Animal Science Papers and Reports vol. 22 (2004) no. 4, 533-541
Institute of Genetics and Animal Breeding, Jastrzębiec, Poland

Identification of six BAC clones containing the bovine *CSN1S1* and *CSN3* genes: construction of the contig spanning the casein gene cluster*

Alexey Sazanov[1], Tadeusz Malewski[2], Stanisław Kamiński[3], Lech Zwierzchowski[2]

[1] Laboratory of Molecular Genome Organization,
Institute of Farm Animal Genetics and Breeding, Russian Academy of Agricultural Sciences,
Moskovskoye sh. 55A, St. Petersburg-Pushkin 196601, Russia

[2] Department of Molecular Biology,
Polish Academy of Sciences Institute of Genetics and Animal Breeding,
Jastrzębiec, 05-552 Wólka Kosowska, Poland

[3] Department of Animal Genetics, University of Warmia and Mazury in Olsztyn,
10-719 Olsztyn, Michała Oczapowskiego 5, Poland

(Received October 20, 2004; accepted November 25, 2004)

Regulatory sequences of the bovine casein genes are important for getting molecular markers of milk production traits and creation of high-effective vectors for transgenesis. Comparative analysis of these sequences requires continuous bovine DNA sequences covering the 5'-flanking region and coding part of the bovine casein gene cluster. Large insert clones containing *CSN1S1* (αS1-casein) and *CSN3* (κ-casein) genes with average insert size 180 kb were detected by screening of the bovine genomic BAC library CHORI-240 (6.1-fold genome coverage) with radioactively labelled DNA-probes specific for those genes. Six and thirteen BAC clones were identified by hybridization with *CSN1S1*- and *CSN3*-specific probes, respectively. The CH240-130F23 clone was confirmed to contain *CSN1S1* gene by PCR amplification with specific primers; the CH240-046B08, CH240-115O19, CH240-258E02, and CH240-200N21 clones were shown to be specific for *CSN3*. The BAC clone CH240-035O12 was found to contain both casein genes as determined with PCR. These results show that the BAC contig (number of overlapping clones) was established for the whole casein cluster, including *CSN1S1, CSN2, CSN1S2*, and *CSN3* genes. Based on the known insert sizes of

* Supported by NATO grant LST.CLG.979802.

BAC clones in the CHORI-240 library, we conclude that this contig contains up to 270 kb of the 5' flanking region and up to 270 kb of the 3' flanking region. A maximum 5-fold coverage of the contig is assumed in the 3' region of the cluster nearby the *CSN3* gene.

KEY WORDS: BAC library / casein genes / cattle / clones / contig

Gene expression is regulated mostly at a transcriptional level. Regulation has complex character involving several elements: promoters, enhancers, matrix/scaffold attachment regions (MAR/SAR) and *locus* control regions (LCRs). The mammalian milk protein genes are very attractive models to investigate tissue- and stage-specific expression due to their biological features. Epithelium cells of a mammary gland are strongly specialized to synthesize a number of milk proteins during late pregnancy, lactation and early involution. Transcription level of these protein genes is extremely high [Hobbs *et al*. 1982]. Sequences responsible for such transcriptional expression could be used as components of vectors for transgenesis allowing to increase transgene expression sufficiently. Besides that, data on regulatory elements of the bovine casein gene cluster, especially their polymorphic variants with highest influence on milk composition, could be applied for marker-assisted selection (MAS) in cattle [Tanskley 1993, Georges *et al*. 1995, Phillips 1999].

In most cases the levels of expression of transgenes are different, even when the same gene constructs are used [Houdebine 2002]. Beside this, transgene expression (TE) could also be observed in unexpected tissues [Wen *et al*. 1995]. It means that TE depends on the gene integration site and is thought to be affected by the surrounding chromosome environment, what is referred to as the "position effect". LCRs are supposed to be necessary for position-independent and copy number-dependent expression of transgenes [Li *et al*. 1999]. Recent research of upstream and downstream regions of milk protein genes suggested that their expression could also be regulated by LCR. Transgenic rats with both 50-kb upstream and downstream regions of human α-lactalbumin gene showed a position-independent expression [Fujiwara *et al*. 1999]. High expression of bovine αS1 casein gene flanked by 50 kb upstream and downstream regions in transgenic rats also suggests that presence of LCRs skips the position effect [Zuelke *et al*. 1999].

The gridded genomic bacterial artificial chromosome (BAC) libraries allow to get large-insert genomic clones that contain the sequence of interest. Such libraries are known to be applicable for genome analysis using different approaches and being applied for whole genome sequences assembling (http://www.ncbi.nlm.nih.gov). Now, some of them are widely available (www.chori.org) that makes possible both identification and study of the casein cluster regulatory elements. Also, some data on the bovine whole genome shotgun (WGS) sequence (http://www.ncbi.nlm.nih.gov/genomes) could be applied to reveal regulatory elements in casein gene cluster. Unfortunately, the whole bovine genome sequence is still not assembled, that does not allow making comparative analysis based on it.

Here we report an obtaining of continuous bovine DNA-sequences covering the

5'-flanking region and coding part of bovine casein gene cluster *via* hybridization of the CHORI-240 bovine library with *CSN1S1* and *CSN3* specific DNA-probes.

Material and methods

Probes for BAC library screening

Probes were generated by polymerase chain reaction (PCR) on the template of bovine genomic DNA. The PCRs were performed in 100-µl reaction mix comprising primers, each at 10 µM, approximately 200 ng genomic DNA template, and 50 µl of Readymix Redtaq PCR Reaction Mix (SIGMA-ALDRICH). Probe for *CSN1S1* gene was 310 bp long and covered the gene region from -274 to +36 [Koczan *et al*. 1993].

Primer sequences for *CSN1S1* probe were as follows:

forward 5'-GAAGAAGCAGCAAGCTGG-3' and
reverse 5'-TGCATGTTCTCATAATAACC-3'.

The temperature profile for PCR amplification of the fragment of *CSN1S1* gene was as follows: initial denaturation for 3 min at 94°C followed by 32 cycles of 30 s at 94°C; 30 s at 60°C; 30 s at 72°C, and the final extension for 10 min at 72°C.

Primer sequences for *CSN3* probe were as follows [Figiel and Kaminski 1993]:

forward 5'-GAAATCCCTACCATCAATACC-3'
reverse 5'-CCATCTACGCTAGTTTAGATG-3'

The temperature profile for PCR amplification of fragment of *CSN3* gene was as follows: initial denaturation for 3 min at 94°C followed by 35 cycles of 30 s at 94°C; 30 s at 61°C; 30 s at 72°C, and the final extension for 10 min at 72°C. The length of the amplified DNA was 273 bp, and covered part of intron 4 and part of exon 4 of the *CSN3* gene.

The identity of the amplified products was checked electrophoretically in 1.5% agarose gel containing ethidium bromide in TBE buffer. The length of the PCR products was estimated using 100-bp marker (SIGMA-ALDRICH). Amplified product was purified using GenElute PCR DNA Purification Kit (SIGMA-ALDRICH) according to manufacturer's instruction.

Probe labelling

The *CSN1S1* and *CSN3* probes were labelled using random primer DNA labelling kit (SiGMA-ALDRICH) according to manufacturer's instruction. Briefly, DNA was denatured by heating for 10 min in boiling water bath, and then immediately cooled on ice. To 25 ng of denatured probe added was 3 µl of dATP, dGTP and dTTP mixture, 2 µl reaction mixture, 5 µl [a-^{32}P]dCTP, 3000 Ci/mmol (ICN), and water to a total volume of 19 µl. After mixing the reaction components, 1 µl of Klenow enzyme (2 U) was added and the mix was incubated at 37°C for 30 min. The labelled probes were purified from unbound nucleotides using QIAquick nucleotide removal kit (QUIAGEN) following the manufacturers' prescription.

Screening of bovine BAC library

Hybridization of six nylon filters CHORI-240 was performed in rotisserie bottles, with two filters in each bottle. Filters were wetted in minimum volume of "Church Buffer" (1% BSA, 1 mM EDTA, 7% SDS, 0.5 M $NaHPO_4$, pH 7.2). Then, filters were placed in the bottles with 25 ml additional "Church Buffer" and pre-hybridized for one hour at 65°C with rotation in the Bachofer hybridization oven. Next, 10-15 ng of the labelled DNA probes were added to each bottle. Hybridization was performed at 65°C overnight (16-18 hours). The filters were washed twice with 50 ml 2×SSC, 0.1% SDS at 65°C for 7 min, and then washed with 50 ml 0.5×SSC, 0.1% SDS at room temperature for 7 min. This procedure was repeated three times. After that, the filters were rinsed in water and wrapped individually. The filters were exposed individually with Imaging Screen K, 35 × 43 cm (BIO-RAD) for 1 hour and analysed with Molecular Imager FX (BIO-RAD) using Quantity One software package. Positions of the positive clones were interpreted following the protocol supplied with the filters using the transparent overlay as an orientation guide according to the recommendations of the gene library supplier (http://bacpac.chori.org).

Isolation of BAC DNA-clones and PCR-confirmation

Bacteria containing recombinant clones were grown overnight in 5 ml of standard LB medium. BACs were isolated using PhasePrep BAC DNA Kit (SIGMA-ALDRICH) as described by the manufacturer. Briefly, bacterial cells were harvested by centrifugation at 5000 × g for 10 minutes. Supernatant was removed and cells resuspended in 250 μl of resuspension solution. To resuspended cells 250 μl of lysis solution was added and gently mixed. Next, aded was 250 μl of neutralization solution and mixed again. The mix was incubated on ice for 5 minutes and then centrifuged at 16000 × g to pellet the cell debris. Cleared lysate was transferred to new tubes and 450 μl of room-temperature isopropanol was added to precipitate DNA. Nucleic acids were collected by centrifugation at 16000 × g for 20 min, and the pellet was washed with 100 μl of 70% ethanol. DNA was dissolved in 100 μl of TE buffer.

BAC clones were checked for the presence of the *CSN1S1* and *CSN3* gene inserts by PCR reaction using the same primers and amplification conditions as for probe preparation.

Results and discussion

Large insert clones containing *CSN1S1* and *CSN3* with average insert size of 200 kb were detected by screening of the bovine genomic CH240 BAC library (6.1-fold genome coverage) with radioactively labelled DNA-probes specific for the casein genes. Six and thirteen BAC clones were identified by hybridization with *CSN1S1*- and *CSN3*-specific probes, respectively (Tab. 1). Photo 1 presents an example of CHORI-240 screening. The positive clones were ordered from Children's Hospital Oakland Research Institute, Oakland, CA, USA. Using PCR amplification the clone CH240-130F23 was confirmed

Table 1. The library screening and BAC PCR-confirmation results

No	Clone identified	DNA-probe	PCR-confirmation for presence of *CSN1S1*	PCR-confirmation for presence of *CSN3*
1	CH240-046B08	*CSN1S1*	–	+
2	CH240-035O12	*CSN1S1*	+	+
3	CH240-115O19	*CSN1S1*	–	+
4	CH240-130F23	*CSN1S1*	+	–
5	CH240-124F17	*CSN1S1*	–	–
6	CH240-126C05	*CSN1S1*	–	–
7	CH240-258E02	*CSN3*	–	+
8	CH240-200N21	*CSN3*	–	+
9	CH240-254P04	*CSN3*	–	–
10	CH240-260O08	*CSN3*	–	–
11	CH240-274P22	*CSN3*	–	–
12	CH240-270N05	*CSN3*	–	–
13	CH240-250M04	*CSN3*	–	–
14	CH240-262B19	*CSN3*	–	–
15	CH240-200L09	*CSN3*	–	–
16	CH240-228M10	*CSN3*	–	–
17	CH240-233L10	*CSN3*	–	–
18	CH240-215M02	*CSN3*	–	–
19	CH240-205K17	*CSN3*	–	–

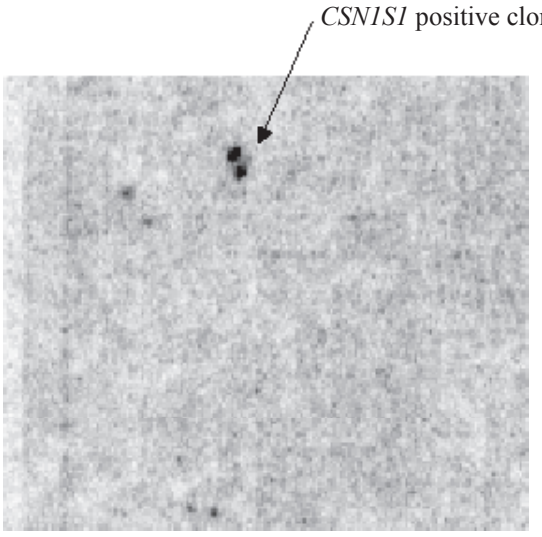

Photo 1. Autoradiogramme of the bovine genomic library CHORI-240 filter fragment after hybridization with radioactively labelled *CSN1S1* DNA-probe.

to contain *CSN1S1* gene, while the CH240-046B08, CH240-115O19, CH240-258E02, and CH240-200N21 clones were shown to be specific for *CSN3* (Photo 2). The BAC clone CH240-035O12, according to PCR-amplification results (Tab. 1), contained genes of both caseins.

Mean insert size of BAC clones from bovine genomic library CHORI-240, segment I is 180 kb, ranging from <100 to 270 kb (http://bacpac.chori.org). In the casein

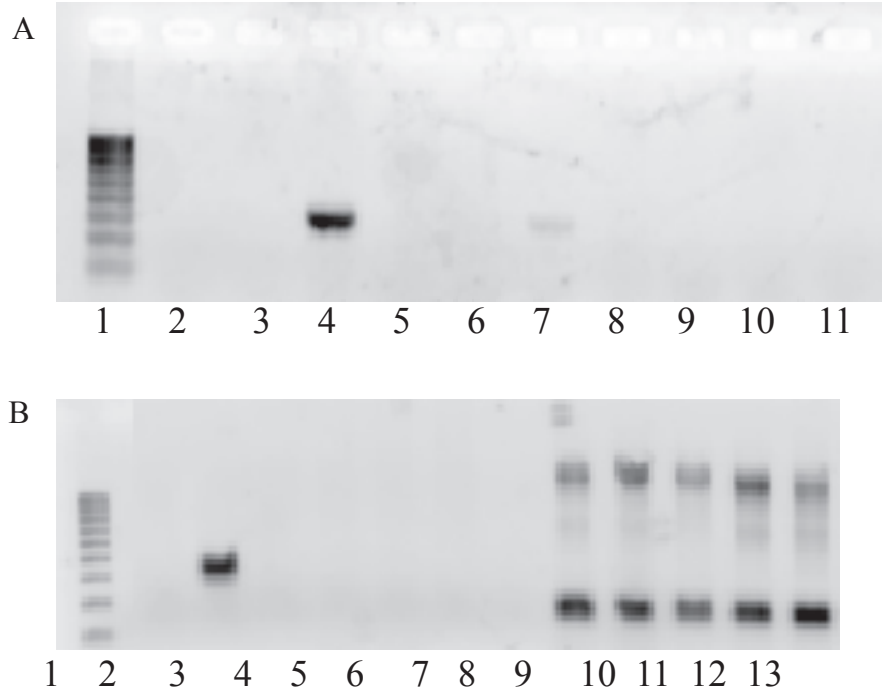

Photo 2. Confirmation with PCR of the identity of BAC clones with *CSN1S1*-specific primers (A) or *CSN3*-specific primers (B).
(A) lane 1 – molecular weight marker (100 bp); 2 – clone CH240-124J17; 3 – CH240-126F23; 4 – CH240-130F23, 5 – CH240-138C05; 6 – CH240-174B06; 7 – CH240-200L09; 8 – CH240-205K17; 9 – CH240-215M02; 10 – CH240-233L10; 11 – CH240-250M04.
(B) lane 1 – molecular weight marker (100 bp); lanes 2 and 9 – clone CH046B08; lanes 3 and 10 – CH240-035012; lanes 4 and 11 – CH240-115019; lanes 5 and 12 – CH240-258E02; lanes 6 and 13 – CH240-200N21; lanes 7 and 8 – H_2O (control).

gene cluster the distance between bovine *CSN1S1* and *CSN3* is about 250 kb [Rijnkels *et al*. 1997]. The *CSN1S1* is situated at the 5'end of the cluster [Rijnkels *et al*. 1997]. Getting the CH240-130F23 clone, which contains *CSN1S1* gene and does not include *CSN3*, means that up to 270 kb of 5' flanking sequence of bovine casein gene cluster was cloned. Search for putative LCRs within the clone CH240-130F23 seems very promising, as such elements are usually located in 5' flanking regions up to -35 kb

upstream of coding genes [Li *et al*. 1999].

Four bovine casein genes are located within a single, multi-gene *locus* of approximately 250 kb in length [Rijnkels *et al*. 1997]. Our data allow to conclude, that clones CH240-130F23, CH240-035O12, CH240-046B08, CH240-115O19, CH240-258E02, and CH240-200N21 are overlapped consequently – 5'-CH240-130F23 – CH240-035O12 – (CH240-046B08, CH240-115O19, CH240-258E02, CH240-200N21) – 3' and form the contig spanning the whole cluster of four bovine casein genes (270 kb upstream from the start codon of *CSN1A1* gene, and up to 270 kb downstream from the terminal codon of *CSN3*) with 5-fold maximum coverage in the 3' region surrounding the *CSN3* gene.

Dairy cattle are the major milk producers for human consumption. Despite the unquestioned economical value of milk, molecular mechanisms regulating expression of its proteins' genes in cattle are still poorly understood. Indeed, only expression and structure of 5'-upstream region of β-casein gene have been systematically studied [Schmidhauser *et al*. 1990,1992, Myers *et al*. 1998]. Some flanking sequences of the αS1- [Platenburg *et al*. 1996], αS2- [Groenen *et al*. [1992] and κ-kasein [Adachi *et al*. 1996] were ascertained, but their importance for the gene expression has not been studied so far.

Moreover, the most investigated regulatory elements of casein genes, such as promoters and enhancers, were shown to be not sufficient for proper expression of genes in transgenic animals [Fujiwara *et al*. 1999]. Polymorphism within coding and 5' flanking (about 1000 bp) regions of the bovine casein *loci* described earlier by Kamiński [2000] and Kamiński and Zabolewicz [2000] does not explain the observed variation in the milk casein biosynthesis satisfactorily.

Isolation of six BAC clones containing *CSN1S1* and *CSN3* genes and establishing of the BAC contig covering the bovine casein gene cluster could provide a new assembled sequence information on its structural organization. These data could also be considered as a tool for comparative analysis of regulatory elements of mammalian casein genes.

REFERENCES

1. ADACHI T., AHN J-Y., YAMAMOTO K., AOKI N., NAKAMURA R., MATSUDA T., 1996 – Characterisation of the bovine κ-casein gene promoter. ***Bioscience, Biotechology, and Biochemistry*** 60(12), 1937-1940.
2. FUJIWARA Y., TAKAHASHI R., MIWA M., KAMEDA M., KODAIRA K., HIRABAYASHI M., SUZUKI T., UEDA M., 1999 – Analysis of control elements for position-independent expression of human α-lactalbumin YAC. ***Molecular Reproduction and Development*** 54, 17-23.
3. GEORGES M., NIELSEN D., MACKINNON M., MISHRA A., OKIMOTO R., PASQUINO A.T., SARGEANT L.S., SOERENSEN A., STEELE M.R., ZHAO X., 1995 – Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. ***Genetics*** 139(2), 907-920.
4. GROENEN M.A.M., DIJKHOF R.J.M., POEL J.J., DIGGELEN R., VERSTEGE E., 1992 – Multiple octamer binding sites in the promoter region of the bovine αS2-casein gene. ***Nucleic Acids Research*** 20(16), 4311-4318.

5. HOBBS A.A., RICHARDS D.A., KESSLER D.J., ROSEN J.M., 1982 – Complex hormonal regulation of rat casein gene expression. ***Journal of Biological Chemistry*** 257 (7), 3598-3605.
6. HOUDEBINE L.M., 2002 – Transgenes and their medical applications. ***Pathological Biology*** (Paris) 50(6), 380-387.
7. KAMIŃSKI S., 2000 – Associations between polymorphism within regulatory and coding fragments of bovine kappa-casein gene and milk performance traits. ***Journal of Animal and Food Sciences*** 9, 435-446.
8. KAMIŃSKI S., FIGIEL I., 1993 – Kappa-casein genotyping of Polish Black-and-White × Holstein Friesian bulls by polymerase chain reaction. ***Genetica Polonica*** 34, 65-72.
9. KAMIŃSKI S., ZABOLEWICZ T., 2000 – Association between bovine beta-lactoglobulin polymorphism within coding and regulatory sequences and milk performance traits. ***Journal of Applied Genetics*** 41, 91-99.
10. KOCZAN D., HOBORN G., SEYFERT H-M., 1993 – Characterisation of the bovine αs1-casein gene C-allele, based on a Mae III polymorphism. ***Animal Genetics*** 24, 74
11. LI Q., HARJU S., PETERSON K.R., 1999 – Locus control regions: coming of age at a decade plus. ***Trends in Genetics*** 15(10), 403-408.
12. MYERS C.A., SCHMIDHAUSER CH., MELLENTIN-MICHELOTTI J., FRAGOSO G., ROSKELLEY C.D., CASPERSON G., MOSSI R., PUJUGUET P., HAGER G., BISSELL M.J., 1998 – Characterisation of BCE-1, a transcriptional enhancer regulated by prolactin and extracellular matrix and modulated by the state of histone acetylation. ***Molecular and Cellular Biology*** 18(4), 2184-2195.
13. PHILLIPS P.C., 1999 – From complex traits to complex alleles. ***Trends in Genetics*** 15 (1), 6-9.
14. PLATENBURG G.J., VOLLENBREGT E.J., RATZAS C.N.K., KOOTWIJK P.A., DE BOER H.A., STRIJKER R., 1996 – Mammary gland-specific hypomethylation of Hpa II sites flanking the bovine αS1-casein gene. ***Transgenic Research*** 5, 421-431.
15. RIJNKELS M., KOOIMAN P.M., DE BOER H.A., PIEPER F.R., 1997 – Organization of the bovine casein gene locus. ***Mammalian Genome*** 8 (2), 148-152.
16. SCHMIDHAUSER CH., BISSELL M.J., MYERS C.A., CASPERSON G.F., 1990 – Extracellular matrix and hormones transcriptionally regulate bovine β-casein 5' sequences in stably transfected mouse mammary cells. ***Proceedings of the National Academy of Sciences, USA*** 87, 9118-9122.
17. SCHMIDHAUSER CH., CASPERSON G.F., MYERS C.A., SANZO K.T., BOLTEN S., BISSELL M.J., 1992 – A novel transcriptional enhancer is involved in the prolactin- and extracellular matrix-dependent regulation of β-casein gene expression. ***Molecular Biology of the Cell*** 3, 699-709.
18. TANKSLEY S.D., 1993 – Mapping polygenes. ***Annual Reviews in Genetics*** 207, 205-233.
19. WEN J., KAWAMATA Y., TOJO H., TANAKA S., TACHI C., 1995 – Expression of whey acidic protein (WAP) genes in tissues other than the mammary gland in normal and transgenic mice expressing mWAP/hGH fusion gene. ***Molecular Reproduction and Development*** 41, 399-406.
20. ZUELKE K., WALKER C., PRICE K., CARFI J., CHRISTOPOULOS H., MORRISSEY W., HARRIS S., TRAVERS A., WEBSTER T., 1999 – Expression of a 100 KB bacterial artificial chromosome (BAC) genomic clone of bovine αs1 casein in transgenic mice. ***Theriogenology*** 51 (1), 430.

Alexey Sazanov, Tadeusz Malewski,
Stanisław Kamiński, Lech Zwierzchowski

Identyfikacja sześciu klonów BAC z genami *CSN1A1* i *CSN3* bydła: konstrukcja ciągłej sekwencji, obejmującej rejon genów kazein

S t r e s z c z e n i e

Sekwencje regulacyjne genów kazein można wykorzystywać jako markery genetyczne użytkowości mlecznej bydła oraz stosować w konstruowaniu wydajnych wektorów dla uzyskiwania zwierząt transgenicznych. Do analizy porównawczej genów kazein (np. różnych gatunków ssaków) potrzebna jest znajomość ciągłej sekwencji rejonu 5'-flankującego oraz części kodującej tych genów. Bibliotekę klonów BAC CHORI-240 (6,1-krotne pokrycie genomu) hybrydyzowano z radioaktywnie znakowanymi sondami do genów *CSN1S1* i *CSN3*, co pozwoliło zidentyfikować klony zawierające geny *CSN1S1* i *CSN3* o średniej długości wklonowanego fragmentu wynoszącej 180000 par zasad (pz). Zidentyfikowano sześć klonów BAC z genem *CSN1S1* i trzynaście z genem *CSN3*. Obecność sekwencji genu *CSN1S1* w klonie CH240-130F23 oraz genu *CSN3* w klonach CH240-046B08, CH240-115019, CH240-258E02 i CH240-200N21 potwierdzono analizą PCR, wykorzystując primery specyficzne dla tych genów. Analiza PCR wykazała obecność sekwencji obu genów kazein w klonie BAC CH240-035012. Uzyskane wyniki wskazują, że znaleziono klony, których sekwencje pokrywają cały *locus* kazeinowy z genami *CSN1S1, CSN2, CSN1S2* i *CSN3*. Na podstawie długości fragmentów DNA, znajdujących się w bibliotece klonów BAC CHORI-240, można wnioskować, że znalezione klony reprezentują do 270 tysięcy pz flankującego rejonu 5' oraz do 270 tysięcy pz rejonu flankującego 3'. Maksymalne – pięciokrotne – pokrycie sekwencji występuje prawdopodobnie w 3'-flankującym rejonie grupy genów kazein.