymeric immunoglobulin receptor	3630	mRNA	X81371	A 575 G		mis, I to V	
18	CSNK2A2	casein kinase 2, alpha prime polypeptide, EC=2.7.1.37	535	DNA	AJ133149	C 302 T	intron 6		Lasa-Benito et al., 1996
18	LHB	luteinizing hormone beta polypeptide	1864	DNA	M11506	G 1156 A	exon 2		
						GC 1169.. 1170 CG	exon 2		
18q24	APOE	apolipoprotein E	1154	mRNA	X64839	G 159 C	CDS		
19q26	GH	growth hormone	2856	DNA	M57764	170.. 171 TG	promoter	indel	Hecht and Geldermann, 1996
						C 253 T	promoter		
						C 303 T	promoter	TRE binding site (comp.)	
						C 313 T	promoter	TRE binding site (comp.)	
						C 502 T	promoter		
						C 591 G	promoter		
						C 943 T	intron 1		
			597	DNA	D30713	C 299 G	exon 5	mis, L 127 V, variant C	Vukasinovic et al., 1999; FD: Sorensen et al., 2002; Zhang et al., 1992; Grochowska and Zwierzchowski, 2000
						T 435 C	exon 5		
19	NOS2	inducible nitric oxide synthase, EC= 1.14.13.39	8504	DNA	AF333248	A 490 C	5’		
						T 566 G	5’		
						(T) 17 or 24 1346..1369	5’	rpt	
						C 1393 T	5’		
						G 1645 A	5’		
						A 3509 G	5’		
						A 3800 G	5’		
						A 4008 G	5’		
						T 4981 C	5’		
						(A)16 or 19 4999..5017	5’	rpt	
						C 5480 T	exon 1		
						G 5494 C	exon 1		
						G 5585 C	intron		
						(T)15 or 16 5954..5968	intron	rpt	
20q17	GHR	growth hormone receptor	541	DNA	AF140284	A 257 G	exon 10		
						T 229 C	exon 10		
						G 200 A	exon 10		
						T 76 C	exon 10		
			325	DNA	AJ000484	tgttgaaa 178..186	3’	indel	Moisio et al., 1998
						tatat 190..194	3’	indel	
						C 233 G	3’		
						ttagtggcagtaat 278.. 291	3’	indel	
20	SDHA	succinate dehydrogenase flavoprotein subunit A	1202	DNA	AF139922	A 863 G			
						C 978 G			
20	ITGA2	integrin alpha 2 subunit	3574	mRNA	L25886	G 1414 C	CDS		
						T 1609 G	CDS		
						G 1740 T	CDS		
						G 1764 A	CDS		
						C 2174 A	CDS		
21	SPC18	signal peptidase subunit 18	4054	DNA	AY017294	A 3456 G	intron 2		Ashwell et al., 2001
22q24	LTF	lactoferrin [lactotransferrin]	293	DNA	AH010864	A 156 C	5’		
						C 216 G	5’		
			194			A 118 G	exon 4	mis, V to I	
			202			C 21 T	intron 4		
			217			G 174 T	exon 8		
			159			T 56 C	exon 9		
						T 131 C	exon 9		
			219			C 166 T	exon 15		
			18808	DNA	Z93399	T 5796 A	exon 11	sil	Seyfert et al., 1996
						C 5805 T	exon 11	mis, H 420 Y	
						G 5837 A	exon 11	sil	
23	BUT	butyrophilin (BTN), type I membrane protein of immunoglobulin super-family	8148	DNA	Z93323	C 2281 A	exon 3	mis, P 35 Q	FD: Pareek et al., 2002
						A 1939 C			Pareek et al., 2002
						A 2074 T			
						A 2076 G			
						G 5495 A			
						A 6804 G			
			10176	DNA	AF005497	G 1902 T			
			2691	mRNA	M35551	A 2304 C			Karall-Albrecht, 1999 Karall et al., 1997
						A 1478 G			
24	ACTH	adrenocorticotropic hormone receptor	2909	mRNA	X74501	T 1776 C			
						T 2734 C			
24	BCL2	Bcl-2 protein (bcl2), apoptosis	4190	DNA	AF515848	A 2492 G			
						G 3485 C			
27q18-19	PLAT	tissue plasminogen activator, EC= 3.4.21.68;	387	DNA	AF230195	G 249 A	intron		
27q13	DEFB1	enteric beta-defensin (EBD)	2704	DNA	AF016539	A 1978 G	intron		Sonstegard et al., 2000
27q18-19	FNTA	farnesyltransferase alpha subunit, EC=2.5.1.21	385	DNA	AF230196	G 55 T			
						C 56 T			
						C 91 T			
						T 102 G			
						T 213 C			
						G 250 T			
						C 289 T			
28	PSAP	prosaposin [sphingolipid activator protein 2]	1685	mRNA	AB036791	A 426 G	CDS	mis, H 127 R	
						caggatcag 826 ^ 827	CDS	ins, QDQ are inserted between 260:H and 261:Q	
29	CAPN1	calpain 1	952	DNA	AF465178	C 162 G	intron 18		
						G 477 T	intron 18		
						C 556 G	intron 18		
						C 812 T	intron 18		
N	ACC1	Acetyl-CoA-Carboxylase Alpha	6203	DNA	AJ276223	467..580		indel, transposon=retroposon	
						1305..1840		indel, transposon=retroposon Art2	
						2984..3229		indel, transposon=retroposon	
N	CYHR1	cysteine and histidine-rich cytoplasmic protein	1375	DNA	AH011753	G 667 or 740 A	intron1		Craig, 2002
N	CKM	creatine kinase M chain, EC 2.7.3.2	1146	mRNA	AF120106	C 1131 G	CDS	mis, I to M	

Abbreviations:
Nucleotides: A Adenine, C Cytosine, G Guanine, T Thymine
Position: 7570^7571 points to a site between bases 7570 and 7571, UTR -untranslated region, CDS - coding sequence
SNPs: Del -deletion, ins -insertion, sil -silent, mis -missense, rpt -repeat, indel insertion/deletion
FD - SNP frequency data available in this paper
exp - experimental, comp - computer analysis
Amino Acids: A - Alanine, R - Arginine, D - Aspartic acid, Q - Glutamine, E - Glutamic Acid,
G - Glycine, H - Histidine, I - Isoleucine, L - Leucine, K - Lysine, M - Methionine,
F - Phenylalanine, P - Proline, S - Serine, W - Tryptophan, Y - Thyrosine, V - Valine
Factors: AP Activator Protein, CREB cAMP-Responsive Element-Binding Protein, GCF GC-binding Factor, GR Glucocorticoid-Receptor, ISGF Interferon-Stimulated Gene Factor, MAF Mammary Cell-Activating Factor, MGF Mammary Gland Factor (STAT5), NF-kB Nuclear Factor kappa B, OCT Octamer-Binding Factor, PMF Pregnancy Specific Mammary Nuclear Factor, SP Promoter-Specific Transcription Factor, TR Thyroid-Receptor, TRE Thyroid Hormone Response Element

sequence information. These were first clarified by the comparison to original data published in the paper or by consultation with the authors and were eventually either included or eliminated from the database.

SNPs were also annotated by adding some important information on the function or significance of certain SNP. Most of these annotations indicate the type of mutation: missense or silent and a SNP location in gene structure: intron, exon, 5'- or 3'-flanking regions. Many SNPs have no information in which part of gene structure they are located. Although this location could be theoretically elucidated, it is preferable to sustain the original data. Some SNPs were located within putative (computational) or experimentally confirmed binding sites of transcription factors. Several other SNPs are localized in epitope for immunoglobulin, suggesting their potential significance in immune response, especially in allergy for milk.

Information on allele frequency is also very useful in planning population experiments. If an allele is very rare or specific for uncommon breed it should probably be eliminated because of the low probability of finding a genotype group of animals for associated studies. Therefore, alleles occurring in rare or endangered breeds of cattle were excluded from the database and SNPs were cataloged only for major dairy cattle breeds (e.g., Holstein, Jersey) because of their economic importance. The SNP database shows that, except for the SNPs of major milk protein variants, the population data for most of the SNPs is very poor (references marked by FD; Table 1).

DISCUSSION

The reason for the current vital interest in SNPs is the hope that they could be used as markers to identify genes associated with multifactoral disorders or quantitative trait *loci* (QTLs) (Coronini et al., 2003). It is assumed that the SNP alleles are inherited together with the QTLs over generations because they are physically close to each other. In contrast to microsatellite markers, SNPs are frequently dispersed throughout the genome and therefore can be used for QTL fine mapping. The rationale would be to genotype a collection of SNPs that occur at regular intervals and cover the whole genome to detect genomic regions in which the frequencies of the SNP allele differ between experimental populations. The genome-wide SNP genotyping is theoretically possible for the human genome, for which almost 2 million SNPs are available in the public database (SNP Consortium, www.snp.schl.org). Celera Genomics also offers commercial SNPs databases for human and mouse genomes (www.celera.com). The throughput required for genotyping even some of the thousands of SNPs and the current cost of genotyping makes such projects impractical. A more feasible alternative to random whole-

genome SNP mapping is to use SNP markers in candidate genes which are thought to be associated with certain QTL. This is the only choice for genomes for which no SNP database has been published, but have numerous detected SNPs dispersed in many publicly available sources. In cattle genome, a good candidate for such an approach are SNPs within genes associated directly, indirectly or potentially, with milk protein biosynthesis. To our knowledge, the database presented in this paper is first publicly available SNP database based on dairy cattle genome processed and described to enable automatic and high throughout SNP genotyping.

Database specificity and limitations

It seems the growing number of mutations within bovine milk protein genes and genes associated with their expression have to be ordered and classified to better design and interpret future experiments with the use of high throughout screening techniques. There is an evident lack of uniform information on the topic. In papers, SNPs are described mostly at the protein level as an amino-acid change with or without relevant nucleic acid sequence information. In contrast, in the GenBank database, sequences are not annotated sufficiently (location and type of SNP) or dispersed within different records. Attempts to use these sequences for multi-*loci* genotyping are very limited or even impossible. Therefore, in this paper all available sequence and research information has been gathered to create a well-organized database of SNPs described in the same format.

Dividing all *loci* into three groups helps to better understand their role in milk protein biosynthesis. For the first and second group (milk protein genes, genes associated with milk protein genes regulation), the associations with milk protein content is obvious and documented in numerous papers (review by Jakob and Puhan, 1992, and Martin et al., 2002). The third group (genes potentially associated with milk protein biosynthesis) contains different genes which are believed to be indirectly or potentially associated with milk protein content in milk. For some of them, these associations are experimentally confirmed, but for others they are not. The latter ones were included in the SNP database because they are involved in basic biochemical processes in the mammary gland or play a fundamental role in the functioning of the whole organism.

It is problematic whether all known SNPs within one gene should be included in the database. On one hand, the more SNPs there are within the *locus*, the more choices there are to design effective primers or probes. But on the other hand, too many synonymous SNPs or repeats within one *locus* which are indirectly or only potentially associated with a phenotype seems to be useless and, in this author's opinion, should be ignored. Although SNPs located within the same gene (or within 20-200 kb) are strongly linked and most of them can be omitted, the reduction of a number of SNPs may lead to missing an interesting genetic phenomenon

- interacting phenotyping effects of co-existing variants located within the 5'- and 3'- flanking region of a single gene (Schwerin et al., 2002). Therefore, in the first and second group all published SNP were included. In the third group, however, a kind of pre-selection was made: from *loci* containing more than 10 SNPs (e.g., PRP, NOS2, CPN1) repeat polymorphism and synonymous SNPs located very close to each other were excluded, leaving only those located in exons and in maximum distance.

Because very short stretches of DNA are inconvenient or even useless in primer design, all sequences shorter than 250 bp were excluded from the database.

The database also contains SNPs determining two genetic diseases (BLAD and DUMPS). The carriers of these disorders are obligatory eliminated from reproductive schemes in many countries to avoid losses in health and reproduction.

A separate group of polymorphism associated with milk protein biosynthesis are microsatellites (Mosig et al., 2001; Boichard et al., 2003). These QTL microsatellite markers were excluded from the database for three reasons: 1. the nature of polymorphism is often unclear (the type of repetitive motif, its location and number of repeats), 2. repetitive sequences are difficult to genotype by primer extension reaction – the most often used method in high throughout-put genotype screening on a chip, 3. each QTL microsatellite marker has approximately 10 alleles, which increase the cost of the genotyping.

A way to represent these genomic regions into the chip is sequencing regions located around a QTL microsatellite marker and then comparing these sequences from a population of animals to find new biallelic SNPs. The probability of finding SNP could be lower than in human (1/1250 bp) because of the higher homogeneity of cattle. Such SNPs may substitute QTL microsatellite markers to enable their implementation in high throughout put genotyping.

The only publicly available livestock SNP database (SNPZoo, www.snpzoo.de) is maintained for the development of paternity control. Because this SNP database contains only anonymous SNPs (randomly dispersed in the genome), its records were not included to our SNP database.

The database should be continuously updated by new data, and a potential source of new SNPs are bovine mammary gland expressed sequence tags (ESTs). They can be found by the use of bovine ESTs data and human genomic sequences (Band et al., 2000; Stone et al., 2002) or by *in silico* mapping of DNA sequences to cattle genome (Farber and Medrano, 2003). Cheung and Spielman (2002) suggest that expression profiling with the use of microarray may reveal data on variation of gene expression indicating genes containing causative SNPs. Picoult-Newberg et al. (1999) published a method of SNPs mining from the EST database. Unfortunately, publicly available bovine mammary gland EST databases are dispersed in different resources (GenBank, TIGR: www.tigr.org/tdb/tg/btgi; Looft et al., 2001) and therefore are not suitable to such experiments.

Database applications

The primary application of this SNP database is for designing a chip for the simultaneous genotyping of hundreds of SNPs to reveal the genetic background of milk protein biosynthesis. Protein content in cow milk is one of the most important criterion in bull selection and also in cow milk pricing. This trait has been improved in recent decades and is still the most desirable milk performance trait. It is believed to be possible to find a combination of SNPs to acts as very effective genetic markers in the selection of milk protein content.

In several genes, many mutations were cataloged which create intragenic haplotypes (many SNPs within one gene or very strong linked genes). These kind of haplotypes were first described for bovine LGB and CSN3 genes by Wagner et al. (1994), Ehrmann et al. (1997) and Kamiński (2000). By using the database it is possible to find SNPs located in different, but functionally associated genes. Good examples are SNPs within PRL, RPRL, STAT5 and SNPs identified within STAT5 binding sites located within milk protein promoters. A combination of these intergenic SNPs can give a new insight into relationships between these genes responsible for the major signal transmitting pathway regulating milk protein gene expression.

The collected SNPs represent most of the 29 cattle chromosomes. Eight chromosomes, namely: 2, 3, 7, 9, 12, 17, 25, 26, are not represented in the SNP database. Most of the collected SNPs may play a role as a marker of certain chromosome region, while others should be treated as a causative mutations. All these sequence variants are located within functional genes. Because functional genes are sometimes organized in groups and located close together because of their function, the SNPs described in the catalog may turned to be more efficient genetic markers and may shorten the way to find causative mutations influence milk protein content.

Another possible application of the SNP database is dairy cattle identification and paternity analysis. Compared with most popular DNA marker (microsatellites), SNPs are attractive because they are abundant, genetically stable and amenable to high-throughput automated technology (Vignal, 2002). They are considered as a realistic alternative in livestock identification and kinship analysis (Fries and Durstewitz, 2001; Heaton et al., 2002). Before it, however, a wide population screening must be conducted to validate frequency of SNPs in major dairy cattle breeds.

The SNPs database can also be used for evolutionary studies, evaluation of genetic distances between wild and domestic cattle breeds and the domestification history of bovine species.

Although the SNPs database does not contain all existing variations associated with milk protein content, its originality, current and future applicability make it a valuable resource for designing different experiments, especially with the use of microarray technology.

REFERENCES

Accorsi P.A., Pacioni B., Pezzi C., Forni M., Flint D.J., Seren E., 2002. Role of prolactin, growth hormone and insulin-like growth factor 1 in mammary gland involution in the dairy cow. J. Dairy Sci. 85, 507-513

Alexander J.L., Hayes G., Bawden W., Stewart F.A., Mackinlay G.A., 1993. Complete nucleotide sequence of the bovine beta-lactoglobulin gene. Anim. Biotechnol. 4, 1-10

Antoniou E., Grosz M., Skidmore J.C., 1998. Cloning, analysis and SSCP characterization of the SH2 domain of the bovine gene STAT5. J. Anim. Sci. 76, Suppl. 1, 379 (Abstr.)

Aschaffenburg R., Drewry J., 1955. Occurence of different beta-lactoglobulin in cow's milk. Nature 176, 218-219

Ashwell M.S., Ashwell C.M., Garrett W.M., Bennett G.L., 2001. Isolation, characterization and mapping of the bovine signal peptidase subunit 18 gene. Anim. Genet. 32, 231-233

Ashwell S.M., Rexroad E.C. Jr., Miller H.R., VanRaden M.P., Da Y., 1997. Detection of loci affecting milk production and health traits in an elite US Holstein population using microsatellite markers. Anim. Genet. 28, 216-222

Band M.R., Larson J.H., Rebeiz M., Green C.A., Heyen D.W., Donovan J., Windish R., Steining C., Mahyuddin P., Womack J.E., Lewin H.A., 2000. An ordered comparative map of the cattle and human genomes. Genome Res. 10, 1359-1368

Barroso A., Dunner S., Canon J., 1999. Technical note: use of PCR-SSCP analysis for detection of bovine beta-casein variants A1, A2, A3, and B. J. Anim. Sci. 77, 2629-2632

Bleck T.G., Bremel D.R., 1993a. Sequence and single-base polymorphisms of the bovine alfa-lactalbumin 5'-flanking region. Gene 126, 213-218

Bleck T.G., Bremel D.R., 1993b. Correlation of the alfa-lactalbumin (+15) polymorphism to milk production and milk composition of Holsteins. J. Dairy Sci. 76, 2292-2298

Blott S., Kim J.J., Moisio S., Schmidt-Küntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vilkki J.H., Georges M., Farnir F., Coppieters W., 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163, 253-266

Boichard D., Grohs C., Bourgeois F., Cerquiera F., Faugeras R., Neau A., Rupp R., Amigues Y., Boscher M.Y., Leveziel H., 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genet. Sel. Evol. 35, 77-101

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

Brookes A.J., 1999. The essence of SNPs. Gene 234, 177-186

Chatchatee P., Järvinen K.M., Bardina L., Vila L., Beyer K., Sampson H.A., 2001. Identification of IgE and IgG binding epitopes on β- and κ-casein in cow's milk allergic patients. Clin. Exp. Allergy 31, 1256-1262

Cheung V.G., Spielman R.S., 2002. The genetics of variation in gene expression. Nature Genet. 32, 522-525

Coronini R., de Looze M.A., Puget P., Bley G., Ramani S.V., 2003. Decoding the literature on genetic variation. Nat. Biotechnol. 21, 21-29

Craig J.A., Jansen G.B., Jiang Z.H., 2002. The bovine cysteine- and histidine-rich cytoplasmic gene: EST assembly, genomic organization and association analysis with milk production traits in Canadian Holstein cattle. In: Proceedings of XXVIII International Conference on Animal Genetics, Gettingen (Germany), pp. 184-185

Damiani G., Pilla F., Leone P., Caccio S., 1992. Direct sequencing and bidirectional allele specific polymerase chain reaction of the bovine β-casein B variant. Anim. Genet. 23, 561-566

David A.V., Deutch H.A., 1992. Detection of bovine α_{s1}-casein genomic variants using the allele-specific polymerase chain reaction. Anim. Genet. 23, 425-429

Debeljak M., Susnik S., Marinsek-Logar R., Medrano F.J., Dovc P., 2000. Allelic differences in bovine kappa-CN gene which may regulate gene expression. Eur. J. Physiol. 439, Suppl. 3, R4-6

Dziuba J., Minkiewicz P., Nalecz D., Iwaniak A., 1999. Database of biologically active peptide sequences. Nahrung 43, 190-195

Ehrmann S., Bartenschlager H., Geldermann H., 1997. Quantification of gene effects on single milk proteins in selected groups of dairy cows. J. Anim. Breed. Genet. 114, 121-132

Eigel N.W., Butler E.J., Ernstrom A.C., Farrel M.H. Jr., Harwalkar R.V., Jenness R., Whitney M.R., 1984. Nomenclature of proteins of cow's milk: fifth revision. J. Dairy Sci. 67, 1599-1631

Erhardt G., 1993. A new α_{s1}-casein allele in bovine milk and its occurerence in different breeds. Anim. Genet. 24, 65-66

Falaki M., Gengler N., Sneyers M., Prandi A., Massart S., Formigoni A., Burny A., Portetelle D., Renaville R., 1996. Relationships of polymorphisms for growth hormone and growth hormone receptor genes with milk production traits for Italian Holstein-Friesian bulls. J. Dairy Sci. 79, 1446-1453

Farber C.R., Medrano J.F., 2003. Putative in silico mapping of DNA sequences to livestock genome maps using SSLP flanking sequences. Anim. Genet. 34, 11-18

Flisikowski K., Zwierzchowski L., 2002. Single-strand conformation polymorphism within exon 7 of the bovine STAT5A gene. Anim. Sci. Pap. Rep. 20, 133-137

Fries R., Durstewitz G., 2001. Digital DNA signatures: SNPs for animal tagging. Nature Biotechnol. 19, 508

Georges M., Nielsen D., Mackinnon M., Mishra A., Okimoto R., Pasquino T., Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E., Hoeschele I., 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139, 907-920

Gobbetti M., Stepniak L., De Angelis M., Corsetti A., Di Cagno R., 2002. Latent bioactive peptides in milk proteins: Proteolytic activation and significance in dairy processing. Crit. Rev. Food Sci. Nutr. 42, 223-236

Gorodetsky I.S., Kershulyte D.R., Korobko V.G., 1983. Nucleotide sequence of cDNA of κ-casein macropeptide of *Bos taurus*. Bioorg. Khimia 9, 1693-1695

Grochowska R., Zwierzchowski L., 2000. Polymorphism of growth hormone gene, secretion of somatotropic hormones and milk production in cattle (in Polish). Prz. hod. 68 (8), 15-18

Hart G.L., Bastiaansen J., Dentine M.R., Kirkpatrick B.W., 1993. Detection of a four-allele single-strand conformation polymorphism (SSCP) in the bovine prolactin gene 5` flank. Anim. Genet. 24, 149

Heaton M.P., Harhay G.P., Bennett G.L., Stone R.T., Grosse W.M., Casas C., Keele J.W., Smith T.P., Chitko-McKown C.G., Laegreid W.W., 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. Mamm. Genome 13, 272-281

Hecht C., Geldermann H., 1996. Variants within the 5'-flanking region and the intron I of the bovine growth hormone gene. Anim. Genet. 27, 329-332

Heinzmann A., Blattmann S., Spuergin P., Forster J., Deichmann K.A., 1999. The recognition pattern of sequential B cell epitopes of beta-lactoglobulin does not vary with the clinical manifestations of cow's milk allergy. Int. Arch. Allergy Immunol. 120, 280-286

Hills D., Comincini S., Schlaepfer J., Dolf G., Ferretti L., Williams J.L., 2001. Complete genomic sequence of the bovine prion gene (PRNP) and polymorphism in its promoter region. Anim. Genet. 32, 231-233

Jacob E., Puhan Z., 1992. Technological properties of milk as influenced by genetic polymorphism of milk proteins (a review). Int. Dairy J. 2, 157-178

Kamiński S., 1999. Simultaneous SSCP genotyping of two bovine casein loci. Anim. Sci. Pap. Rep. 17, 29-33

Kamiński S., 2000. Associations between polymorphism within regulatory and coding fragments of bovine kappa-casein gene and milk performance traits. J. Anim. Feed Sci. 9, 435-446

Kamiński S., 2002. DNA microarrays - a methodological breakthrough in genetics. J. Appl. Genet. 43, 123-130

Kamiński S., Zabolewicz T., 1998. SSCP polymorphism within 5' region of bovine beta-lactoglobulin (LGB) gene. J. Appl. Genet. 39, 97-102

Kamiński S., Zabolewicz T., 2000. Association between bovine beta-lactoglobulin polymorphism within coding and regulatory sequences and milk performance traits. J. Appl. Genet. 41, 91-99

Karall-Albrecht C., 1999. Kartierung und Charakterisierung exprimierter Sequenzen der bovinen Milchdrüse. Schriftenreihe des Instituts für Tierzucht und Tierhaltung der Christian-Albrechts-Universität zu Kiel. Dissertation

Karall C., Looft C., Barendse W., Kalm E., 1997. Detection and mapping of a point mutation in the bovine buthyrophilin gene using F-SSCP-analysis. Anim. Genet. 28, 58-71

Kaupe B., Winter A., Ibeagha E., Ozbeyaz C., Williams J.L., Ajmone-Marsan P., Fries R., Erhardt G., 2002. Screening of *Bos indicus* and *Bos taurus* cattle breeds for *DGAT1* polymorphism. In: Proceedings of XXVIII International Conference on Animal Genetics, Gettingen (Germany), p. 149

Kazmer W.G., Zhou P., Troyer J.L., Strausbaugh L., 2001. Haplotype analysis involving a novel polymorphism in bovine alpha-lactalbumin 5`-flanking region. J. Dairy Sci. 84, 1542-1544

Klauzińska M., 2002. Polymorphism within 5' region sof GH, GHRH, prolactin and myostation in cattle (in Polish). Ph.D. Thesis. Institute of Animal Genetics and Breeding, Jastrzębiec (Poland)

Koczan D., Hobom G., Seyfert M.-H., 1993. Characterization of the bovine alfaS1-casein gene C-allele, based on a Mae III polymorphism. Anim. Genet. 24, 74

Lageziel A., Lipkin E., Soller M., 1996. Associations between SSCP haplotypes at the bovine growth hormone gene and milk protein percentage. Genetics 142, 945-951

Lasa-Benito M., Marin O., Meggio F., Pinna L.A., 1996. Golgi apparatus mammary gland casein kinase: monitoring by specific peptide substrate and definition of specifity determinants. FEBS Lett. 382, 149-152

Leveziel H., Metenier L., Mahe F.M., Choplain J., Furet P.J., Pabeuf G., Mercier C.J., Grosclaude F., 1988. Identification of the two common alleles of the bovine kappa-casein locus by the RFLP technique, using the enzyme Hind III. Genet. Sel. Evol. 20, 247-254

Liefers S.C., Te Pas M.F.W., Veerkamp R.F., Delavaud C., Chilliard Y., Van der Lende T., 2002. Association of bovine leptin polymorphisms with circulating leptin concentrations. In: Proceedings of International Conference on Animal Genetics, Gettingen (Germany), p. 172

Lien S., Alestrom P., Steine T., Langsrud T., Vegarud G., Rogne S., 1990. A method for β-lactoglobulin genotyping of cattle. Livest. Prod. Sci. 25, 173-176

Lien S., Karlsen A., Klemetsdal G., Väge D.I., Olsaker I., Klungland H., Aasland M., Heringstad B., Ruane J., Gomez-Raya L., 2000. A primary screen of the bovine genome for quantitative trait loci affecting twinning rate. Mamm. Genome 11, 877-882

Lipkin E., Mosig M.O., Darvasi A., Ezra E., Shalom A., Friedmann A., Soller M., 1998. Quantitative trait locus mapping in dairy cattle by means of selective milk DNA poling using dinucleotide microsatellite markers: analysis of milk protein percentage. Genetics 149, 1557-1567

Looft C., Reinsch N., Karall-Albrecht C., Paul S., Brink M., Thomsen H., Brockmann G., Kühn C., Schwerin M., Kalm E., 2001. A mammary gland EST showing linkage diseqilibrium to a milk production QTL on bovine chromosome 14. Mamm. Genome 12, 646-650

Lum S.L., Dovc P., Medrano F.J., 1997. Polymorphisms of bovine beta-lactoglobulin promoter and differences in the binding affinity of activator protein-2 transcription factor. J. Dairy Sci. 80, 1389-1397

Malewski T., Zwierzchowski L., 2002. Genes expressed in cattle mammary gland - computational analysis of 5' -upstream sequences in search for factors conferring a tissue- and stage-specific transcription. Anim. Sci. Pap. Rep. 20, 5-20

Mao L.I., Buttazzoni G.L., Aleandri R., 1992. Effects of polymorphic milk protein genes on milk yield and composition traits in Holstein cattle. Acta Agr. Scand., Sect. A, Anim. Sci. 42, 1-8

Mariani P., Summer A., Di Gregorio P., Rando A., Fossa E., Pecorari M., 2001. Effects of the $CSN1A^G$ allale of the clotting time of cow milk and on the rheological properties of rennet-curd. J. Dairy Res. 68, 63-70

Martin P., Szymanowska M., Zwierzchowski L., Leroux C., 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. Reprod. Nutr. Develop. 42, 433-459

McCracken J.Y., Molenaar A.J., Snell R.J., Davey H.W., Wilkins R.J., 1997. A polymorphic TG repeat present within the bovine STAT5A gene. Anim. Genet. 28, 459

Medrano F.J., Debeljak M., Islas A., Dovc P., 1999. Sequence of bovine kappa-casein promoter. GenBank Database, Acc. no AF097400

Miller K., Meredith C., Selo I., Wal J.M., 1999. Allergy to bovine beta-lactoglobulin: specificity of immunoglobulin E generated in the Brown Norway rat to tryptic and synthetic peptides. Clin. Exp. Allergy 29, 1696-1704

Mohr U., Koczan D., Linder D., Hobom G., Erhardt G., 1994. A single point mutation results in A allele-specific exon skipping in the bovine alphas1-casein mRNA. Gene 143, 187-192

Moisio S., Elo K., Kantananen J., Vilkki J.H., 1998. Polymorphism within the 3' flanking region of the bovine growth hormone receptor gene. Anim. Genet. 29, 55-57

Mosig M.O., Lipkin E., Khutoreskaya G., Tchourzyna E., Soller M., Friedmann A., 2001. A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. Genetics 157, 1683-1698

Pareek C.S., Pareek R.S., Walawski K., Seyfert M.H., 2002. Association of bovine butyrophilin encoding gene polymorphism with milk fat percentage in Polish Black-and-White and German Holstein-Friesian cattle. In: Proceedings of International Conference on Animal Genetics, Gettingen (Germany), pp. 187-188

Picoult-Newberg L., Ideker T.E., Pohl M.G., Taylor S.L., Donaldson M.A., Nickerson D.A., Boyce-Jacino M., 1999. Mining SNPs from EST databases. Genome Res. 167-174

Prinzenberg M.E., Krause I., Erhardt G., 1999. 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, A_1). Anim. Biotechnol. 10, 49-62

Rando A., Di Gregorio P., Masina P., 1988. Identification of bovine kappa-casein genotypes at the DNA level. Anim. Genet. 19, 51-54

Rando A., Di Gregorio P., Ramunno L., Mariani P., Fiorella A., Senese C., Marletta D., Masina P., 1998. Characterization of the $CSN1A^G$ allele of the bovine alfaS1-casein locus by the insertion of a relict of a long interspersed element. J. Dairy Sci. 81, 1735-1742

Sasavage N.L., Nilson J.H., Horowitz S., Rottman F.M., 1982. Nucleotide sequence of bovine prolactin messenger RNA. J. Biol. Chem. 257, 678-681

Schild A.T., Geldermann H., 1996. Variants within the 5'-flanking regions of bovine milk-protein-encoding genes. III. Genes encoding the Ca-sensitive caseins αs1, αs2 and β. Theor. Appl. Genet. 93, 887-893

Schild A.T., Wagner V., Geldermann H., 1994. Variants within the 5'-flanking regions of bovine milk protein genes: I. kappa-casein-encoding gene. Theor. Appl. Genet. 89, 116-120

Schwerin M., Maak S., Fuerbass R., 2002. Interacting phenotypic effects of co-existing variants within a single gene - cellular stress response is significantly affected by interactions between

promoter and 3'-UTR variants of the porcine *hsp70.2* gene. In: Proceedings of XXVIII International Conference on Animal Genetics, Gettingen (Germany), p. 64

Seyfert M.-H., Klumann U., Steinhoff U.M., Vanselow J., Koczan D., Hobom G., 1996. Variants and biotechnological use of the bovine lactoferrin-encoding gene. In: T W. Hutchens., B. Lönnerdal (Editors). Lactoferrin: Interactions and Biological Functions. Humana Press Inc., Totowa, NJ, pp. 61-79

Shillingford J.M., Hennighausen L., 2001. Experimental mouse genetics - answering fundamental questions about mammary gland biology. Trends Endocrinol. Metab. 12, 402-408

Sonstegard T.S., Garrett W.M., Ashwell M.S., Bennett G.L., Kappes S.M., Van Tassell C.P., 2000. Comparative map alignment of BTA27 and HSA4 and 8 to identify conserved segments of genome containing fat deposition QTL. Mamm. Genome 11, 682-688

Sorensen P., Grochowska R., Holm L., Henryon M., Lovedahl P., 2002. Polymorphism in the bovine growth hormone gene affects endocrine release in dairy calves. J. Dairy. Sci. 85, 1887-1893

Stewart A.F., Willis I.M., Mackinlay G.A., 1984. Nucleotide sequences of bovine s1 and κ-casein cDNAs. Nucl. Acid. Res. 12, 3895-3907

Stone R.T., Grosse W.M., Casas E., Smith T.P.L., Keele J. W., Bennet G.L., 2002. Use of bovine EST data and human genomic sequences to map 100 gene-specific bovine markers. Mamm. Genome 13, 211-215

Syvänen A.C., 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat. Rev., Genetics 2, 930-942

Szymanowska M., 2001. Influence of polymorphism within 5' region on biding of transcription factors and on gene expression in mammary gland (in Polish). Ph.D. Thesis, Institute of Animal Genetics and Breeding, Jastrzębiec (Poland)

Tchourzyna E., Grosman G., Friedmann A., Soller M., Lipkin E., 2002. Detection of multiple QTL on a single chromosome by haplotype analysis with selective DNA pooling. In: Proceedings of 7th World Congress on Genetics Applied Production, Montpellier (France)

Vignal A., 2002. SNP technology. In: Proceedings of XXVIII International Conference on Animal Genetics, Gettingen (Germany), pp. 76-77

Vilkki J.H., de Koning J.D., Elo K., Velmala R., Maki-Tanila A., 1997. Multiple marker mapping of quantitative trait loci of Finnish dairy cattle by regression. J. Dairy Sci. 80, 198-204

Voelker R.G., Bleck T.G., Wheeler B.M., 1997. Single-base polymorphisms within the 5' flanking region of the bovine alfa-lactalbumin gene. J. Dairy Sci. 80, 194-197

Vukasinovic N., Denise S.K., Freeman A.E., 1999. Association of growth hormone loci with milk yield traits in Holstein bulls. J. Dairy Sci. 82, 788-794

Wagner J.S., Schild A.T., Geldermann H., 1994. DNA variants within the 5'-flanking region of milk-protein-encoding genes. II. The β-lactoglobulin-encoding gene. Theor. Appl. Genet. 89, 121-126

Walawski K., Sowinski G., Czarnik U., Zabolewicz T., 1994. Beta-lactoglobulin and kappa-casein polymorphism in relation to production traits and technological properties of milk in the herd of Polish Black-and-White cows. Genet. Pol. 35, 93-108

Winter A., Alzinger A., Fries R., 2002. Physical mapping of a bovine chromosome region with an effect on milk fat content. In: Proceeding of XXVIII Conference on Animal Genetics, Gettingen (Germany), pp. 76-77

Yao J., Aggrey S.E., Zadworny D., Kühnlein U., Hayes J.F., 1998. Restriction fragment length polymorphisms at the ornithine decarboxylase locus associated with milk protein yield in Holsteins. J. Dairy Res. 65, 341-345

Zhang H.M., Brown D.R., Denise S.K., Ax R.L., 1992. Nucleotide sequence determination of a bovine somatotropin allele. Anim. Genet. 23, 578

STRESZCZENIE

Baza polimorficznych sekwencji nukleotydowych w obrębie genów kandydujących związanych z biosyntezą białka mleka krów

Wzrastająca liczba mutacji w obrębie genów białek mleka oraz genów związanych z biosyntezą białek mleka nie została sklasyfikowana i opisana w sposób umożliwiający projektowanie doświadczeń i ich interpretację z użyciem metody multiplex PCR lub wysokowydajnych technik przesiewowych. Celem podjętych badań była „obróbka" i skatalogowanie wszystkich dostępnych informacji na temat polimorficznych sekwencji nukleotydowych (SNPs) zlokalizowanych w genach bezpośrednio, pośrednio lub potencjalnie związanych z biosyntezą białka mleka krowiego. Zgromadzone sekwencje DNA zostały podzielone na 3 grupy: geny białek mleka, geny związane z regulacją ekspresji genów białek mleka i geny potencjalnie związane z biosyntezą białek mleka.

Skonstruowano bazę zawierającą 339 SNPs w obrębie 49 genów. Spośród 339 SNPs, 316 okazało się substytucjami pojedynczych nukleotydów, 8 delecjami, 5 powtórzeniami, 7 mutacjami typu indel i 3 insercjami. Wszystkie polimorfizmy zostały opisane w sposób umożliwiający automatyczne pobieranie sekwencji w formatach dostępnych w GenBank do specjalistycznego oprogramowania służącego do jednoczesnego projektowania wielu starterów PCR oraz allelo-specyficznych sond używanych w technologii mikropłytek (microarray technology).

Przedstawiona w pracy kolekcja SNPs może służyć jako wiarygodne źródło do studiów nad genetycznym podłożem zmienności biosyntezy białek mleka, a po przeprowadzeniu szerokich badań populacyjnych także do kontroli pochodzenia i badań filogenetycznych bydła.