

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CÂMPUS DE JABOTICABAL**

**ANÁLISE COMPARATIVA DE GENES DAS CASEÍNAS DE
BÚFALO**

Bruna Cristina Machado Naressi

Bióloga

2015

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CÂMPUS DE JABOTICABAL**

**ANÁLISE COMPARATIVA DE GENES DAS CASEÍNAS DE
BÚFALO**

Bruna Cristina Machado Naressi

Orientadora: Profa. Dra. Maria Elisabete Jorge Amaral

Coorientadora: Dra. Nedenia Bonvino Stafuzza

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias - Unesp, Câmpus de Jaboticabal, para a obtenção do título de Mestre em Genética e Melhoramento Animal.

2015

N226a Naressi, Bruna Cristina Machado
Análise comparativa de genes das caseínas de búfalo / Bruna
Cristina Machado Naressi. – – Jaboticabal, 2015
x, 58 p. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual Paulista,
Faculdade de Ciências Agrárias e Veterinárias, 2015
Orientadora: Maria Elisabete Jorge Amaral
Banca examinadora: Paola Jocelan Scarin Provazzi, Rodrigo
Pelicioni Savegnago
Bibliografia

1. Biblioteca genômica. 2. *Bubalus bubalis*. 3. Polimorfismos. 4.
Pirosequenciamento. 5. Sequências repetitivas I. Título. II.
Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.293.2:636.082

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação –
Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.



UNIVERSIDADE ESTADUAL PAULISTA

CAMPUS DE JABOTICABAL

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS DE JABOTICABAL

CERTIFICADO DE APROVAÇÃO

TÍTULO: ANÁLISE COMPARATIVA DE GENES DAS CASEÍNAS DE BÚFALO

AUTORA: BRUNA CRISTINA MACHADO NARESSI

ORIENTADORA: Profa. Dra. MARIA ELISABETE JORGE AMARAL

CO-ORIENTADORA: Profa. Dra. NEDENIA BONVINO STAFUZZA

Aprovada como parte das exigências para obtenção do Título de MESTRE EM GENÉTICA E MELHORAMENTO ANIMAL , pela Comissão Examinadora:

Profa. Dra. NEDENIA BONVINO STAFUZZA

Pós-doutoranda / Departamento de Ciências Exatas / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. RODRIGO PELICIONI SAVEGNAGO

Pós-doutorando / Departamento de Ciências Exatas / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Profa. Dra. PAOLA JOCELAN SCARIN PROVAZZI

Departamento de Biologia / Instituto de Biociências Letras e Ciências Exatas de São José do Rio Preto

Data da realização: 02 de fevereiro de 2015.

DADOS CURRICULARES DA AUTORA

BRUNA CRISTINA MACHADO NARESSI - nascida em 25 de outubro de 1989, na cidade de Riolândia-SP, Brasil, filha de Vera Aparecida de Souza Machado e Francisco Valdeci Florian Naressi. Iniciou em fevereiro de 2008 o curso de Ciências Biológicas no Instituto de Biociências, Letras e Ciências Exatas - IBILCE/UNESP - Câmpus de São José do Rio Preto/SP, onde em janeiro de 2012 obteve o título de Licenciada em Ciências Biológicas. Em 2012, foi bolsista Treinamento Técnico - TT3 do projeto “Região MHC do búfalo - organização molecular utilizando ferramentas genômicas de última geração”, financiado pela FAPESP e desenvolvido sob a coordenação da Profa. Dra. M. Elisabete J. Amaral. Ingressou no Programa de pós-graduação em Genética e Melhoramento Animal da Faculdade de Ciências Agrárias e Veterinárias (FCAV-UNESP), Câmpus de Jaboticabal/SP como aluna de mestrado e bolsista CAPES, em março de 2013. Desenvolveu seu projeto de mestrado, intitulado “Análise comparativa de genes das caseínas entre búfalo e bovino”, sob a orientação da Profa. Dra. M. Elisabete J. Amaral e a coorientação da Dra. Nedenia Bonvino Stafuzza.

“Na vida, não vale tanto o que temos, nem tanto importa o que somos. Vale o que realizamos com aquilo que possuímos e, acima de tudo, importa o que fazemos de nós.”

Chico Xavier

Dedico este trabalho à minha FAMÍLIA!

AGRADECIMENTOS

Ao meu amado Pai de Bondade e Misericórdia Infinita pela oportunidade de aprimoramento intelectual e crescimento em tantos outros aspectos.

À minha orientadora Profa. Dra. Elisabete Amaral pela paciência e orientação ao longo do desenvolvimento deste trabalho. Obrigada professora pelas oportunidades que foram imprescindíveis para o meu amadurecimento profissional e pessoal.

À minha coorientadora Dra. Nedenia Bonvino Stafuzza pelos ensinamentos, pela amizade, dedicação, incentivo e apoio incondicional ao longo desta jornada. Meu imenso carinho, admiração e gratidão.

À CAPES pela concessão da bolsa de estudos, tornando possível o desenvolvimento das atividades relacionadas ao meu mestrado.

À FAPESP pelo auxílio concedido a pesquisa (Processo: 2011/11889-3) possibilitando a realização dos experimentos vinculados ao meu projeto de pesquisa.

Aos professores Dr. Danísio Prado Munari e Dra. Vera Fernanda Martins Hossepian de Lima pelas contribuições prestadas enquanto membros da banca examinadora do projeto de mestrado.

Aos membros da banca de qualificação, a professora Dra. Vera Fernanda Martins Hossepian de Lima e o Dr. Rodrigo Pelicioni Savegnago pelas valiosas sugestões.

À Dra. Paola Jocelan Scarin Provazzi e o Dr. Rodrigo Pelicioni Savegnago por terem aceitado de prontidão o convite para participar da comissão examinadora da presente dissertação de mestrado. Meu agradecimento mais do que especial à Dra. Paola por ter aceitado meu convite de última hora e por não ter medido esforços para participar da minha banca. Agradeço também à Dra. Paola e ao Dr. Rodrigo as correções e valiosas sugestões prestadas para o enriquecimento do meu trabalho.

A todos os funcionários da Seção de Pós-graduação da FCAV-UNESP pela atenção e por sempre se mostrarem dispostos a resolver as questões burocráticas do mestrado à distância.

Ao coordenador do Programa de Genética e Melhoramento Animal, Prof. Dr. Danísio Prado Munari por toda atenção e assistência prestadas.

Aos meus colegas de laboratório Daniela Bizari, Mariana Victoretti e Lucas Olivatto pelas trocas de experiências e pelos momentos amizade e diversão. À Mariana Maciel Borges, um agradecimento especial pela amizade, por sempre me ajudar com problemas cotidianos, pelas risadas e brincadeiras.

Às minhas ex-companheiras de apartamento Tayane Machado e Aline Monteiro pela amizade e companheirismo ao longo da minha estadia em Rio Preto. Obrigada meninas pelos momentos de partilha e conversas aleatórias que tornaram os meus dias ainda mais felizes.

A todos os amigos que fiz em Jaboticabal por terem me recebido com hospitalidade e carinho no período que morei nesta cidade para cursar as disciplinas do mestrado. Um agradecimento especial à Jaqueline Rosa, Emília Barreto e Tatiane Chud por terem me acolhido e por sempre terem me ajudado nos momentos que mais precisei. Sou muito grata a tudo que fizeram por mim, meninas!

Às minhas eternas amigas Cleiliane S. Peixoto e Clarinda Santos pela importância de cada uma em minha vida. Sou muito grata a Deus por ter encontrado vocês ao longo da minha caminhada!

Ao meu companheiro e amigo, William Pfeifer, por todo incentivo e por acreditar no meu potencial. Obrigada por me ensinar a ver o mundo de uma perspectiva diferente e por sempre estar do meu lado nos momentos que mais preciso!

À minha mãe Vera, exemplo de luta e determinação, pelo amor, incentivo e apoio incondicional. Obrigada mãe por sempre estar do meu lado e por ajudar na realização dos meus sonhos.

Ao meu pai Francisco, que mesmo distante e com o seu jeito simplório de ser, sempre torce pelo meu sucesso.

Às minhas irmãs Beatriz e Betânia pelo carinho, incentivo e por acreditarem no meu potencial.

A todas as pessoas que, mesmo não mencionadas, contribuíram de alguma forma para a realização deste trabalho. Obrigada!

SUMÁRIO

	Página
RESUMO.....	xii
ABSTRACT.....	xiii
1. INTRODUÇÃO.....	1
2. REVISÃO DA LITERATURA.....	4
3. OBJETIVOS.....	11
4. MATERIAL E MÉTODOS.....	12
4.1 Seleção dos clones positivos para genes das caseínas.....	12
4.2 Análises das sequências de DNA obtidas por pirosequenciamento com as sequências correspondentes em bovino.....	17
4.3 Predição de genes.....	18
4.4 Identificação de sequências de DNA conservadas entre búfalo e boi.....	19
4.5 Identificação de sequências repetitivas.....	20
4.6 Análise comparativa com outros bovídeos.....	21
4.7 Comparação da sequência de DNA dos genes da raça Murrah com aquelas correspondentes na raça Mediterrâneo.....	21
5. RESULTADOS E DISCUSSÃO.....	22
5.1 Seleção dos clones positivos para genes das caseínas.....	22
5.2 Análises das sequências de DNA obtidas por pirosequenciamento com as sequências correspondentes em bovino.....	26
5.3 Predição de genes.....	28
5.4 Identificação de sequências de DNA conservadas entre búfalo e boi.....	29
5.5 Identificação de sequências repetitivas.....	36
5.6 Análise comparativa com outros bovídeos.....	40
5.7 Comparação da sequência de DNA dos genes da raça Murrah com aquelas correspondentes na raça Mediterrâneo.....	43
6. CONCLUSÃO.....	47
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	49

ANÁLISE COMPARATIVA DE GENES DAS CASEÍNAS DE BÚFALO

RESUMO - Dentre as proteínas do leite, as caseínas (alfa-s1, alfa-s2, beta- e kappa-caseína) assumem papel de destaque devido ao alto valor nutritivo e às características físico-químicas que favorecem a fabricação de derivados do leite. Essas proteínas são codificadas pelos genes CSN1S1, CSN1S2, CSN2 e CSN3. A fim de realizar análise comparativa dos genes das caseínas de búfalo, o presente trabalho teve como objetivo a identificação, caracterização e sequenciamento de clones da biblioteca genômica de búfalo, visando analisar a estrutura molecular de genes das caseínas. Dentre os 33.792 clones avaliados, foram identificados dois clones positivos para genes das caseínas, um para o gene CSN1S1 (clone A/2) e outro para o gene CSN3 (clone L/8). Na sequência de DNA obtida a partir do clone A/2, foram identificados os genes CSN1S1 inteiro e CSN2 parcial, enquanto que nas sequências de DNA do clone L/8 identificou-se o gene CSN3 parcial. O gene CSN1S1 apresentou 17.008 bp organizados em 19 éxons com tamanhos variando de 24 bp a 380 bp e 18 íntrons com tamanhos de 90 bp a 1.710 bp. As análises comparativas revelaram que os éxons e íntrons desse gene apresentaram conservação acima de 85% entre búfalo e boi. As porções do gene CSN2 identificadas incluíram o éxon 9 e parte do íntron 8, os quais mostraram conservação acima de 98% com as sequências correspondentes em boi. Já as sequências parciais do gene CSN3 abrangeram parte dos íntrons 2 e 3 e o íntron 4 completo, além dos éxons 3, 4 e 5. Estas sequências apresentaram conservação acima de 94% com as correspondentes em boi. As análises de identificação de sequências repetitivas mostraram que 43,83% e 44,98% das sequências de DNA do clone A/2 e L/8, respectivamente, são representadas por elementos retrotransposons. Nas análises comparativas, tanto o gene CSN1S1 quanto o gene CSN3 parcial apresentaram sequências repetitivas búfalo específicas. A sequência codificante do gene CSN1S1 de búfalo apresentou 98%, 93% e 90% de identidade com as sequências homólogas de boi, cabra e ovelha, respectivamente. Já a sequência de aminoácidos da caseína alfa-s1 de búfalo mostrou de 95%, 87% e 84% de identidade com as respectivas sequências correspondentes em boi, cabra e ovelha. A comparação entre as sequências do gene CSN1S1 de animais das raças Murrah e Mediterrâneo revelou 42 potenciais polimorfismos de nucleotídeo único (SNP, "Single Nucleotide Polymorphism").

Palavras-chave: Biblioteca genômica, *Bubalus bubalis*, polimorfismos, pirosequenciamento, sequências repetitivas

COMPARATIVE ANALYSIS OF THE BUFFALO CASEIN GENES

ABSTRACT - Among milk proteins, the caseins (alpha-s1, alpha-s2, beta- and kappa-casein) play a crucial role considering their high nutritional value and physicochemical characteristics which contribute to the manufacture of dairy products. These proteins are encoded by the CSN1S1, CSN1S2, CSN2 and CSN3 genes, respectively. In order to analyze the buffalo casein genes and compare the sequences with other species, the goal of the present study was to identify, characterize and sequence clones from a buffalo genomic library. A total of 33,792 clones were evaluated, and two clones were identified as positive, one for the CSN1S1 gene (clone A/2) and other for the CSN3 gene (clone L/8). The DNA sequence from clone A/2 identified the whole CSN1S1 and a partial sequence from the CSN2 genes. The DNA sequence from clone L/8 revealed a partial sequence from the CSN3 gene. The CSN1S1 gene presented a total of 17,008 bp organized in 19 exons ranging from 24 bp to 380 bp and 18 introns ranging from 90 bp to 1,710 bp. Comparative analysis showed sequence conservation higher than 85% on exons and introns of the CSN1S1 gene when compared with the cattle gene sequence. The partial sequence from the CSN2 gene included exon 9 and part of intron 8, with conservation higher than 98% when compared with the cattle sequence. The partial sequences of the CSN3 gene included parts of the introns 2 and 3, the whole sequence of intron 4 and exons 3, 4 and 5. These sequences showed conservation higher than 94% with cattle. The identification of repetitive sequences showed that 43.83% of DNA sequence from clone A/2 and 44.98% from clone L/8 were represented by retrotransposable elements. Further comparative analysis showed buffalo specific repetitive sequences in the CSN1S1 gene and the partial CSN3 gene with when compared with other bovids species. The coding sequence of the buffalo CSN1S1 gene showed 98%, 93%, and 90% of identity with the correspondent sequences in cattle, goat and sheep, respectively. The amino acids sequence of the buffalo alpha-s1 casein had 95%, 87% and 84% of identity with the respective corresponding sequences in cattle, goat and sheep. The comparison between the CSN1S1 genes of animals from Murrah and Mediterranean breeds revealed the occurrence of 42 potentials single nucleotide polymorphisms (SNPs).

Key-words: *Bubalus bubalis*, genomic library, polymorphisms, pyrosequencing, repetitive elements

1. INTRODUÇÃO

O búfalo é um animal originário do sudeste da Ásia que foi introduzido na África, Europa e, mais recentemente, no continente americano. Esses animais pertencem à família Bovidae, a mais diversa das nove famílias da ordem Artiodactyla, cuja subfamília Bovinae compreende seis gêneros: *Angra*, *Bibos*, *Bison*, *Bos*, *Bubalus* e *Syncerus*. O gênero *Syncerus* é representado pelos búfalos selvagens de origem africana e *Bubalus* pelos búfalos de origem asiática. Deste último, somente os búfalos da espécie *Bubalus bubalis* são criados para fins econômicos (búfalos domésticos) (LENSTRA; BRADLEY, 1999).

Estudos morfológicos (MACGREGOR, 1939), bioquímicos (BARKER et al., 1997a; AMANO, 1983), moleculares (LAU et al., 1998; BARKER et al., 1997b; KIKKAWA et al., 1997) e genéticos (IANNUZZI, 1994; BERARDINO; IANNUZZI, 1981; FISCHER; ULBRICH, 1968), mostraram que os búfalos da espécie *Bubalus bubalis* podem ser divididos em dois grupos: os búfalos de rio (2n=50) e os búfalos de pântano (2n=48). Dentre as 22 raças de búfalo de rio citadas por Moioli e Bhorguese (2005), somente as raças Jafarabadi, Mediterrâneo e Murrah são encontradas no Brasil.

Esses animais são conhecidos pela rusticidade e elevada capacidade de adaptação, podendo sobreviver em diversos ambientes de grandes variações de clima, relevo e vegetação. A adaptabilidade em áreas impróprias para bovinos contribui para que o produtor explore o búfalo tanto para produção de carne quanto de leite bubalino (BERNARDES, 2007; NARDI JUNIOR et al., 2012)

Segundo estimativas da FAO (do inglês “Food and Agriculture Organization of the United Nations”), a população mundial de búfalos abrange mais de 199 milhões de animais, distribuídos em 42 países (FAO, 2013). O Brasil possui o maior rebanho bubalino das Américas, compreendendo 1,18 milhões de animais distribuídos da seguinte forma: 63,5% na região Norte, 10,2% no Nordeste, 10,5% no Sul, 10,3% no Sudeste e 5,4% no Centro-Oeste. Dados do censo agropecuário brasileiro de 2010 mostraram que esse rebanho apresentou crescimento de 7,8%, em relação ao ano anterior, enquanto que o rebanho bovino aumentou 1,6% (IBGE, 2010).

O alto valor agregado dos produtos derivados do leite de búfala e o aumento da demanda pelos mesmos têm motivado criadores brasileiros a investirem na

bubalinocultura para exploração de leite, particularmente no sudeste do país, onde se encontra os maiores centros consumidores (BERNANDES, 2007; SENO; CARDOSO; TONHATI, 2007). De acordo com Bernardes (2007), a produção nacional de leite bubalino foi de 92,3 milhões de litros em 2006, produzidos por aproximadamente 82 mil animais pertencentes a 2.500 rebanhos. Esse autor ainda revelou a existência de pelo menos 150 laticínios transformando anualmente 45 milhões de litros de leite de búfala em 18,5 mil toneladas de derivados, os quais geram um faturamento bruto de 55 milhões de dólares aos laticínios e 17 milhões aos criadores.

No Brasil, a produção média de leite bubalino tem alcançado 1.583,7 litros por lactação com média diária de 7,3 litros por animal, sendo que cerca de 30% das búfalas produzem mais de 2.000 litros de leite por lactação. Já a produção média das fêmeas bovinas encontra-se ao redor de 1.290 litros por lactação com média diária de 4,3 litros por animal (ANUALPEC, 2009; NARDI JUNIOR; SANTOS, 2013).

O leite de búfala, quando comparado com o leite bovino, apresenta maiores teores de gordura (8,16% *versus* 3,68%), proteína (4,5% *versus* 3,70%) e sólidos totais (17% *versus* 12%). Embora apresente alto teor de gordura, o leite bubalino possui menor concentração de colesterol por litro quando comparado com o leite de outros animais de produção (VERRUDA; SALGADO, 1994). Na elaboração de derivados, os altos teores de proteína e gordura do leite de búfala fazem com que seu rendimento seja 40% superior ao leite bovino (TEIXEIRA; BASTIANETTO; OLIVEIRA, 2005). Tal superioridade no rendimento, aliada ao maior valor agregado de seus produtos finais, têm estimulado laticínios a remunerarem essa matéria prima com valores duas vezes maiores àqueles pagos pelo leite bovino (MACEDO et al., 2001; BERNANDES, 2007).

A produção de leite de búfala é completamente absorvida pelo mercado interno brasileiro, sendo utilizada principalmente na fabricação de queijo tipo “mozzarella”. Além deste queijo, o leite bubalino é utilizado na fabricação de outros derivados, tais como queijo tipo minas frescal, provolone, ricota, manteiga, requeijão, doce de leite e iogurte (BERNANDES, 2007). De acordo com informações da ABCB (2010), a procura por produtos derivados do leite de búfala tem aumentado 25% por ano e está longe de ser atendida. Por isso, é necessário o desenvolvimento de

medidas que otimizem a produção do leite bubalino de modo que a demanda de mercado seja atendida.

O aumento da produção e qualidade do leite pode ser alcançado por meio de estratégias como nutrição adequada, boas práticas de manejo, controle de doenças e melhoramento genético (DOMINGUES; LANGONI, 2001; KIPLAGAT; LIMO; KOSGEY, 2012). No Brasil, iniciativas envolvendo projetos e programas de melhoramento genético voltados à bubalinocultura, tal como a PROMEBUL (Programa de Melhoramento Genético de Bubalinos - 1997), têm sido implementadas desde a década de 90. No entanto, essas iniciativas não promoveram progresso genético intensivo dos rebanhos bubalinos leiteiros em decorrência das limitações biológicas destes animais (sazonalidade reprodutiva e menor porcentagem de fertilização por inseminação artificial), bem como da ausência de ferramentas de seleção embasadas em informações genéticas próprias da espécie (MARCONDES, 2011).

Diante desse panorama, o conhecimento sobre os genes responsáveis pelo processo de lactação em búfalas mostra-se de extrema valia, uma vez que essas informações podem auxiliar no entendimento das particularidades genéticas da espécie que refletem na produção e qualidade do leite. Além disso, essas informações podem contribuir para estudos de seleção genômica por meio da inclusão de marcadores moleculares nas predições dos valores genéticos de rebanhos bubalinos leiteiros.

2. REVISÃO DE LITERATURA

2.1 Caseínas

O leite é uma secreção nutritiva constituída pela combinação de elementos sólidos e água. Os elementos sólidos (proteínas, lipídios, carboidratos, sais minerais e vitaminas) compreendem aproximadamente 13% do leite, dos quais as proteínas representam entre 3% a 4%, constituídas por dois grupos distintos: as proteínas do soro e as caseínas (BRITO et al., 2007).

As proteínas do soro representam aproximadamente 20% das proteínas do leite e possuem característica de permanecerem solúveis no soro do leite após a precipitação das caseínas em pH 4,6 e a 20°C, ou sob a ação de clivagem enzimática. As principais proteínas desse grupo são: beta-lactoglobulina, alfa-lactoalbumina, albumina sérica e imunoglobulinas (FARREL et al., 2004; KRISSANSEN, 2007).

Em bovídeos, as caseínas representam cerca de 80% das proteínas presentes no leite, sendo formadas por quatro proteínas (caseína alfa-s1, alfa-s2, beta-caseína e kapa-caseína) sintetizadas nas células epiteliais das glândulas mamárias (RIJNKELS, 2002). Essas proteínas são precursoras de diversos peptídeos bioativos, os quais se encontram inativos dentro da sequência da proteína original, sendo liberados após a proteólise enzimática (KORHONEN et al., 1998; KRÜGER et al., 2006). Alguns desses peptídeos agem como opióides, imunomoduladores e inibidores enzimáticos, enquanto outros apresentam atividade antitrombótica e atuam como carreadores de minerais (KITTS et al., 1992; ZUCHT et al., 1995; KORHONEN et al., 1998; MEISEL; BOCKELMANN, 1999).

As caseínas são classificadas em dois tipos, as denominadas caseínas sensíveis ao cálcio (caseína alfa-s1, alfa-s2 e beta-caseína) e as caseínas insensíveis ao cálcio (kapa-caseína) (SWAISGOOD, 2003). Na forma natural, essas proteínas se agregam em estruturas termodinamicamente estáveis denominadas micelas, as quais apresentam as caseínas sensíveis ao cálcio na porção hidrofóbica central e as caseínas insensíveis ao cálcio na camada hidrofílica periférica, conferido estabilidade físico-química à micela (RASMUSSEN et al., 1999). As micelas de caseínas têm a função de sequestrar partículas de fosfato de cálcio no leite,

caracterizando-se como a principal fonte desses minerais para a nutrição de neonatos. O transporte de fosfato de cálcio por meio dessas micelas contribui tanto para os processos de mineralização dos ossos de recém-nascidos quanto para o controle da concentração de cálcio e fosfato no leite, impedindo a calcificação das glândulas mamárias (HOLT et al., 2013).

Do ponto de vista econômico, as caseínas assumem papel de destaque devido ao alto valor nutritivo e às características físico-químicas que favorecem a fabricação de produtos derivados do leite (HUPERTZ et al., 2006). Segundo Fox e McSweeney (1998), só é possível submeter o leite a processos em altas temperaturas devido à elevada estabilidade dessas proteínas ao calor. Em função dessas características, os genes das caseínas têm se tornado alvo de intensos estudos, principalmente em bovinos, caprinos e ovinos. Nestas espécies, os genes das caseínas têm apresentado grande número de variações alélicas (CERIOTTI et al., 2004; MARLETTA et al., 2007; CAROLI; CHESSA; ERHARDT, 2009).

As caseínas de bovídeos são codificadas pelos genes CSN1S1, CSN1S2, CSN2 e CSN3 que estão agrupados em uma região específica do genoma, cujo tamanho varia de 250 a 350 kilobases (Kb), dependendo da espécie (RIJNKELS, 2002). Em bovino, os quatro genes das caseínas estão caracterizados e se encontram localizados no cromossomo 6 (BTA6) em uma região de 200 Kb, arranjados na seguinte ordem: CSN1S1, CSN2, CSN1S2 e CSN3 (LEMAY et al., 2009).

O gene CSN1S1 codifica a proteína caseína alfa-s1 que tem o papel de captar e transportar partículas de fosfato e cálcio no leite. Em bovino, esse gene apresenta 17.536 bp e 19 éxons com tamanhos que variam de 24 bp a 385 bp. Até o momento, estudos de polimorfismos identificaram nove alelos nesse gene (A, B, C, D, E, F, G, H e I), dos quais o alelo B apresenta relação com uma maior concentração de proteínas e sólidos totais no leite, resultando em maiores rendimentos de produtos derivados do leite (CAROLI; CHESSA; ERHARDT, 2009).

O gene CSN1S2 codifica a proteína caseína alfa-s2 que desempenha a mesma função da caseína alfa-s1, captação e transporte de fosfato e cálcio no leite. O gene CSN1S2 bovino abrange uma região de 18.480 bp contendo 18 éxons com tamanhos que variam de 21 bp a 266 bp. Estudos de polimorfismos identificaram

quatro alelos (A, B, C e D) nesse gene, os quais não apresentaram associação expressiva com características do leite (CAROLI; CHESSA; ERHARDT, 2009).

O gene CSN2 codifica a proteína beta-caseína cuja função é alterar propriedades específicas da superfície da micela. O gene CSN2 bovino apresenta 9.355 bp e nove éxons com tamanhos que variam de 29 bp a 498 bp. Até o momento, foram identificados 12 alelos nesse gene (A^1 , A^2 , A^3 , B, C, D, E, F, G, H^1 , H^2 e I), dos quais o alelo B tem se mostrado mais favorável à coagulação do leite (DI STASIO; MARIANI, 2000; CAROLI; CHESSA; ERHARDT, 2009).

O gene CSN3 codifica a proteína kapa-caseína que tem como papel a estabilização e determinação do tamanho da micela (DALGLEISH, 2011). Em bovino, o gene compreende uma região de 13.062 bp, com cinco éxons variando de 33 bp a 517 bp. Grande parte da sequência codificante da kapa-caseína está contida no exón 4, onde 12 alelos (A^1 , A^2 , A^3 , B, C, D, E, F, G, H^1 , H^2 e I) foram identificados. O alelo B mostra-se associado com o aumento na concentração da proteína kapa-caseína no leite (HALLÉN et al., 2007; CAROLI; CHESSA; ERHARDT, 2009), gerando micelas menores, o que resulta em coágulos com consistência mais firme, contribuindo com aumento de até 10% na produção de queijos (WALSH et al., 1998).

Em bovino, o sequenciamento dos genes das caseínas foi imprescindível para realização de estudos de transgenia e de genética de populações (KAMINSKI, 2004). No estudo desenvolvido por Brophy et al. (2003), o conhecimento da estrutura molecular dos genes CSN2 e CSN3 foi essencial para que cópias desses genes fossem introduzidas em fibroblastos de 11 fêmeas bovinas, as quais passaram a produzir leite com o dobro de kapa-caseína e com 8% a 20% a mais de beta-caseína. No estudo realizado por Kaminski et al. (2008), o microarranjo de DNA denominado *MilkProtChip* foi construído a partir das sequências de 42 genes envolvidos no processo de lactação em bovino, incluindo os genes das caseínas. Essa ferramenta possibilitou a genotipagem simultânea de um conjunto de 71 SNPs em 300 animais da raça “Polish Black and White”, sendo que os SNPs localizados no gene CSN3 foram associados com maior concentração de gordura e proteína no leite.

Em búfalo, os estudos com os genes das caseínas têm avançado aos poucos nos últimos anos. Estudos citogenéticos por meio da técnica de hibridização *in situ* por fluorescência - FISH (do inglês “Fluorescence *in situ* Hybridization”), mapearam os genes CSN1S2 e CSN3 no cromossomo 7 bubalino (BBU7), o qual é homólogo ao cromossomo 6 bovino (BTA6) (IANNUZZI et al., 2003). Posteriormente, mapas de alta resolução do genoma bubalino corroborou a localização do gene CSN3 no BBU7 (AMARAL et al., 2008). No final de março de 2014, foi disponibilizado no banco de dados americano NCBI uma versão da sequência do genoma bubalino, UMD_CASPUR_WB_2.0 (<http://www.ncbi.nlm.nih.gov/assembly/?term=Bubalus+bubalis>), representada por um animal da raça Mediterrâneo. Nesta versão, os genes das caseínas estão distribuídos na mesma ordem de seus correspondentes no genoma bovino, sendo CSN1S1, CSN2, CSN1S2 e CSN3 (nº de acesso no Genbank: NW_005785429.1).

A sequência UMD_CASPUR_WB_2.0 do genoma total bubalino mostra o gene CSN1S1 bubalino com 17.010 bp distribuídos em 19 éxons com tamanhos variando de 24 bp a 380 bp. O gene CSN2 com 8.892 bp abrangendo 9 éxons com tamanhos que variam de 24 bp a 498 bp. O gene CSN1S2 com 18.631 bp organizados em 18 éxons com tamanhos de 21 bp a 267 bp, e o gene CSN3 com 13.204 bp distribuídos em cinco éxons com tamanhos que variam de 33 bp a 517 bp.

Embora a sequência dos genes das caseínas de búfalo esteja definida, não existem até o momento, estudos comparativos da sequência dos mesmos com aquelas correspondentes em outras espécies de bovídeos, principalmente considerando as regiões intergênicas, as quais geralmente apresentam alta variabilidade entre espécies próximas, contribuindo para a identificação de sequências não codificantes conservadas com funções regulatórias. Além disso, os estudos comparativos são fundamentais para obtenção de informações sobre os mecanismos evolutivos do genoma que resultaram na especiação, adaptação e diversificação de espécies (THOMAS; TOUCHMAN, 2002; LINBLAD-TOH et al., 2005; MIKKELSEN et al., 2007; AFOLDI; LINDBLAD-TOH, 2013).

Poucos trabalhos sobre polimorfismos nos genes das caseínas de búfalo estão disponíveis na literatura. Até o momento, foram identificados dois alelos no gene CSN1S1 (A e B) (CHIANESE et al., 2009; BONFATTI et al., 2012; EL NAHAS

et al., 2013), três alelos no gene CSN1S2 (A, B e C) (CHIANESE et al., 1996), três alelos no gene CSN2 (A, A² e B) (FERRANTI et al., 1998; MISHRA et al., 2009) e dois alelos no gene CSN3 (A e B) (GOUDA; GALAL; ABDELAZIZ, 2013).

Tendo em vista o número ainda limitado de informações sobre os genes das caseínas em búfalo, torna-se pertinente o desenvolvimento de estudos de resequenciamento destes genes em animais de diferentes raças, visando a obtenção de potenciais polimorfismos que possam estar associados às características de produção e qualidade do leite.

Estudos de resequenciamento de genes podem ser realizados a partir de clones de bibliotecas genômicas do tipo BAC (do inglês “Bacterial Artificial Chromosome”). Essas bibliotecas são consideradas uma ferramenta genômica indispensável para estudos de isolamento de genes específicos (LIU et al., 2010), mapeamento físico de regiões genômicas, cromossomos únicos ou de genomas inteiros (CHEN et al., 2004). Em 2012, as pesquisas desenvolvidas pelo grupo “Genômica de Búfalos” do CNPq resultaram na construção de uma biblioteca genômica do tipo BAC de um macho da raça Murrah (STAFUZZA et al., 2012).

A disponibilidade de tal ferramenta genômica permite a identificação de clones contendo genes das caseínas para posterior isolamento e resequenciamento dos mesmos. Análises comparativas das sequências de DNA obtidas com aquelas correspondentes em outros bovídeos possibilita identificar sequências repetitivas e sequências conservadas com possíveis funções regulatórias, assim como sequências únicas do genoma de búfalo. Essas informações podem contribuir para o entendimento dos elementos genéticos responsáveis pelas diferenças na produção e qualidade do leite entre os bovídeos. Já a comparação da sequência dos genes das caseínas de Murrah, com aquelas correspondentes na raça Mediterrâneo disponíveis no banco de dados NCBI, auxilia na identificação de potenciais SNPs, os quais podem ser utilizados em estudos de associação com fenótipos de interesse.

2.2 Sequências repetitivas no genoma de eucariotos

Sequências repetitivas são fragmentos de DNA que estão presentes em múltiplas cópias no genoma. Essas sequências são representadas por dois grupos distintos de repetições: repetições esparsas e regiões de baixa complexidade.

Repetições esparsas são sequências de DNA que se repetem em várias posições da cadeia de DNA, sendo representadas principalmente por cópias incompletas e inativas de elementos transponíveis. Já as regiões de baixa complexidade são sequências de DNA onde existe uma alta incidência de certas bases ou de pequenas sequências de bases, sendo representadas principalmente pelos microssatélites (JURKA, 2007).

Elementos transponíveis são fragmentos de DNA que possuem a habilidade de se moverem entre regiões do genoma. Estes elementos são agrupados de acordo com mecanismo de transposição em duas classes principais: classe I e classe II. Os elementos da classe I também são chamados de retrotransposons, uma vez que se movimentam por meio de um intermediário de RNA, cuja transcrição reversa resulta em cópias de DNA complementar que são inseridas em outro local do genoma. Já os elementos da classe II se deslocam a partir de um intermediário de DNA por um mecanismo de excisão do sítio original e reinserção em outro sítio no genoma, sendo denominados de transposons (REBOLLO; ROMANISH, MAGER, 2012). A classe dos retrotransposons pode ainda ser subdividida em duas subclasses: os retrotransposons LTR (do inglês “Long Terminal Repeat”) e os retrotransposons não-LTR. Dentro da subclasse dos retrotransposons não-LTR são encontrados os elementos transponíveis conhecidos como LINEs (do inglês “Long Interspersed Element”) e SINEs (do inglês “Short Interspersed Element”), os quais diferem em decorrência do primeiro codificar a enzima transcriptase reversa enquanto que o último não, sendo transportados com o auxílio das enzimas codificadas pelos elementos LINEs (CORDAUX; BATZER, 2009).

Os elementos transponíveis apresentam componentes cis-regulatórios necessários para a sua própria expressão, sendo eles: promotores, sítios de processamento alternativo (do inglês “alternative splicing”) e terminação, acentuadores e isoladores (DAVIS et al., 1998; MAGER et al., 1999; REBOLLO et al., 2012). Tais componentes podem não só influenciar a própria atividade dos elementos transponíveis, como também a expressão de genes adjacentes (DUNN et al., 2006). No estudo desenvolvido por Romanish et al. (2007) foi observado que a região promotora dos genes NAIP (do inglês, “Apoptosis Inhibitory Protein”), os quais estão associados com a inibição da apoptose de células neuronais, originou do

retrotransposons LTR em humano e roedores. No estudo conduzido por Pérez et al. (1994), a inserção do retrotransposon não-LTR L1_BT (LINE) na região 3'UTR (UTR, do inglês "Untranslated Region") do gene CSN1S1 de cabra foi associada com redução na síntese de proteínas. Analisando um conjunto de fatores de transcrição de humano (ESR1, TP53 e RELA) e camundongo (POUSF1 e CTCF) associados a diversos sistemas biológicos, Bourque et al. (2008) observaram que esses fatores frequentemente se ligam a sítios contidos em elementos transponíveis, o que evidencia o importante papel destes elementos na expansão do repertório de sítios ligantes de fatores de transcrição.

Segmentos derivados de elementos transponíveis têm sido frequentemente encontrados em regiões codificantes de genes hospedeiros em boi (IWASHITA et al., 2006; ALMEIDA et al., 2007; ALMEIDA et al., 2008), camundongo (BEST et al., 1996) e humano (BRITTEN, 2004). A inserção desses elementos em regiões codificantes pode modificar a função de um gene, caso esta mudança seja adaptativa e conservada, resultando no fenômeno conhecido como domesticação molecular ou exaptação (MILLER; MCDONALD; PINSKER, 1997; BRANDT et al., 2005). Almeida et al. (2008) observaram que o último éxon do gene FASTKD3 bovino, o qual está associado a processos celulares como a apoptose, originou do elemento retrotransposon ART2A, cuja inserção resultou em uma proteína diferente daquelas codificadas por humano, chimpanzé e cachorro.

Os microssatélites apresentam unidades de repetição que variam de dois a seis nucleotídeos, as quais se repetem de 20 a 50 vezes, podendo alcançar tamanhos que variam de 60 a 300 bp (ELLEGRÉN, 2004; LIU; CORDES, 2004). Esse tipo de sequência repetitiva ocorre principalmente em regiões não codificantes do genoma e apresenta as seguintes características: elevado nível de polimorfismos, altas frequências de ocorrência, facilidade na identificação e distribuição em todo genoma. Tais características fazem com que os microssatélites sejam amplamente utilizados em estudos de diversidade genética, controle de origem, caracterização da estrutura e grau de endogamia da população e também de identificação de QTLs (do inglês "Quantitative Trait Loci") (CITEK et al., 2006, BARLOWSK et al., 2012).

3. OBJETIVOS

O presente trabalho teve os seguintes objetivos:

- ✓ Realizar varredura na biblioteca genômica de búfalo com o propósito de identificar clones positivos para os genes das caseínas para o processo de sequenciamento;
- ✓ Analisar as sequências de DNA obtidas visando caracterizar os genes das caseínas quanto à sua organização estrutural (regiões 5' e 3'UTR, número e tamanho de éxons e íntrons);
- ✓ Realizar a comparação das sequências de DNA obtidas com aquelas correspondentes em boi para avaliar a porcentagem de conservação das mesmas entre búfalo e boi;
- ✓ Avaliar as sequências de DNA obtidas quanto à presença de sequências repetitivas;
- ✓ Comparar as sequências codificantes e de aminoácidos dos genes de búfalo com as sequências homólogas de outros bovídeos visando estabelecer a porcentagem de identidade entre as diferentes espécies;
- ✓ Realizar a comparação das sequências de DNA dos genes das caseínas do animal da raça Murrah com aquelas de um animal da raça Mediterrâneo, a fim de identificar potenciais polimorfismos do tipo SNPs.

4. MATERIAL E MÉTODOS

As soluções citadas no item Material e Métodos estão descritas em detalhe no Apêndice 1, em ordem alfabética.

4.1 Seleção dos clones positivos para genes das caseínas

A seleção dos clones foi realizada a partir da biblioteca genômica de búfalo, armazenada no Laboratório de Genômica Comparativa - IBILCE/UNESP, câmpus de São José do Rio Preto/SP, cuja metodologia de construção e caracterização preliminar encontram-se descritas em detalhes em Stafuzza et al. (2012).

Os clones da biblioteca genômica de búfalo estão armazenados segundo o sistema de organização tridimensional (*singlepool*, *superpool* e linha/coluna) (STAFUZZA et al., 2012), tornando possível a seleção de clones da biblioteca por meio da Reação em Cadeia da Polimerase (PCR).

Um *singlepool* inclui os clones de uma única placa de armazenamento (384 poços), enquanto o *superpool* representa o conjunto de clones de oito placas de 384 poços (3.072 clones). Os clones referentes a cada *singlepool* e a cada *superpool* foram armazenados em microtubos de 2 mL.

O sistema linha/coluna de cada uma das placas de 384 poços refere-se em armazenar todos os clones de uma determinada linha em um único microtubo de 1,5 mL (16 linhas: A→P), e todos os clones presentes em uma determinada coluna da placa de armazenamento em outro microtubo de 1,5 mL (24 colunas: 1→24). Dessa maneira, o sistema de organização linha/coluna de cada placa pode ser organizado em 40 microtubos, sendo que cada clone está representado duas vezes (uma vez na linha e outra na coluna).

Para otimizar a busca pelos clones, as reações de PCR foram realizadas de acordo com os seguintes passos:

- **1º passo:** reação de PCR com DNA do conjunto de clones que compõem os *superpools* (oito placas), visando identificar qual conjunto de oito placas poderia ser positivo para o gene específico;
- **2º passo:** reação de PCR com DNA dos clones das oito placas (*singlepools*), visando identificar a placa que contém o clone positivo para o gene específico;

- **3º passo:** reação de PCR com DNA dos clones das linhas e colunas da placa de 384 poços que mostrou amplificação positiva para o gene específico permitindo identificar a posição do clone positivo na placa (sistema linha/coluna).

A Figura 1 apresenta como exemplo, um esquema da sequência de experimentos de PCR realizados na avaliação de três *superpools* quanto à presença de um gene de interesse.

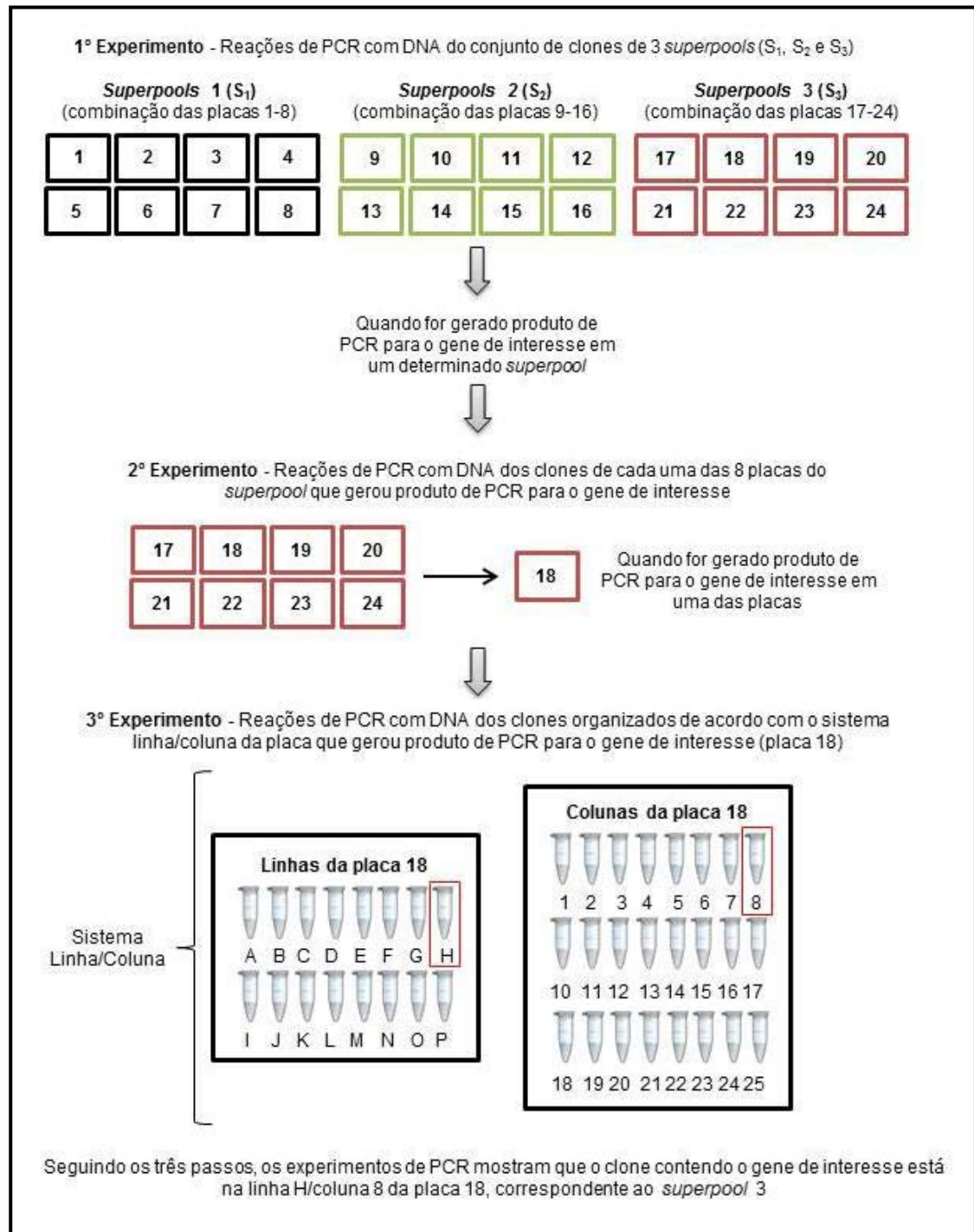


Figura 1. Sequência dos experimentos de PCR realizados na busca por clones contendo um gene de interesse em três *superpools* da biblioteca BAC de búfalo

Para seleção dos clones positivos para os genes das caseínas foram utilizados iniciadores para PCR dos quatro genes das caseínas, previamente mapeados no cromossomo 6 bovino. A Tabela 1 apresenta a relação dos iniciadores para PCR que foram utilizados com suas respectivas sequências (senso e anti-senso), temperatura de anelamento e referência bibliográfica ou banco de dados para os mesmos.

Tabela 1. Informações sobre os genes utilizados, incluindo símbolo do gene, as sequências senso (F) e anti-senso (R) dos iniciadores para PCR, temperatura de anelamento e referência bibliográfica.

Símbolo do gene	Sequência senso (F) e anti-senso (R) dos iniciadores para PCR (5'-3')	Temperatura de anelamento	Referência Bibliográfica
CSN1S1	F: CCTTACCACTTCACTATTGCCAC R: CTTTCTTAAAGCATAGAGCATATCC	54°C	UniSTS: 253703*
CSN1S2	F: TCTCCTCCTAGGATTGGAAAGA R: TTGAGACGGTTTGGGAATCT	54°C	SAZANOV et al., 2006
CSN2	F: ACAGCCTCCCACAAAACATC R: AGACTGGAGCAGAGGCAGAG	65°C	SAZANOV et al., 2006
CSN3	F: CGCTGTGAGAAAGATGAAAGATT R: TTTGGGCAGGTGACCTAACT	63°C	UniSTS: 519151*

*UniSTS refere-se ao número de identificação da respectiva sequência de iniciadores para PCR no banco de dados NCBI (<http://www.ncbi.nlm.nih.gov/probe>).

As reações de PCR foram realizadas em um volume total de 25 µL, utilizando 2 µL do material biológico dos clones, os quais foram previamente submetidos à 98°C por 15 minutos para a lise da parede celular. Em seguida foram adicionados 0,2 mM de cada iniciador para PCR (senso e anti-senso); 10 mM de dNTPs; 10 mM de Tris-HCl; 50 mM de KCl; 1,5 mM de MgCl₂ e 0,5 U/µL da DNA polimerase *AmpliTaQGoldTM* (Life Technologies). A amplificação foi realizada com desnaturação inicial a 95°C por 10 minutos, seguida de 35 ciclos com desnaturação a 95°C por 30 segundos, anelamento a 54°C-65°C por 30 segundos (dependendo do iniciador para PCR) e extensão a 72°C por 30 segundos, seguida de extensão final a 72°C por 7 minutos. Os produtos de PCR foram visualizados em gel de agarose 2%, corado com brometo de etídio e submetidos à eletroforese horizontal a 120V por 50 minutos, em tampão TBE 0,5X. Os resultados foram fotodocumentados com câmera digital *DC290 KodakTM* em luz ultravioleta.

Após a identificação dos clones positivos para os genes das caseínas de búfalo, foram realizadas as etapas de multiplicação celular dos clones e purificação dos DNA BAC.

Com o auxílio de um palito de madeira, uma porção da solução contendo os clones selecionados foi transferida e espalhada em placas com meio LB-ágar acrescido de cloranfenicol (12,5 mg/mL - *SERVA*), 20 µL de IPTG a 20% (Isopropil-β-D-tiogalactosídeo - *Eppendorf*) e 100 µL de XGal a 2% (5-bromo-4-cloro-3-indolil-β-D-galactopiranoside - *Life Technologies*), a qual foi mantida em estufa a 37°C por 12 horas para o crescimento dos clones. Cada clone foi colocado em um Erlenmeyer contendo 80 mL de meio LB (*Luria-Broth*) acrescido de cloranfenicol (12,5 mg/mL), sendo mantido a 37°C sob agitação de 200 rpm durante 21 horas para replicação dos clones. Após esse período, 75 mL da cultura foram transferidos para 10 tubos com fundo cônico de 15 mL (7,5 mL de cultura/tubo) para a realização da extração do DNA.

A extração do DNA dos clones foi realizada com o *kit PhasePrep™ BAC DNA* (*Sigma-Aldrich*), seguindo indicações do fabricante, as quais estão brevemente descritas a seguir. Após a propagação dos clones, os mesmos foram submetidos à centrifugação durante 10 minutos a 4.000 rpm, onde o sobrenadante obtido foi descartado e o precipitado ressuspensionado em 250 µL de Solução de Ressuspensão. Posteriormente, os clones ressuspensionados foram transferidos para um microtubo de 1,5 mL, onde foi realizada a lise celular por meio da adição de 250 µL de Solução de Lise, seguida de incubação à temperatura ambiente por cinco minutos.

A neutralização da solução foi realizada pela adição de 250 µL de Solução de Neutralização, seguida de incubação em gelo por cinco minutos. O volume obtido foi submetido à centrifugação durante cinco minutos a 14.000 rpm, a 4°C e o sobrenadante foi transferido para um novo microtubo de 1,5 mL.

A etapa de precipitação do DNA foi realizada pela adição de 450 µL de isopropanol, seguida de centrifugação por 20 minutos a 14.000 rpm e a 4°C. O sobrenadante foi removido e o precipitado lavado com 100 µL de etanol 70% à temperatura ambiente, seguido de centrifugação por cinco minutos com as mesmas condições descritas acima. O sobrenadante foi novamente descartado e o precipitado foi mantido à temperatura ambiente por cinco minutos para a secagem.

Em seguida, o RNA residual foi removido pela adição de 500 µL de Solução de Eluição e 1 µL de Coquetel RNase ao precipitado, o qual foi mantido em banho-maria à 60°C por cinco minutos.

Para ajuste da concentração de sais, foram adicionados 40 µL de acetato de sódio (3M e pH 7,0). As endotoxinas e outras impurezas foram retiradas por meio de uma remoção mediada por temperatura. Para isso, foram adicionados 100 µL de Solução de Remoção de Endotoxinas e a solução obtida foi aquecida em banho-maria à 37°C por cinco minutos, sendo posteriormente centrifugada por três minutos à 14.000 rpm em temperatura ambiente, para que ocorresse a separação em duas fases. A fase superior continha o DNA do clone e a fase inferior, as endotoxinas e impurezas. Dessa maneira, a fase superior foi transferida para um novo microtubo de 1,5 mL e a fase inferior foi descartada. A precipitação do DNA foi realizada pela adição de 540 µL de Solução de Precipitação, seguida de centrifugação por 20 minutos a 14.000 rpm e a 4°C. O sobrenadante foi removido e o precipitado lavado com 150 µL de etanol 70% em temperatura ambiente. A etapa de centrifugação foi repetida por mais cinco minutos e o sobrenadante obtido foi descartado. Foi realizada uma nova lavagem com 50 µL de etanol 70% em temperatura ambiente e mais uma etapa de centrifugação por 10 minutos. Ao final dessas etapas todo o sobrenadante foi removido, e o precipitado mantido em temperatura ambiente para a secagem.

Após a secagem, o DNA foi eluído em tampão TE (Tris-EDTA, pH 8,0) e sua concentração foi medida em espectrofotômetro (*Nanodrop ND1000TM* - *Thermo Scientific*) e ajustada para 100 ng/µL. Para verificar a integridade do DNA extraído, foi realizada uma eletroforese horizontal em gel de agarose 2% corado com brometo de etídio, contendo 100 ng de cada amostra de DNA. A eletroforese foi realizada a 120V por 50 minutos, em tampão TBE 0,5X. Os géis foram fotodocumentados por meio de câmera digital (*DC290 KodakTM*) em luz ultravioleta.

Para verificar se os clones identificados não eram falsos positivos, foram realizadas reações de PCR com o DNA purificado dos mesmos. As reações foram realizadas em um volume total de 25 µL utilizando 1 µL de DNA extraído e purificado do clone, 0,2 mM de cada iniciador para PCR (senso e anti-senso); 100 mM de dNTPs; 10 mM de Tris-HCl; 1,5 mM de MgCl₂ e 0,05U/µL da DNA polimerase

AmpliTaqGoldTM (Life Technologies). A amplificação foi realizada com desnaturação inicial à 95°C por 10 minutos, seguida de 35 ciclos com desnaturação a 95°C por 30 segundos, anelamento a 54°C-65°C por 30 segundos (dependendo do iniciador para PCR) e extensão a 72°C por 30 segundos, seguida de uma extensão final a 72°C por 7 minutos. Os produtos de PCR também foram visualizados em gel de agarose 2%, corado com brometo de etídio e submetidos à eletroforese horizontal a 120V por 30 minutos, em tampão TBE 0,5X. Os resultados foram fotodocumentados com câmera digital (*DC290 KodakTM*) em luz ultravioleta.

As amostras de DNA dos clones foram enviadas para o sequenciamento, o qual foi realizado por meio da plataforma de pirosequenciamento *GS FLX Titanium* (Roche 454) via serviço de terceiros no “454 Sequencing Center”, em Connecticut (EUA).

4.2 Análises das sequências de DNA obtidas por pirosequenciamento com as sequências correspondentes em bovino

A fim de verificar possíveis regiões de homologia entre as sequências de DNA dos clones e o genoma bovino, foram realizados alinhamentos entre as sequências de DNA de cada clone e as duas anotações do genoma bovino (UMD_3.1 e Btau_4.6.1), disponíveis no banco de dados americano NCBI (National Center for Biotechnology Information - <http://www.ncbi.nlm.nih.gov/assembly/?term=bos+taurus>). A anotação UMD_3.1 do genoma bovino foi realizada pelos pesquisadores do “Center for Bioinformatics and Computational Biology”, da universidade de Maryland em 2009. Já a anotação Btau_4.6.1 foi desenvolvida pelos pesquisadores do “Cattle Genome Sequencing International Consortium - Baylor College of Medicine (EUA)”, no ano de 2011.

Para a realização dos alinhamentos foi utilizado a ferramenta *BLASTn* (do inglês, “Basic Local Alignment Search Tool of Nucleotide”), a qual promove buscas por sequências de nucleotídeos correspondentes às sequências de DNA em estudo. Nas análises foram considerados os alinhamentos que obtiveram maiores valores de “score”, cobertura e identidade e menor valor de *e-value*. O “score” trata-se de um valor numérico que reflete a qualidade do alinhamento, sendo calculado a partir de

uma matriz de pontuação que considera as penalidades dadas aos alinhamentos errôneos (do inglês “mismatch”) e aos “gaps” (lacunas), os quais são inseridos quando ocorrem alterações nas sequências de nucleotídeos causadas por substituições e *indels* (deleções ou inserções), respectivamente. A cobertura da sequência (do inglês, “query cover”) representa a porcentagem da sequência de DNA de interesse que está incluída no alinhamento. Já a identidade trata-se de um valor percentual que expressa o número de nucleotídeos idênticos resultantes do alinhamento entre sequências de DNA. Por fim, o *e-value* expressa a probabilidade da sequência correspondente à sequência de DNA em estudo ter sido alinhada ao acaso.

4.3 Predição de genes

A predição de genes nas sequências de DNA foi realizada com o auxílio do programa *AUGUSTUS* (<http://augustus.gobics.de/> - HOFF; STANKE, 2013). Este programa faz predições de genes de acordo com a abordagem *ab initio*, a qual utiliza os sinais biológicos da própria sequência de DNA para a definição da estrutura do gene predito. Os sinais biológicos comumente utilizados são: sítio de início de tradução (ATG), sítios de fim de tradução (TAA, TAG e TGA), sítio doador (GT) e sítio receptor (AG).

Antes de realizar a predição de genes, as sequências de DNA foram mascaradas quanto à presença de elementos transponíveis, com o auxílio do programa *RepeatMasker* (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>). Sequências repetitivas do tipo repetições em “*tandem*” (microsatélites) não foram mascaradas, uma vez que podem estar presentes em regiões de éxons dos genes preditos. Assim, o parâmetro padrão “*mask interspersed and simple repeats*” foi substituído pelo “*don't mask simple repeats and low complexity DNA*” no programa *RepeatMasker*. Esse procedimento foi realizado porque determinadas sequências repetitivas presentes na sequência de DNA podem ser erroneamente preditas como genes (HOFF; STANK, 2013).

As sequências codificantes dos genes preditos para cada sequência de DNA foram avaliadas com a ferramenta *BLASTn*, utilizando o banco de dados referência

de RNA de *Bos taurus* (RefSeq RNA) para identificação de sequências de RNAs correspondentes em bovino. Posteriormente, as sequências de aminoácidos previstas pelo mesmo programa foram avaliadas com a ferramenta *BLASTp* (do inglês, “Basic Local Alignment Search Tool of Protein”), utilizando o banco referência de proteínas de bovino (RefSeq Protein) para identificação de sequências de aminoácidos homólogas entre as sequências de búfalo e boi.

Como o programa *AUGUSTUS* utilizou informações moleculares de um banco de dados de humano para realizar a predição de genes na sequência de DNA bubalino, foi realizada a confirmação dos genes preditos para as sequências de DNA obtidas a partir dos clones. Para tanto, essas sequências de DNA foram alinhadas com as sequências dos genes correspondentes em bovino, por meio da ferramenta *BLAST2seq*. As sequências dos genes em bovino foram obtidas no banco de dados “UCSC Genome Bioinformatics - University of California Santa Cruz” (<http://genome.ucsc.edu/>).

4.4 Identificação de sequências de DNA conservadas entre búfalo e boi

O grau de conservação das sequências de DNA de búfalo com aquelas correspondentes no genoma de boi foi identificado com auxílio do programa *wgVISTA* (<http://genome.lbl.gov/cgi-bin/WGVistaInput> - FRAZER et al., 2004). Para isso, a sequência de DNA foi alinhada com o “scaffold” do cromossomo 6 de boi (nº de acesso no *GenBank*: NW_001495210.2), pertencente à versão de anotação Btau_4.6.1.

O *wgVISTA* compara sequências de DNA de duas espécies, com até 10 mega bases, sendo uma delas anotada. As informações contidas na sequência de DNA anotada são utilizadas para direcionar a identificação de regiões conservadas na sequência de DNA em análise. De acordo com informações fornecidas pelo programa *wgVISTA*, as regiões consideradas conservadas apresentam janelas de 100 nucleotídeos com conservação acima de 70%.

4.5 Identificação de sequências repetitivas

As sequências repetitivas foram identificadas com auxílio do programa RepeatMasker (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>), o qual utiliza o banco de dados RepBase Update (JURKA et al., 2005), mantido pelo “Genetic Information Research Institute (GIRI)”. Esse programa permite a identificação sequências repetitivas do tipo elemento transponíveis, microssatélites e DNA satélite. Além dessas sequências repetitivas, o programa RepeatMasker auxilia na identificação de pequenos RNAs.

As classes de pequenos RNAs identificadas pelo programa RepeatMasker incluem os RNAs transportadores (tRNA), pequenos RNAs nucleares (do inglês “small nuclear RNA” - snRNA), RNAs ribossômicos (rRNA), srpRNA (do inglês “signal recognition particle RNA”) e scRNA (do inglês “small cytoplasmatic RNA”). As moléculas de tRNA têm o papel de transportar aminoácidos específicos para o ribossomo durante a síntese de proteínas. Os snRNAs são curtas moléculas de RNA que auxiliam no processamento de outros RNAs, enquanto que os rRNAs são constituintes estruturais e funcionais dos ribossomos (GRIFFITHS et al., 2006). Tanto os srpRNA quanto os scRNA são moléculas que participam do processo denominado “signal recognition particle”, o qual é responsável pelo transporte de proteínas dos ribossomos para membrana do retículo endoplasmático (NAKAMURA., 1995).

Na identificação das sequências repetitivas com o programa RepeatMasker foram utilizados os seguintes parâmetros:

- “cross_match” para a opção “search engine”;
- “slow sensibility”, a qual permite uma busca mais sensível;
- “mammal other than below” para especificar a fonte de DNA da sequência de interesse;
- “show alignments in query sequence orientation” para obter o alinhamento dos elementos identificados na sequência de interesse;
- “mask repetitive sequences replaced by strings of N”, onde as sequências repetitivas identificadas foram substituídas pela letra N;

- “mask interspersed and simple repeats”, para que as sequências de pequenos RNAs e repetições em “tandem” identificadas também fossem substituídas pela letra N;
- “skip bacterial insertion element check”, onde não foi verificada a presença de elementos pertencentes às sequências de DNA de bactérias.

4.6 Análise comparativa com outros bovídeos

Alinhamentos múltiplos das sequências codificantes e de aminoácidos dos genes preditos para búfalo com as sequências homólogas de outros bovídeos (boi, caprino e ovino) foram realizados com o auxílio do programa *Clustal Omega* (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), visando identificar as diferenças e a porcentagem de identidade entre essas sequências.

4.7 Comparação da sequência de DNA dos genes da raça Murrah com aquelas correspondentes na raça Mediterrâneo

Com o auxílio do programa *Clustal Omega*, as sequências dos genes de Murrah obtidas no presente estudo foram comparadas com as correspondentes em Mediterrâneo (UMD_CASPUR_WB_2.0), disponíveis no banco de dados público NCBI, visando identificar potenciais SNPs entre as mesmas.

5. RESULTADOS E DISCUSSÃO

5.1 Seleção dos clones positivos para genes das caseínas

Foram avaliados 33.792 clones da biblioteca genômica de búfalo, dentre os quais foram identificados dois clones positivos para os genes das caseínas CSN1S1 e CSN3. O clone localizado na linha A e coluna 2 da placa 33 (*Singlepool* 33) mostrou-se positivo para o gene CSN1S1. Já o clone presente na linha L e coluna 8 da placa 9 (*Singlepool* 9) mostrou-se positivo para o gene CSN3. Os demais genes das caseínas (CSN1S2 e CSN2) não foram encontrados na etapa de seleção dos clones. Há três explicações possíveis para estes genes não terem sido identificados na biblioteca BAC de búfalo:

- **Primeira:** a biblioteca não apresenta clones que contenham os genes CSN1S2 e CSN2 de búfalo;
- **Segunda:** os clones da biblioteca apresentam regiões parciais dos genes CSN1S2 e CSN2 de búfalo que não foram complementares aos pares de iniciadores para PCR utilizados durante a etapa de seleção;
- **Terceira:** a grande quantidade de componentes celulares provenientes do conjunto de clones que compõem os *superpools* (oito placas com 384 clones) impediu a identificação dos clones positivos para os genes CSN1S2 e CSN2 de búfalo durante as reações de PCR.

A Figura 2 apresenta os resultados obtidos nas reações de PCR realizadas para a seleção dos clones positivos para o gene CSN1S1. A Figura 2A ilustra um gel de agarose com os resultados obtidos nas reações de PCR com os 11 *superpools* da biblioteca, onde o *superpool* 5 conteve o clone positivo para o gene CSN1S1, indicado pela seta vermelha. Na Figura 2B pode ser observado que dentre as oito placas que constituem o *superpool* 5, a placa 33 foi positiva para o mesmo gene, como indicado pela seta verde. Os resultados obtidos no sistema linha/coluna podem ser observados na Figura 2C, onde o clone positivo para o gene CSN1S1 se encontra na linha A e coluna 2, indicado pela seta azul.

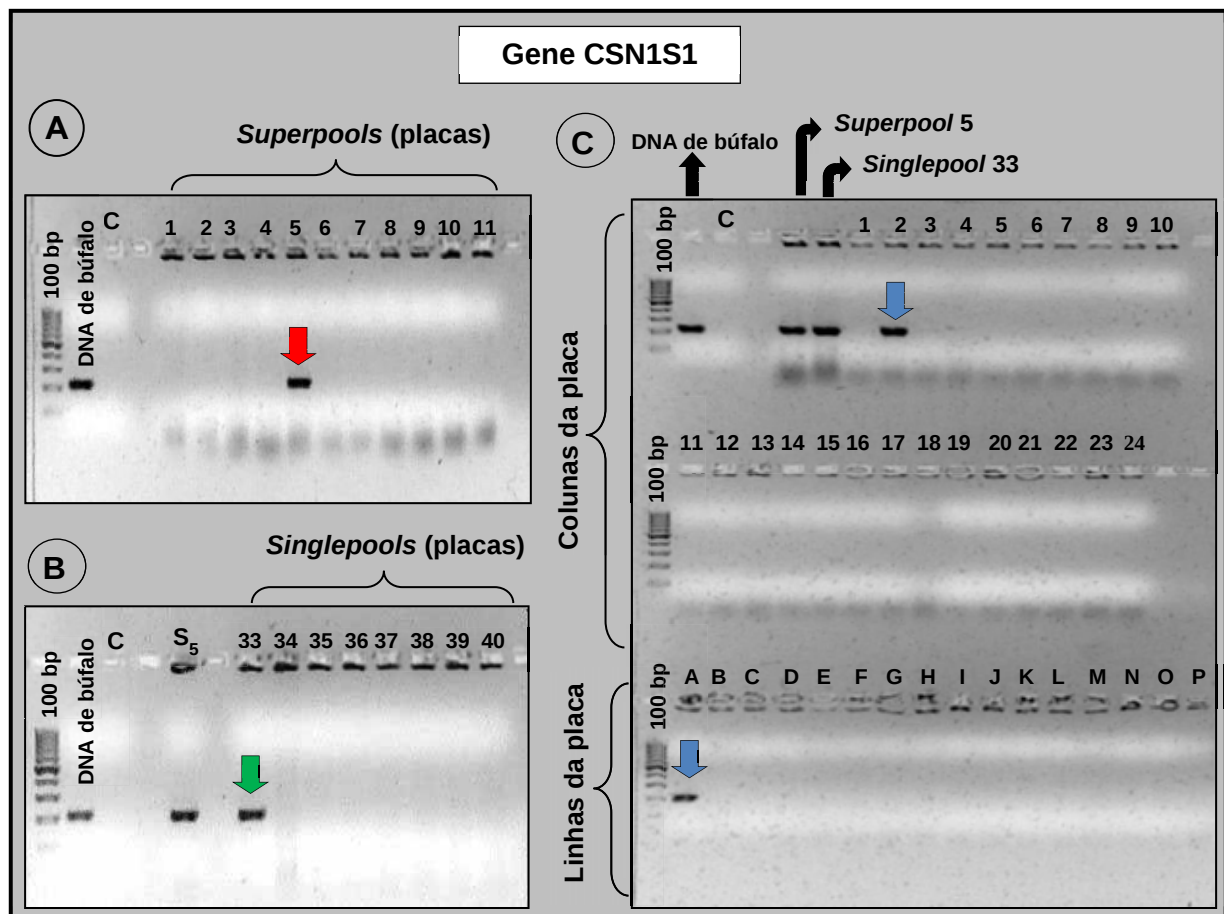


FIGURA 2. Representação de géis de agarose 2% corados com brometo de etídio, mostrando o resultado das reações de PCR realizadas para identificação de clones positivos para o gene CSN1S1. **(A)** Reação de PCR com 11 *superpools*. A seta vermelha indica que o *superpool* 5 é positivo. **(B)** Reação de PCR com os *singlepools* que compunham o *superpool* 5. A seta verde indica que o *singlepool* 33 é positivo. **(C)** Reação de PCR com o sistema linha/coluna do *singlepool* 33. **A-P:** linhas da placa 33; **1-24:** colunas da placa 33. As setas azuis indicam que o clone da linha A e coluna 2 é positivo. **DNA de búfalo:** controle positivo da reação; **C:** controle negativo da reação (sem DNA); **100bp:** marcador molecular (*Fermentas*).

Os resultados obtidos na seleção dos clones positivos para o gene CSN3 podem ser observados na Figura 3. Na Figura 3A, o *superpool* 2 mostrou-se positivo para o gene CSN3. Na Figura 3B, observou-se que dentre as oitos placas que compõem o *superpool* 2, a placa 9 foi positiva para tal gene, como indicado pela seta verde. Nas reações de PCR com os clones do sistema linha/coluna da placa 9, o clone localizado na linha L e coluna 8 foi positivo para o gene CSN3, como indicado pela seta azul (Figura 3C).

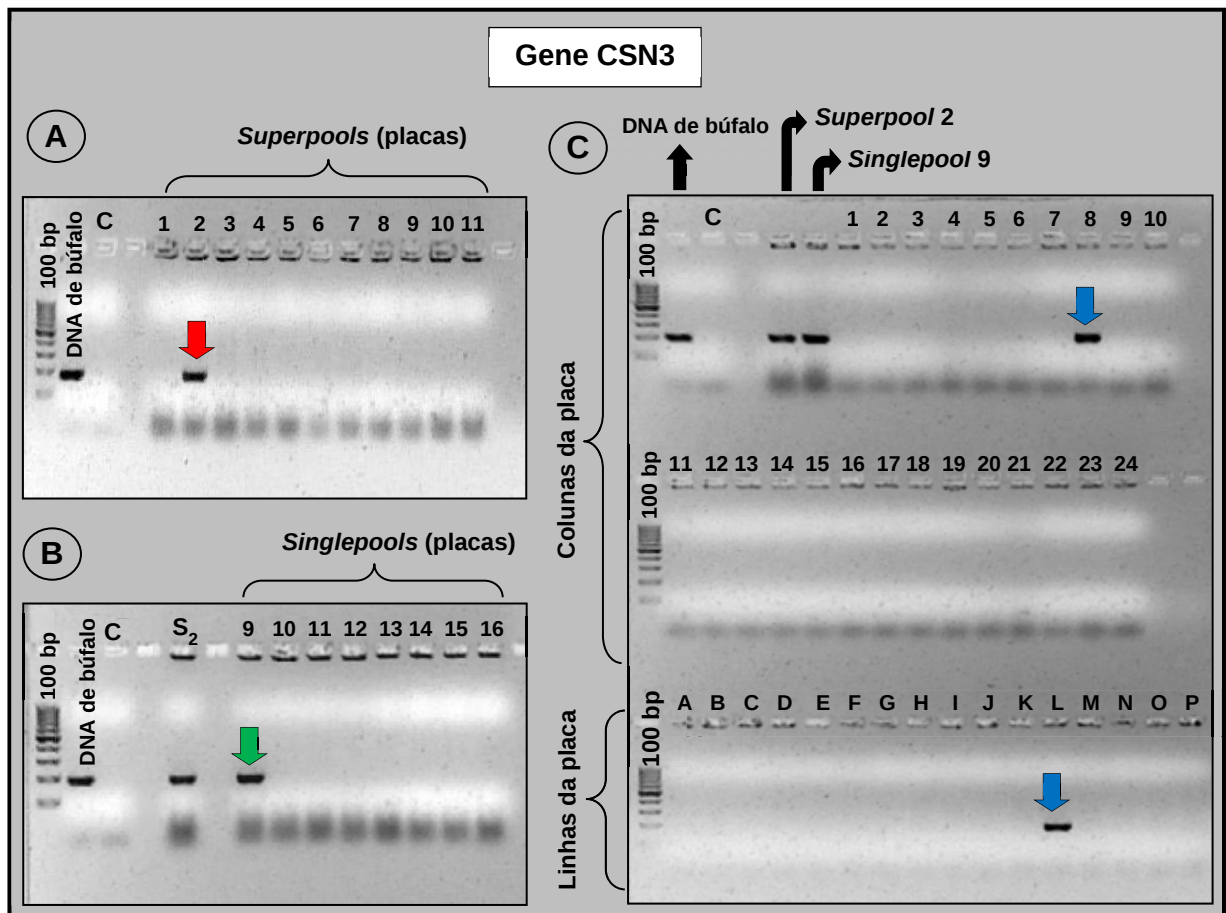


FIGURA 3. Representação de géis de agarose 2% corados com brometo de etídio, mostrando o resultado das reações de PCR realizadas para identificação de clones positivos para o gene CSN3. **(A)** Reação de PCR com 11 *superpools*. A seta vermelha indica que o *superpool 2* é positivo. **(B)** Reação de PCR com os *singlepools* que compunham o *superpool 2*. A seta verde indica que o *singlepool 9* é positivo. **(C)** Reação de PCR com o sistema linha/coluna do *singlepool 9*. **A-P:** linhas da placa 9; **1-24:** colunas da placa 9. As setas azuis indicam que o clone da linha L e coluna 8 é positivo. **DNA de búfalo:** controle positivo da reação; **C:** controle negativo da reação (sem DNA); **100bp:** marcador molecular (*Fermentas*).

A Figura 4 representa um gel de agarose contendo os resultados obtidos com o experimento de avaliação da integridade do DNA extraído dos dois clones selecionados da biblioteca. A partir da figura, pode ser observada a integridade do DNA extraído dos clones A/2 (gene CSN1S1) e L/8 (gene CSN3) devido à presença de DNA com alto peso molecular e à ausência de rastros no gel, o que indicaria DNA degradado.

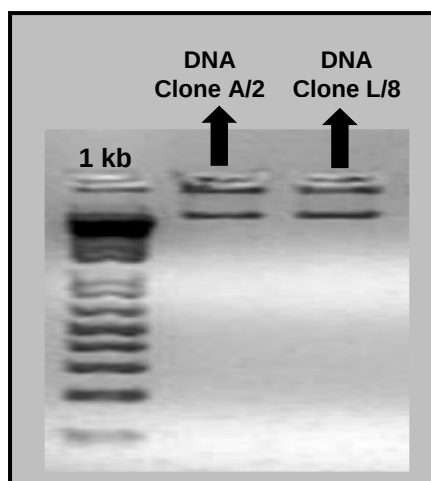


Figura 4. Representação de gel de agarose mostrando os resultados obtidos com o experimento de avaliação da integridade do DNA extraído dos clones positivos para os genes CSN1S1 e CSN3 de búfalo.

A Figura 5 apresenta os resultados obtidos nas reações de PCR realizadas para confirmação da presença dos genes CSN1S1 e CSN3 no DNA obtido a partir dos clones A/2 da placa 33 e L/8 da placa 9, respectivamente. Após a confirmação da presença dos genes, a concentração do DNA foi medida em espectrofotômetro (*Nanodrop ND1000TM – Thermo Scientific*), o qual forneceu uma concentração de 29,4 ng/μL para o DNA do clone A/2 (gene CSN1S1) e de 56,6 ng/μL para o L/8 (gene CSN3).

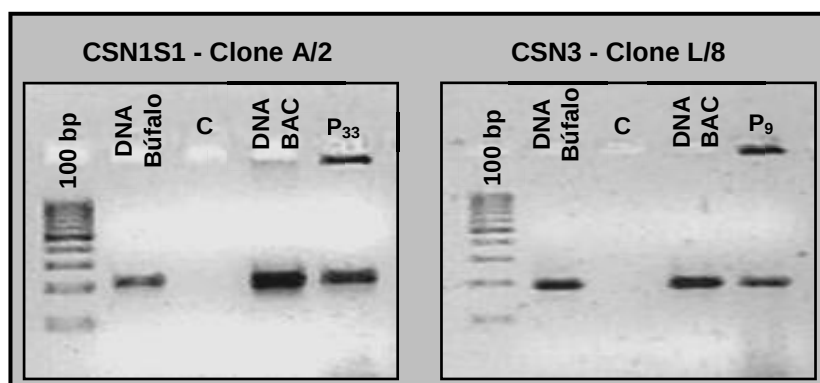


Figura 5. Representação de géis de agarose 2% corados com brometo de etídio, mostrando o resultado das reações de PCR realizadas para confirmar se os clones identificados eram positivos para os genes CSN1S1 e CSN3. **100 bp:** marcador molecular (*Fermentas*); **DNA Búfalo:** controle positivo da reação; **C:** controle negativo da reação (sem DNA); **DNA BAC:** DNA extraído do clone; **Px:** placa que continha o respectivo clone.

5.2 Análises das sequências de DNA obtidas por pirosequenciamento com as sequências correspondentes em bovino

O DNA dos clones foi clivado em fragmentos de aproximadamente 700 bp, os quais foram sequenciados resultando nas denominadas “reads”. Estas são as sequências de nucleotídeos que foram alinhadas por sobreposição para montagem da sequência de DNA consenso de cada clone, de acordo com a metodologia “*de novo assembly*”. No total, foram geradas 39.891 “reads” com tamanho médio de 693 bp para o clone A/2 e 42.323 “reads” de aproximadamente 698 bp para o clone L/8. A sobreposição das “reads” resultou em uma sequência de DNA consenso de 60.089 bp para o clone A/2. Já para o clone L/8 foram gerados duas sequências consenso, a do “contig” 1 de 46.443 bp e a do “contig” 2 de 6.383 bp. Como a montagem das sequências de DNA dos clones foi realizada sem o auxílio de uma genoma referência (metodologia “*de novo assembly*”), a ocorrência de duas sequências de DNA consenso (“contigs” 1 e 2) para o clone L/8, ao invés de uma única sequência, foi resultante de falhas no processo de montagem.

A sequência de DNA do clone A/2 apresentou um Q40+ de 99,87% e uma cobertura de 279,31X. Já as duas sequências de DNA do clone L/8 apresentaram um Q40+ de 99,66% e uma cobertura de 51,13X. O Q40+ trata-se de um parâmetro que mensura a porcentagem de bases na sequência consenso que apresentam um “phred-equivalent quality score” maior ou igual a 40. Um “score” de 40 equivale a uma taxa de erro de 10^{-4} ou acurácia de 99,99%. A cobertura, por sua vez, se refere ao número médio de “reads” em que um determinado nucleotídeo da sequência de DNA consenso encontra-se representado.

As análises realizadas com as sequências de DNA de cada clone e as versões de anotação do genoma bovino UMD_3.1 e Btau_4.6.1, por meio da ferramenta BLASTn, apresentaram os alinhamentos mais significativos com as sequências correspondentes do cromossomo 6 bovino (BTA6) (nº de acesso no *GenBank*: AC_000163.1 e NC_007304.5). Tais resultados estão de acordo com o esperado, uma vez que os genes das caseínas encontram-se localizados nesse cromossomo, o qual é homólogo ao cromossomo 7 bubalino (BBU7).

Na Tabela 2 estão descritos os resultados obtidos nos alinhamentos das sequências de DNA de cada clone com as sequências correspondentes do BTA6 das anotações UMD_3.1 e Btau_4.6.1.

Tabela 2. Informações obtidas a partir do alinhamento das sequências de DNA de cada clone com as sequências correspondentes do cromossomo 6 bovino (BTA6) de ambas as versões de anotação do genoma desta espécie (UMD_3.1 e Btau_4.6.1).

Sequência de DNA do Clone A/2					
Anotação	Cobertura	Identidade	“Score”	E-value	Intervalo_BTA6
UMD_3.1	93%	96%	21.322	0.0	87.118.876 - 87.180.554
Btau_4.6.1	93%	96%	21.329	0.0	88.218.368 - 88.279.554
Sequências de DNA do Clone L/8					
“Contig” 1					
Anotação	Cobertura	Identidade	“Score”	E-value	Intervalo_BTA6
UMD_3.1	97%	96%	27.563	0.0	87.389.748 - 87.434.432
Btau_4.6.1	98%	96%	28.393	0.0	88.480.224 - 88.525.648
“Contig” 2					
Anotação	Cobertura	Identidade	“Score”	E-value	Intervalo_BTA6
UMD_3.1	97%	95%	9.954	0.0	87.381.376 - 87.387.567
Btau_4.6.1	98%	96%	9.959	0.0	88.473.871 - 88.480.057

A partir das informações contidas na Tabela 2, observou-se que 93% da sequência de DNA do clone A/2 foi alinhada com a sequência correspondente do BTA6 de ambas as anotações do genoma bovino, com 96% de identidade entre os nucleotídeos. No alinhamento com ambas as anotações foram obtidos valores de “score” próximos (21.322 *versus* 21.329) e *e-value* igual à zero. Tanto para a anotação UMD_3.1 quanto para a Btau_4.6.1, as posições do intervalo do alinhamento da sequência de DNA do clone A/2 nas sequências correspondentes do BTA6 indicaram a presença do gene CSN1S1 bovino.

Com relação à sequência de DNA do “contig” 1 do clone L/8 foi observado que 97% desta sequência alinhou com a sequência correspondente do BTA6 da anotação UMD_3.1, com 95% de identidade entre os nucleotídeos. Já com a anotação Btau_4.6.1 foi verificada uma pequena diferença nos resultados do alinhamento, pois 98% da sequência de DNA desse “contig” se sobrepôs à sequência correspondente do BTA6 com 96% de identidade entre os nucleotídeos. Esses alinhamentos apresentaram valores de “score” relativamente diferentes (27.563 *versus* 28.393) e de “*e-value*” igual à zero. Em ambas as anotações do genoma bovino, as posições do intervalo do alinhamento da sequência de DNA do

“contig” 1 nas sequências correspondentes do BTA6 indicaram a presença do gene CSN3 bovino.

Também foi observado que 97% da sequência de DNA do “contig” 2 do clone L/8 alinhou com a sequência correspondente do BTA6 da anotação UMD_3.1, onde 95% dos nucleotídeos se mostraram idênticos entre ambas as sequências de DNA (Tabela 2). Já com a anotação Btau_4.6.1 foi observado que 98% da sequência de DNA do “contig” 2 se sobrepôs à sequência correspondente do BTA 6, com 96% de identidade entre nucleotídeos. No alinhamento com ambas as anotações, os valores de “score” foram próximos (9.954 *versus* 9.959) e de *e-value* igual à zero. As posições do intervalo do alinhamento da sequência de DNA do “contig” 2 nas sequências correspondentes do BTA6 de ambas as anotações do genoma bovino também indicaram a presença do gene CSN3 bovino.

5.3 Predição de genes

Foi identificado um gene inteiro na sequência de DNA do clone A/2 e um gene parcial na sequência de DNA do “contig” 1 do clone L/8. Nenhum gene foi predito para a sequência de DNA do “contig” 2 deste clone.

O gene predito para a sequência de DNA do clone A/2 apresentou 28.098 bp e nove éxons com tamanhos variando de 24 bp a 155 bp. Para identificar o gene que se tratava, a sequência codificante do gene predito foi alinhada com o banco de dados referência de RNAs (*RefSeq RNA*) do genoma bovino, por meio da ferramenta BLASTn, o que resultou na identificação do gene CSN1S1. Este alinhamento apresentou 63% de cobertura, 98% de identidade, “score” de 316 e “e-value” de 4e-85. Além disso, a sequência de aminoácidos predita foi alinhada com o banco de dados referência de proteínas (*RefSeqProtein*) do genoma bovino, com o auxílio da ferramenta BLASTp, resultando na identificação da proteína caseína alfa-s1. Tal alinhamento apresentou 69% de cobertura, 55% de identidade, “score” de 191 e *e-value* de 1e-60.

Para a sequência de DNA de “contig” 1 do clone L/8 foi predito um gene parcial de 14.890 bp com quatro éxons. O alinhamento da sequência codificante predita com o banco referência de RNAs de bovino (*RefSeqRNA*), por meio da

ferramenta BLASTn, resultou na predição do gene CSN3. Este alinhamento apresentou 5% de cobertura, 96% de identidade, “score” de 122 e *e-value* igual a 5e-26. O alinhamento da sequência de aminoácidos predita com o banco referência de proteínas de bovino (*RefSeqProtein*) resultou na identificação da proteína kapa-caseína. Neste alinhamento foi observada uma cobertura de 6% com 71% de identidade, “score” de 60.1 e *e-value* igual a 5e-10.

Como a predição de genes nas sequências de DNA dos clones A/2 e L/8 foi realizada com base em informações do genoma humano, os baixos valores de cobertura obtidos nos alinhamentos das sequências codificantes e de aminoácidos dos genes preditos com as sequências correspondentes em bovino estão de acordo o esperado, uma vez que humano e bovino são espécies filogeneticamente distantes.

Para confirmar os genes que foram preditos para sequências de DNA obtidas a partir dos clones, as mesmas foram alinhadas com as sequências dos genes correspondentes em bovino, com o auxílio da ferramenta *BLAST2seq*. O alinhamento da sequência de DNA do clone A/2 com a sequência do gene CSN1S1 bovino (17.536 bp) apresentou 98% de cobertura, 96% de identidade, “score” de 19.769 e *e-value* igual a 0.0. Esse gene alinhou no intervalo entre os nucleotídeos 21.381 e 38.388 da sequência de DNA desse clone. Tais resultados indicam que o gene CSN1S1 está inteiramente representado na sequência de DNA de búfalo obtida a partir do clone A/2. Já o alinhamento da sequência de DNA do clone L/8 com a sequência de do gene CSN3 bovino (13.062 bp) apresentou 37% de cobertura, 96% de identidade, “score” de 6.049 e *e-value* igual a 0.0. O intervalo do alinhamento ocorreu entre os nucleotídeos 1 e 3.714 da sequência de DNA de tal clone. Neste caso, a baixa porcentagem de cobertura obtida no alinhamento indicam que o gene CSN3 está parcialmente representado na sequência de DNA bubalino obtida a partir clone L/8.

5.4 Identificação de sequências de DNA conservadas entre búfalo e boi

O alinhamento da sequência de DNA de búfalo obtida a partir do clone A/2 com o “scaffold” do cromossomo 6 de boi (nº de acesso: NW_001495210.2), por

meio do programa *wgVISTA*, resultou na identificação dos genes CSN1S1 completo e CSN2 parcial, bem como de duas regiões intergênicas, uma localizada antes do gene CSN1S1 e a outra entre os genes CSN1S1 e CSN2, como ilustrado na Figura 6.

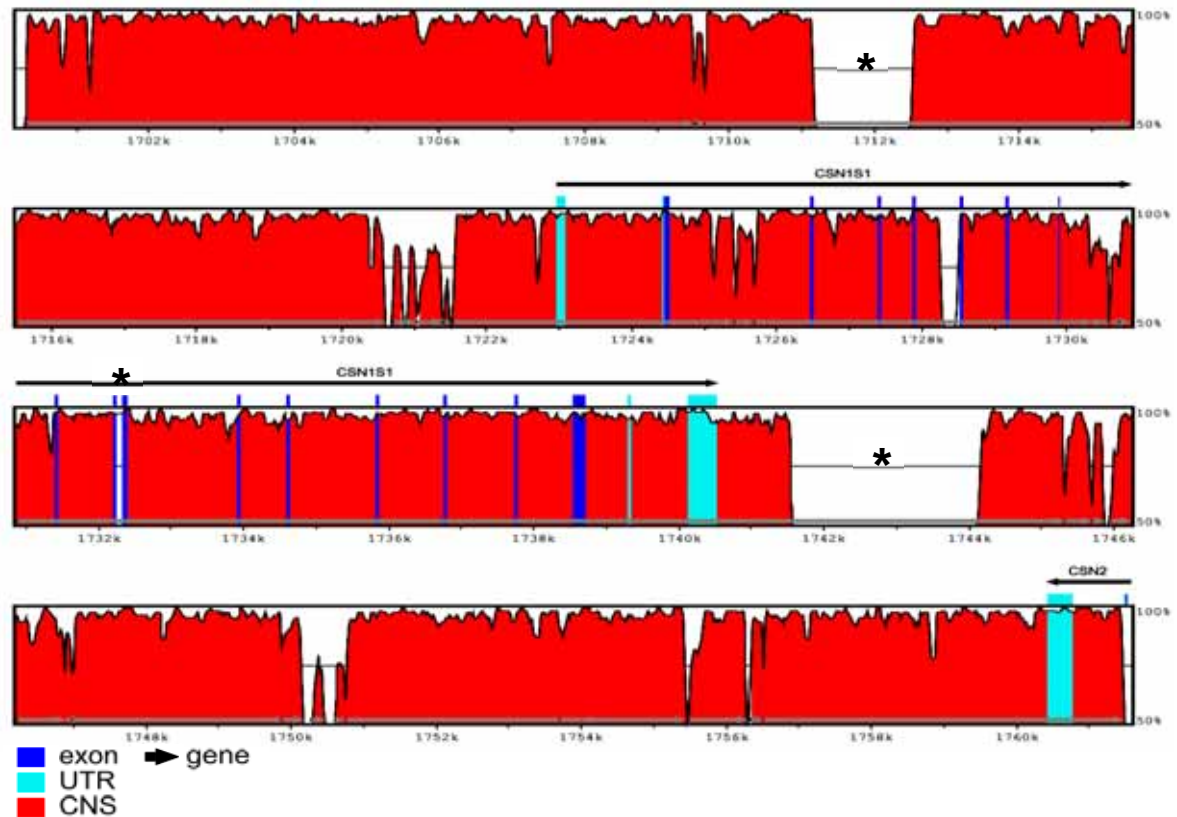


Figura 6. *wgVISTA* plot gerado a partir do alinhamento da sequência de DNA do clone A/2 com o “scaffold” do cromossomo 6 de boi (nº de acesso: NW_001495210.2), mostrando a conservação entre ambas as sequências, indicada em vermelho. As setas indicam a orientação dos genes CSN1S1 e CSN2, enquanto os blocos azul-escuro correspondem aos éxons e os blocos azul-claros às regiões 5’UTR e 3’UTR. O gene CSN1S1 apresentou conservação acima de 85% e o CSN2 acima de 98% entre búfalo e boi. O íntron 10 do gene CSN1S1 foi o único que não apresentou conservação entre ambas as espécies, como indicado pelo asterisco. A região intergênica localizada entre os genes CSN1S1 e CSN2 apresentou conservação acima de 72% entre ambas as espécies. Já a região intergênica anterior ao gene CSN1S1 mostrou uma conservação acima de 73% entre búfalo e boi. Os espaços indicados pelos asteriscos representam porções do DNA de boi que não apresentaram correspondência em búfalo.

Os resultados do alinhamento mostraram que o gene CSN1S1 de búfalo ocupa o intervalo entre os nucleotídeos 21.381 e 38.388 da sequência de DNA do clone A/2. Esse gene apresentou um tamanho total de 17.008 bp organizados em 19 éxons com tamanhos que variaram de 24 bp a 380 bp e 18 íntrons de 90 bp a 1.710

bp, totalizando 1.156 bp de regiões de éxons e 15.852 bp de regiões de introns. A maioria desses éxons mostrou o mesmo tamanho das sequências de DNA correspondentes em boi, com exceção dos éxons 1, 18 e 19 que apresentaram 77 bp, 43 bp e 380 bp ao invés dos respectivos 88 bp, 44 bp e 385 bp encontrados nessa espécie. Dentre os 18 íntrons do gene CSN1S1 de búfalo, somente os íntrons 7, 10 e 16 apresentaram o mesmo tamanho de seus correspondentes em boi.

No gene CSN1S1 de búfalo, o éxon 1 e os 12 primeiros nucleotídeos do éxon 2 compõem a região 5'UTR, enquanto o éxon 19 e grande parte do éxon 18 (42 bp dos 43 bp) constituem a região 3'UTR de tal gene (Figura 6). Em mamíferos, é comum a região 5'UTR dos genes das caseínas ser formada pelo éxon 1. No entanto, somente nos genes CSN1S1, CSN1S2 e CSN2 foi observado que os 12 primeiros nucleotídeos do éxon 2 também constituem essa região. Outra característica comum dos genes das caseínas decorre da região 3'UTR ser formada pelo último e parte do penúltimo éxon na maioria das espécies de mamíferos (RIJENKLS, 2002).

A Figura 7 apresenta as informações sobre a estrutura do gene CSN1S1 de búfalo e boi quanto ao número e tamanho dos éxons e íntrons. As informações sobre o gene CSN1S1 de boi pertencem à versão de anotação Btau_4.6.1 e foram obtidas a partir do banco de dados *GenBank* (nº de acesso: NC_007304.5).

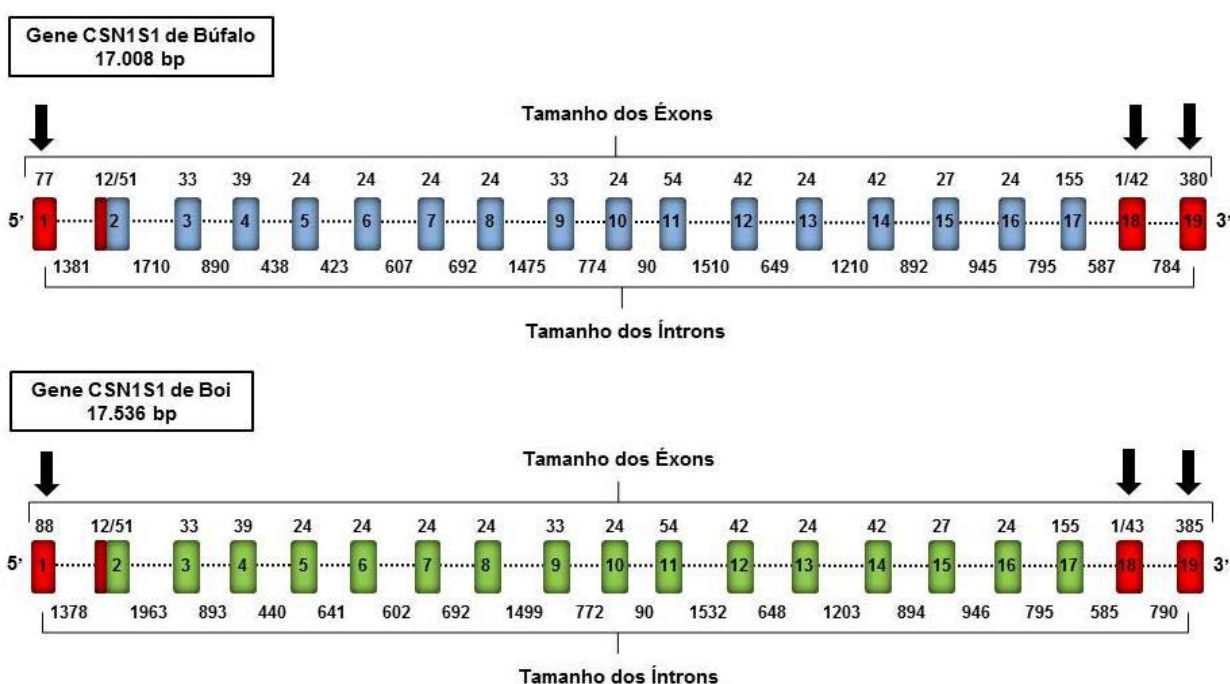


Figura 7. Representação esquemática da estrutura do gene CSN1S1 de búfalo e boi quanto ao número e tamanho dos éxons e íntrons. **Gene CSN1S1 de Búfalo:** Informações fornecidas pelo programa *wgVISTA* a partir do alinhamento da sequência de DNA do clone A/2 com o “scaffold” do cromossomo 6 de boi (NW_001495210.2). **Gene CSN1S1 Boi:** Informações obtidas a partir do banco de dados *Genbank* (nº de acesso: NC_007304.5). As setas indicam os éxons do gene CSN1S1 que apresentam tamanhos diferentes entre búfalo e boi.

Em função da característica estrutural do gene CSN1S1 de búfalo, a sequência codificante deste gene é formada pelos 51 nucleotídeos restantes do éxon 2, os éxons 3 a 17 e o primeiro nucleotídeo do éxon 18, totalizando 645 bp. A tradução dessa sequência, com o auxílio do programa *EMBOSS Transeq* (EMBL-EBI), resultou em uma sequência de 214 aminoácidos. Ambas as sequências, codificante e de aminoácidos, apresentaram o mesmo tamanho das sequências homólogas em boi.

No Quadro 1 estão descritas as porcentagens de conservação dos éxons e íntrons do gene CSN1S1 de búfalo com as sequências correspondentes em boi.

Quadro 1. Porcentagens de conservação de cada éxon e íntron do gene CSN1S1 entre búfalo e boi. Informações fornecidas pelo programa *wgVISTA* a partir do alinhamento da sequência de DNA do clone A/2 com o “scaffold” do cromossomo 6 de boi (NW_001495210.2).

Gene CSN1S1_Búfalo/Boi			
Éxon	Conservação	Íntron	Conservação
1	98,9%	1	96,9%
2	100%	2	94,0%
3	97,0%	3	96,0%
4	94,9%	4	96,6%
5	100%	5	97,8%
6	95,8%	6	97,0%
7	100%	7	96,7%
8	100%	8	86,0%
9	100%	9	97,4%
10	95,8%	10	0,00%
11	100%	11	95,2%
12	95,2%	12	96,0%
13	87,5%	13	96,4%
14	97,6%	14	97,7%
15	100%	15	97,8%
16	95,8%	16	97,1%
17	97,4%	17	96,9%
18	96,3%	18	96,2%
19	97,9%		

De acordo com as informações descritas no Quadro 1, sete dos 19 éxons do gene CSN1S1 foram 100% conservados entre búfalo e boi, enquanto que os demais éxons apresentaram conservação variando de 87,5% a 98,9% entre ambas as espécies. Como a maioria dos éxons do gene CSN1S1 de búfalo e boi apresentam o mesmo tamanho, as diferenças observadas entre essas sequências de DNA resultaram principalmente da ocorrência de substituições de nucleotídeos. O

éxon 13, por exemplo, apresentou a menor porcentagem de conservação (87,5%) entre essas espécies devido à presença de três substituições de nucleotídeos consecutivas na sequência de búfalo em relação à de boi g. [11.116A>T; 11.117G>T; 11.118T>G] (Quadro 1).

Os íntrons do gene CSN1S1 mostraram uma conservação acima de 86,0% entre búfalo e boi, com exceção do íntron 10 que não foi conservado entre ambas às espécies, como pode ser observado na Figura 6 e no Quadro 1. Esses resultados se devem à ocorrência de substituições de nucleotídeos, inserções, deleções e/ou duplicações que alteraram o tamanho e a composição dos íntrons do gene CSN1S1 em búfalo e boi.

A partir dos resultados obtidos, observou-se que o gene CSN1S1 é conservado entre búfalo e boi, uma vez que o mesmo apresentou a mesma organização estrutural (19 éxons e 18 íntrons) e regiões com conservação acima de 70% entre tais espécies. Essa conservação já era esperada, devido à próxima relação filogenética existente entre búfalo e boi, ambos pertencentes à família Bovidae. Somado a isso, estudos comparativos evidenciaram que o número e o tamanho dos éxons desse gene são similares entre as espécies de mamíferos (RIJNKELS, 2002).

Com relação ao gene CSN2 parcial de búfalo, os resultados do alinhamento mostraram que o mesmo está localizado no intervalo entre os nucleotídeos 59.046 e 60.089 da sequência de DNA do clone A/2. A porção do gene identificada incluiu o éxon 9 completo e parte do íntron 8, os quais apresentaram 324 bp e 720 bp, respectivamente. O éxon 9 exibiu um nucleotídeo a mais em relação ao éxon correspondente em boi, de 323 bp. Essa diferença não reflete sobre a sequência codificante do gene CSN2, uma vez que o éxon 9 faz parte da região 3'UTR em ambas as espécies. Como a sequência gerada para o íntron 8 foi parcial, não foi possível fazer comparação do número de nucleotídeos dessa região com a correspondente em bovino. A conservação observada entre o gene CSN2 parcial de búfalo e boi foi de 98,8% para o éxon 9 e 98,5% para íntron 8 parcial (Figura 6).

Entre os genes CSN1S1 e CSN2 parcial foi identificada uma região intergênica de 20.669 bp na sequência de DNA do clone A/2 que apresentou conservação acima de 73% com a correspondente em boi. Dentro dessa região foi observada a

ocorrência de um segmento de DNA bovino de 2.573 bp que não mostrou correspondência em búfalo. A região intergênica encontrada antes do gene CSN1S1 de búfalo apresentou 21.369 bp e conservação acima de 72% com aquela correspondente em boi. Ao longo dessa região também foi identificado um segmento de DNA de boi com tamanho de 1.344 bp que não apresentou correspondência em búfalo (Figura 6).

As sequências de DNA dos “contigs” 1 e 2, obtidas a partir do clone L/8, também foram alinhadas com o “scaffold” do cromossomo 6 de boi (NW_001495210.2), por meio do programa *wgVISTA*, o qual auxiliou na identificação da conservação do gene CSN3 parcial e de uma região intergênica entre búfalo e boi (Figura 8).

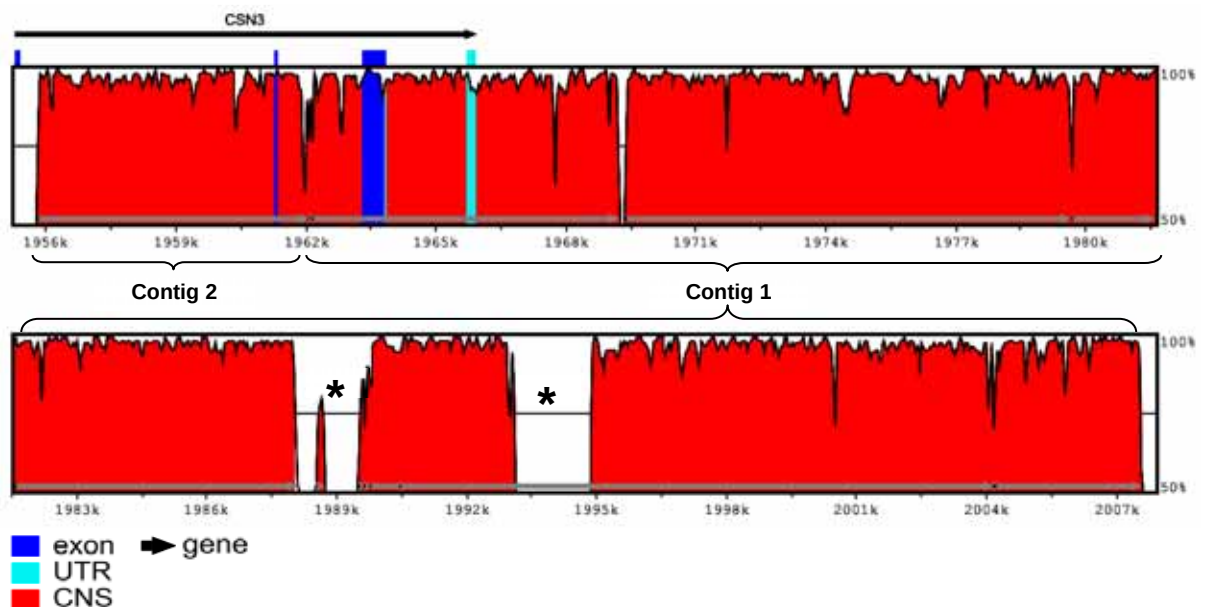


Figura 8. *wgVISTA* plot gerado a partir do alinhamento das sequências de DNA dos “contigs” 1 e 2 do clone L/8 com o “scaffold” do cromossomo 6 de boi (nº de acesso: NW_001495210.2), mostrando a conservação entre as sequências de DNA de búfalo e boi, indicada em vermelho. A seta indica a orientação do gene CSN3, enquanto os blocos azul-escuros representam o éxons e os blocos azul-claros a região 3'UTR. O gene CSN3 parcial de búfalo apresentou uma conservação acima de 94% com o correspondente em boi. A região intergênica posterior ao gene CSN3 mostrou uma conservação acima de 75% entre ambas as espécies. Os espaços em indicados pelos asteriscos representam porções do DNA de boi que não apresentaram correspondência em búfalo.

Os resultados do alinhamento mostraram que a sequência de DNA inteira do “contig” 2 abrange o éxon 3 e parte dos íntrons 2 e 3 do gene CSN3, como ilustrado na Figura 8. Esse éxon faz parte da região codificante do gene CSN3 de búfalo e

apresentou 33 bp, o mesmo tamanho observado em boi. Já os íntrons 2 e 3 parciais exibiram tamanhos de 5.496 bp e 854 bp, respectivamente. Como as sequências geradas para os íntrons 2 e 3 foram parciais, não foi possível fazer comparações do número de nucleotídeos dessas regiões com as correspondentes em boi. A conservação dessas regiões do gene CSN3 entre búfalo e boi foi de 100% para o éxon 3, 95,8% para o íntron 2 parcial e 97,0% para o íntron 3 parcial (Figura 8).

Parte do gene CSN3 também foi identificada na sequência de DNA do “contig” 1, onde ocupa o intervalo entre os nucleotídeos 1 e 3.714. Como pode ser observado na Figura 7, esse “contig” inclui o íntron 3 parcial (1.163 bp) e a região completa do íntron 4 (1.861 bp), do éxon 4 (517 bp) e do éxon 5 (173 bp). Os resultados do alinhamento revelaram que o éxon 5 e os 34 últimos nucleotídeos do éxon 4 constituem a região 3’UTR do gene CSN3 de búfalo. Já os 483 nucleotídeos restantes do éxon 4 compõem a maior parte da sequência codificante deste gene (Figura 8).

Análises comparativas mostraram que os éxons 4 e 5 do gene CSN3 de búfalo apresentam o mesmo tamanho dos correspondentes em boi. Já o íntron 4 possui três nucleotídeos a mais em relação ao mesmo íntron do gene CSN3 dessa espécie. Como a sequência gerada para o íntron 3 foi parcial, não foi possível fazer comparação do número de nucleotídeos dessa região com a correspondente em boi. As porções do gene CSN3 de búfalo identificadas no “contig” 1 apresentaram as seguintes porcentagens de conservação com as correspondentes em boi: 94,6% para o íntron 3 parcial, 96,5% para o éxon 4, 96,9% para o íntron 4 e 94,8% para o éxon 5 (Figura 8).

Os resultados do alinhamento com o “scaffold” do cromossomo 6 de boi (NW_001495210.2) mostraram que há um intervalo de 166 bp entre as sequências de DNA dos dois “contigs” do clone L/8 (Figura 8).

Além do gene CSN3 parcial, foi identificada uma região intergênica de 42.729 bp na sequência de DNA do “contig” 1, que apresentou conservação acima de 75% com aquela correspondente em boi. Dentro dessa região foram encontrados dois segmentos de DNA de boi, um de 833 bp e outro de 1.792 bp, que não apresentaram correspondência em búfalo (Figura 8).

Considerando que elementos regulatórios são conservados principalmente entre espécies próximas filogeneticamente, os resultados de conservação obtidos nessa análise podem nortear estudos de identificação de elementos regulatórios que influenciam traços quantitativos e qualitativos do leite bubalino e bovino.

5.5 Identificação de sequências repetitivas

Os resultados obtidos com o programa *RepeatMasker* mostraram que a sequência de DNA do clone A/2 apresenta 37,48% de conteúdo CG e 103 sequências repetitivas abrangendo 45,55% da sequência de DNA (27,3 Kb). Deste total de sequências repetitivas, foram identificados 91 elementos transponíveis (48 SINEs, 25 LINEs, 13 LTRs e cinco transposons), nove repetições em “tandem”, dois pequenos RNAs e uma sequência de DNA de baixa complexidade. Os elementos retrotransposons foram os mais representativos compreendendo 43,83% da sequência, sendo o número de retrotransposons não-LTR (31,63%) superior ao número de retrotransposons LTR (12,20%). A Tabela 3 apresenta as sequências repetitivas identificadas na sequência de DNA do clone A/2 considerando o tipo e o número de repetições, assim como o tamanho da sequência de DNA representada pelas mesmas em pares de base (bp) e porcentagem (%).

Tabela 3. Tipo e número de sequências repetitivas identificadas na sequência de DNA do clone A/2, assim como o tamanho da sequência de DNA representada por estas repetições em pares de base (bp) e porcentagem (%).

Tipos de sequência repetitiva	Número	Tamanho da sequência	
		(bp)	(%)
Retrotransposon não-LTR (SINE)	48	5.438	9,05
Retrotransposon não-LTR (LINE)	25	13.570	22,58
Retrotransposon LTR	13	7.220	12,20
Transposon	5	744	1,24
Pequenos RNAs	2	261	0,43
Repetições em “tandem”	9	309	0,51
DNA de baixa complexidade	1	91	0,15

A análise da presença de sequências repetitivas no gene CSN1S1 de búfalo revelou 31,96% de conteúdo CG e 18 sequências repetitivas, todas inseridas em regiões de íntrons. As sequências repetitivas identificadas incluíram nove SINEs,

duas LINEs, um transposon, cinco repetições em “tandem” e uma sequência de DNA de baixa complexidade, as quais compreenderam 11,99% do tamanho total do gene.

Análises comparativas com as sequências repetitivas contidas no gene CSN1S1 de outros bovídeos (boi, iaque, cabra e ovelha), camundongo e humano revelaram duas sequências repetitivas búfalo específicas, incluindo uma repetição em “tandem” (GT) no íntron 15 e uma sequência de DNA de baixa complexidade (rico em adenina) no íntron 2.

As análises mostraram também a ocorrência de sete sequências repetitivas comuns em todas as espécies de bovídeos, sendo: quatro SINEs (BOV-A2, MAR1c_Mdo, MIRb e MIR3), uma LINE (BovB) e duas repetições em “tandem” (AATATTA e ATGA). O elemento transposon Charlie16 foi o único comum entre bovídeos e humano. A partir da Figura 9 é possível observar a estrutura do gene CSN1S1 de búfalo com suas respectivas sequências repetitivas, incluindo aquelas que são búfalo específicas ou comuns com outros bovídeos.

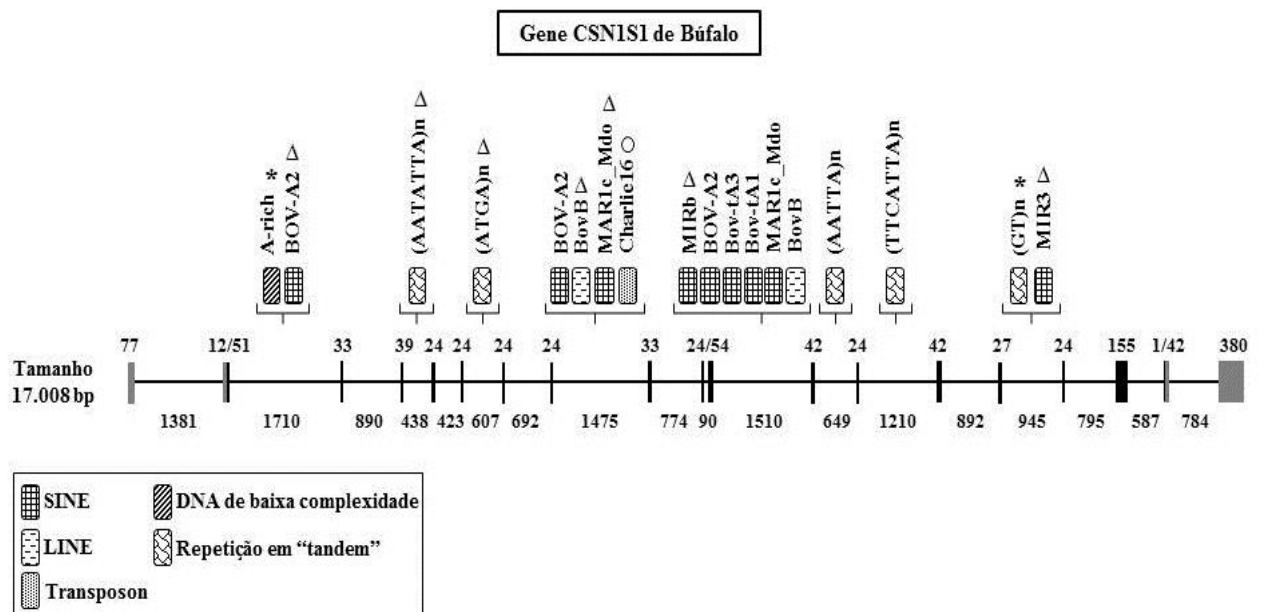


Figura 9. Representação esquemática da estrutura do gene CSN1S1 de búfalo identificado na sequência de DNA do clone A/2, com suas respectivas sequências repetitivas. Os éxons codificantes e não-codificantes de proteína estão representados pelos retângulos pretos e cinzas, respectivamente. Os números acima e abaixo dos retângulos mostram os respectivos tamanhos dos éxons e íntrons. As sequências repetitivas búfalo específicas estão indicadas pelos asteriscos, enquanto que aquelas comuns com outros bovídeos (boi, iaque, cabra e ovelha) estão destacadas pelos triângulos. O círculo indica a sequência repetitiva do gene CSN1S1 comum entre bovídeos e humano.

Patel, Singh e Suryanarayana (2014) analisaram a região promotora do gene CSN1S1 de Murrah e identificaram os seguintes elementos do tipo retrotransposons: um L1_BT (LINE/família L1), um Bov-tA1 e Bov-tA2 (SINE/família BovA), um ART2A (SINE/família RTE-BovB) e dois BovB (LINE/família RTE-BovB). Investigando a região correspondente na sequência de DNA do clone A/2 contendo o gene CSN1S1, observou-se a presença de todos estes elementos, exceto o ART2A.

Na sequência parcial do gene CSN2 de búfalo não foi identificada nenhuma sequência repetitiva.

Com relação aos resultados obtidos com as sequências de DNA do clone L/8, foi observada a ocorrência de 36,47% de conteúdo CG e 120 sequências repetitivas representando 53,91% do tamanho total das mesmas (28,5 Kb). As sequências repetitivas identificadas incluíram 97 elementos transponíveis (51 SINEs, 34 LINEs, nove LTRs e três transposons), 15 repetições em “tandem”, três sequências de DNA satélite, três pequenos RNAs e duas sequências de DNA de baixa complexidade. Assim como foi observado para o clone A/2, os retrotransposons foram os elementos mais representativos abrangendo 44,98% das sequências de DNA do clone L/8, onde o número de retrotransposons não-LTR (40,91%) foi superior ao número de retrotransposons LTR (4,07%). A Tabela 4 apresenta os tipos e o número de sequências repetitivas identificadas nas sequências de DNA do clone L/8, como também o tamanho da sequência de DNA representada por estas repetições em pares de base (bp) e porcentagem (%).

Tabela 4. Tipo e número de sequências repetitivas identificadas nas sequências de DNA do clone L/8, assim como o tamanho da sequência de DNA representada por estas repetições em pares de base (bp) e porcentagem (%).

Tipo de sequência repetitiva	Número	Tamanho da sequência (bp)	sequência (%)
Retrotransposon não-LTR (SINE)	51	6.355	12,03
Retrotransposon não-LTR (LINE)	34	15.256	28,88
Retrotransposon LTR	9	2.150	4,07
Transposon	3	638	1,21
Pequenos RNAs	3	401	0,76
DNA satélite	3	2414	4,57
Repetições em “tandem”	15	1.121	2,12
DNA de baixa complexidade	2	146	0,27

As regiões parciais do gene CSN3 de búfalo apresentaram 36,47% de conteúdo CG e 21 sequências repetitivas localizados em regiões de íntrons. Deste total de sequências repetitivas, foram encontradas 11 SINEs, cinco LINEs, três repetições em “tandem” e duas sequência de DNA de baixa complexidade, representando 42,45% do tamanho total do gene parcial. As análises comparativas realizadas com as sequências de DNA correspondentes em outros bovídeos (boi, iaque, cabra e ovelha), camundongo e humano revelaram três sequências repetitivas búfalo específicas, incluindo duas SINEs, uma presente no íntron 2 (Bov-tA1) e a outra no íntron 4 (MIRb), e uma sequência de DNA de baixa complexidade (rica em adenina) no íntron 3. Além disso, foram identificadas sete sequências repetitivas comuns entre búfalo e aquelas espécies de bovídeos, sendo elas: quatro LINEs (L1_Art, L1-1_Ttr e dois BovB) e três SINEs (Bov-tA2, BOV-A2 e SINE1_BB). A Figura 10 apresenta a estrutura do gene CSN3 parcial de búfalo com suas respectivas sequências repetitivas, incluindo aquelas que são específicas de búfalo ou comuns com outros bovídeos.

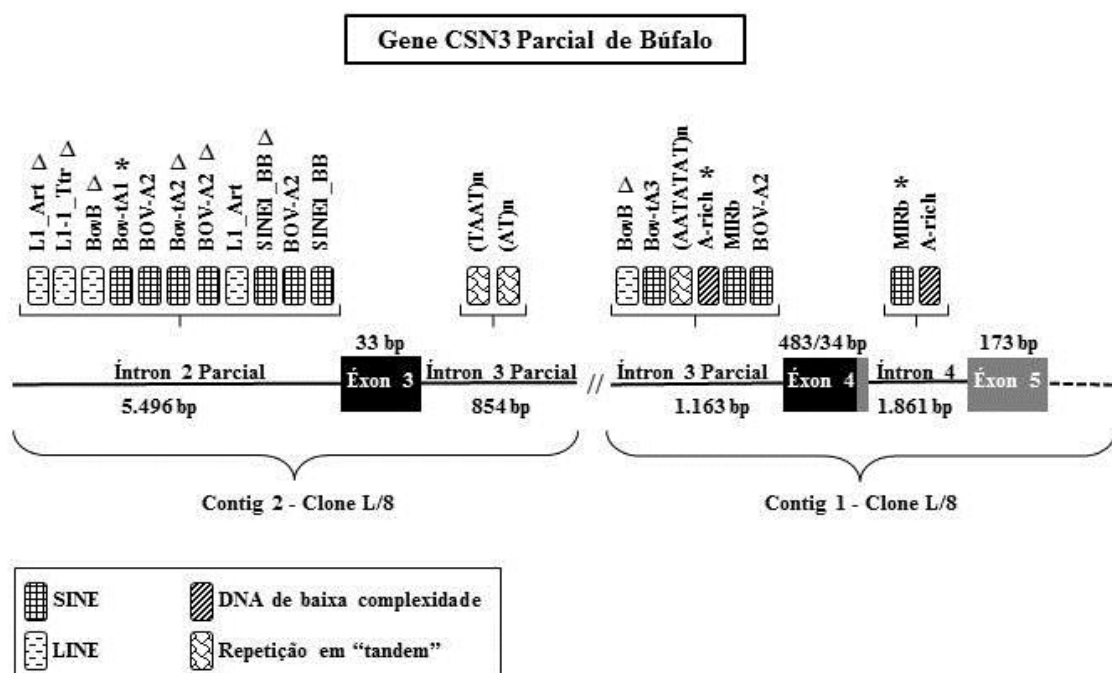


Figura 10. Representação esquemática da estrutura do gene CSN3 parcial de búfalo identificado nas sequências de DNA do clone L/8, com suas respectivas sequências repetitivas. Os éxons codificantes e não-codificantes de proteína estão representados pelos retângulos pretos e cinzas, respectivamente. Os números acima e entre os retângulos mostram os respectivos tamanhos dos éxons e íntrons. As sequências repetitivas búfalo específicas estão indicadas pelos asteriscos, enquanto que aquelas comuns com outros bovídeos (boi, iaque, cabra e ovelha) estão destacadas pelos triângulos.

A partir dos resultados expostos, foi possível observar que os retrotransposons foram as sequências repetitivas mais representativas tanto na sequência de DNA do clone A/2 (43,83%) quanto nas sequências do clone L/8 (44,98%). Estes resultados estão de acordo com o observado na literatura, onde essa categoria de repetição também abrange de 40 a 50% dos genomas de humano, camundongo e boi (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001; MOUSE GENOME SEQUENCING CONSORTIUM, 2002; JURKA et al., 2007; ADELSON et al., 2009).

Os resultados obtidos no presente estudo também mostraram que sequências repetitivas de diversas categorias (elementos transponíveis, microssatélites e DNA de baixa complexidade) têm contribuído para a variabilidade dos genes CSN1S1 e CSN3 (parcial) de búfalo. Em ambos os genes, foram identificadas sequências repetitivas somente em regiões de íntrons, onde os retrotransposons foram predominantes. As análises comparativas possibilitaram a identificação de sequências repetitivas búfalo específicas ou comuns com outros bovídeos tanto no gene CSN1S1 quanto no gene CSN3 parcial bubalino.

5.6 Análise comparativa com outros bovídeos

O alinhamento múltiplo da sequência codificante do gene CSN1S1 de búfalo com as sequências correspondentes de outros bovídeos, por meio do programa *Clustal Omega*, apresentou identidade de 98% com boi (*Bos taurus*) (nº de acesso no *GenBank*: NM_181029.2), 93% com cabra (*Capra hircus*) (nº de acesso no *GenBank*: NM_001285695.1) e de 90% com ovelha (*Ovis aries*) (nº de acesso no *GenBank*: NM_001009795.1). Esse alinhamento se encontra ilustrado na Figura 11, onde é possível observar as alterações de nucleotídeos nas sequências codificantes do gene CSN1S1 de boi, cabra e ovelha, assumindo a sequência correspondente em búfalo como referência.

```

O. Aries    ATGAAACTTCTCATCCTTACCTGTCTTGTGGCNGTTGCTCTTGCCAGGCCTAAACATCCTATCAAGCACCAAGGACT
C. Hircus   ATGAAACTTCTCATCCTTACCTGTCTTGTGGCTGTTGCTCTTGCCAGGCCTAAACATCCTATCAATCACCAGGACT
B. bubalis  ATGAAACTTCTCATCCTTACCTGTCTTGTGGCTGTTGCTCTTGCCAGGCCTAAACAGCCTATCAAGCACCAAGGACT
B. taurus   ATGAAACTTCTCATCCTTACCTGTCTTGTGGCTGTTGCTCTTGCCAGGCCTAAACATCCTATCAAGCACCAAGGACT
*****
O. Aries    CTCTCCAGAAGTCCCTCAATGAAAATTTACTCAGGTTTGTGTGGCGCCTTTTCCAGAAGTGTTTGAAGGAGAACA
C. Hircus   CTCTCCAGAAGTCCCTCAATGAAAATTTACTCAGGTTTGTGTGGCGCCTTTTCCAGAAGTGTTTGAAGGAGAACA
B. bubalis  CCCTCAAGGAGTCCCTCAATGAAAATTTACTCAGGTTTGTGTGGCGCCTTTTCCAGAAGTGTTTGAAGGAGAAGG
B. taurus   CCCTCAAGGAGTCCCTCAATGAAAATTTACTCAGGTTTGTGTGGCGCCTTTTCCAGAAGTGTTTGAAGGAGAAGG
* * * * *
O. aries    TCAATGAAGTGAAGTATTTGGGAGTGAATCAATGAGGATCAAGCCATGGAAGATGCTAAGCAAATGAAGCT
C. hircus   TCAATGAAGTGAAGTATTTGGGAGTGAATCAACTGAGGATCAAGCCATGGAAGATGCTAAGCAAATGAAGCT
B. bubalis  TCAATGAAGTGAAGTATTTGGGAGTGAATCAACTGAGGATCAAGCCATGGAAGATATTAAGCAAATGAAGCT
B. taurus   TCAATGAAGTGAAGTATTTGGGAGTGAATCAACTGAGGATCAAGCCATGGAAGATATTAAGCAAATGAAGCT
*****
O. aries    GGAAGCATTTCGTCAAGTGAAGAAATTGTTCCCAATAGTGCTGAGCAGAAGTACATTCAAAGGAAGATGTGCCCTC
C. hircus   GGAAGCATTTCGTCAAGTGAAGAAATTGTTCCCAATAGTGCTGAG--AAGTACATTCAAAGGAAGATGTGCCCTC
B. bubalis  GAAAGCATTTCGTCAAGTGAAGAAATTGTTCCCAATAGTGTTGAGCAGAAGCACATTCAAAGGAAGATGTGCCCTC
B. taurus   GAAAGCATTTCGTCAAGTGAAGAAATTGTTCCCAATAGTGTTGAGCAGAAGCACATTCAAAGGAAGATGTGCCCTC
* * * * *
O. aries    TGAGCGTTACTTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAACTGCCCCAGCTGGAAATTGTTCCCA
C. hircus   TGAGCGTTACTTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAACTGCCCCAGCTGGAAATTGTTCCCA
B. bubalis  TGAGCGTTACTTGGGTTATCTGGAACAGCTTCTTAGACTGAAAAAATACAACTGCCCCAGCTGGAAATTGTTCCCA
B. taurus   TGAGCGTTACTTGGGTTATCTGGAACAGCTTCTCAGACTGAAAAAATACAACTGCCCCAGCTGGAAATTGTTCCCA
*****
O. aries    AAAGTGTCTGAGGAACAACCTTCAATAGTATGAAAGAGGGAAACCTTGCCCAACAGAAACAGCCTATGATAGCAGTGA
C. hircus   ATAGTGTCTGAGGAACAACCTTCAACAGTATGAAAGAGGGAAACCTTGCCCAACAGAAACAGCCTATGATAGCAGTGA
B. bubalis  ATTTGGCTGAGGAACAACCTTCAACAGTATGAAAGAGGGAAATCCATGCCCAACAGAAAGAACCTATGATAGGAGTGA
B. taurus   ATAGTGTCTGAGGAACAACCTTCAACAGTATGAAAGAGGGAAATCCATGCCCAACAGAAAGAACCTATGATAGGAGTGA
* * * * *
O. aries    -----TCAGCTTTTCAGACAATTCTACCAGCTGGACGCCTATCCATCTGGTGCCTGGTA
C. hircus   CAGGAAGTGGCCTACTTCTACCTCAGCTTTTCAGACAATTCTACCAGCTGGACGCCTATCCATCTGGTGCCTGGTA
B. bubalis  CAGGAAGTGGCCTACTTCTACCTCAGCTTTTCAGACAATTCTACCAGCTGGATGCCTATCCATCTGGTGCCTGGTA
B. taurus   CAGGAAGTGGCCTACTTCTACCTCAGCTTTTCAGACAATTCTACCAGCTGGATGCCTATCCATCTGGTGCCTGGTA
* * * * *
O. aries    TTACCTTCCACTAGGCACACAATACACTGATGCCCCCTCATTCTCTGACATCCCTAATCCCATTGGCTCTGAGAACA
C. hircus   TTACCTTCCACTAGGCACACAATACACTGATGCCCCCTCATTCTCTGACATCCCTAATCCCATTGGCTCTGAGAACA
B. bubalis  TTACGTTCCACTAGGCACGCAATACCCTGATGCCCCATCATTCTCTGACATCCCTAATCCCATTGGCTCTGAGAACA
B. taurus   TTACGTTCCACTAGGCACACAATACACTGATGCCCCATCATTCTCTGACATCCCTAATCCCATTGGCTCTGAGAACA
* * * * *
O. aries    GTGGAAGATTACTATGCCACTGTGGTGA 621
C. hircus   GTGGAAGACTACTATGCCACTGTGGTGA 642
B. bubalis  GTGGAAGACTACTATGCCACTGTGGTGA 645
B. taurus   GTGGAAGACTACTATGCCACTGTGGTGA 645
* * * * *

```

Figura 11. Alinhamento múltiplo da sequência codificante do gene CSN1S1 de búfalo com as sequências correspondentes em boi, cabra e ovelha, realizado com o auxílio da ferramenta *Clustal Omega*. As alterações de nucleotídeos em relação à sequência codificante do gene CSN1S1 de búfalo estão destacadas em cinza.

Como pode ser observado na Figura 11, foi obtida uma identidade de 98% com a sequência codificante do gene CSN1S1 de boi devido à presença de 16 substituições de nucleotídeos ao longo da mesma. Com relação à sequência de cabra, a identidade de 93% resultou da ocorrência de três deleções e 40 substituições de nucleotídeos, enquanto que para a sequência de ovelha, a presença de 24 deleções e 41 substituições de nucleotídeos foi a responsável pela identidade de 90% obtida com a sequência codificante do gene CSN1S1 de búfalo.

O alinhamento múltiplo da sequência de aminoácidos da caseína alfa-s1 de búfalo com as sequências homólogas em boi (nº de acesso no *GenBank*: NP_851372.1), cabra (nº de acesso no *GenBank*: NP_001272624.1) e ovelha (nº de acesso no *GenBank*: NP_001009795.1) revelou identidades de 95%, 87% e 84%, respectivamente. A Figura 12 apresenta o alinhamento múltiplo entre essas sequências, destacando as alterações de aminoácidos nas sequências da proteína caseína alfa-s1 de boi, cabra e ovelha, com base na sequência correspondente em búfalo, escolhida como referência.

```

O. aries  MKLLILTCLVAVALARPKHPIKHQGLSPEVLNENLLRFFVAPFPEVFRKENINELSKDIGSESTEDQAMEDAKQMK
C. hircus  MKLLILTCLVAVALARPKHPINHQGLSPEVLNENLLRFFVAPFPEVFRKENINELSKDIGSESTEDQAMEDAKQMK
B. bubalis MKLLILTCLVAVALARPKQPIKHQGLPQGVNENLLRFFVAPFPEVFGKEKVNELSTDIGSESTEDQAMEDIKQMEA
B. taurus  MKLLILTCLVAVALARPKHPIKHQGLPQEVNENLLRFFVAPFPEVFGKEKVNELSKDIGSESTEDQAMEDIKQMEA
          *****:**:****. * *****.***** **: **:****.***** ***** ***:
O. aries  GSSSSSEIIVPNSAEQKYIQKEDVPSELYLGYLEQLLRLKKYNVPQLEIVPKSAEEQLHSMKEGNEPAHQKQPMIAVN
C. hircus  GSSSSSEIIVPNSAEK-YIQKEDVPSELYLGYLEQLLRLKKYNVPQLEIVPNSAEELHSMKEGNEPAHQKQPMIAVN
B. bubalis ESISSEIIVPISVEQKHIQKEDVPSELYLGYLEQLLRLKKYNVPQLEIVPNLAEEQLHSMKEGIHAQQKEPMIGVN
B. taurus  ESISSEIIVPNSVEQKHIQKEDVPSELYLGYLEQLLRLKKYKVPQLEIVPNSAEERLHSMKEGIHAQQKEPMIGVN
          * ***** *: *: *****:*****:***** **: **:***** *:**:***.***
O. aries  -----QLFRQFYQLDAYPSGAWYYLPLGTQYTDAPSFSDIPNPIGSENSGKITMPLW 206
C. hircus  QELAYFYPQLFRQFYQLDAYPSGAWYYLPLGTQYTDAPSFSDIPNPIGSENSGKITMPLW 213
B. bubalis QELAYFYPQLFRQFYQLDAYPSGAWYYVPLGTQYDPDAPSFSDIPNPIGSENSGKITMPLW 214
B. taurus  QELAYFYPQLFRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSENSEKITMPLW 214
          :*****:*****.*****.***** ***** * *****

```

Figura 12. Alinhamento múltiplo da sequência de aminoácidos da proteína caseína alfa-s1 de búfalo com as sequências de aminoácidos correspondentes em boi, cabra e ovelha, realizado com o auxílio da ferramenta *Clustal Omega*. As alterações de aminoácidos em relação à sequência da proteína caseína alfa-s1 de búfalo estão destacadas em cinza.

A partir do alinhamento múltiplo representado na Figura 12, foi possível observar que a ocorrência de 10 trocas de aminoácidos na sequência da caseína alfa-s1 de boi resultou na identidade de 95% obtida com a sequência correspondente em búfalo. Além disso, o alinhamento mostrou que a identidade de 87% obtida com a sequência de cabra foi resultante da ausência de um aminoácido e presença de 28 trocas de aminoácidos na mesma. Com relação à sequência de ovelha, foi observado que a ausência de oito aminoácidos e a ocorrência de 24 trocas de aminoácidos foram responsáveis pela identidade de 84% obtida com a sequência da caseína alfa-s1 de búfalo.

Em uma análise do alinhamento entre as sequências de aminoácidos da caseína alfa-s1 de búfalo e boi foi observado que das 16 substituições de nucleotídeos identificadas na sequência codificante do gene CSN1S1 de búfalo,

quatro são sinônimas e 10 não sinônimas. A ocorrência destas substituições não sinônimas resultou nas seguintes trocas de aminoácidos na cadeia polipeptídica de búfalo: His19Gln, Glu29Gly, Lys57Thr, Asn89Lle, Lys120Asn, Ser130Leu, Arg134Gln, Glu163Gln, Thr189Lle e Glu207Gly (Figura 12).

Ferranti et al. (1998) identificaram essas mesmas trocas de aminoácidos em uma análise comparativa realizada entre as sequências de aminoácidos da caseína alfa-s1 de búfalo e boi. De acordo com esse estudo, as trocas Glu29Gly, Ser130Leu e Glu163Gln contribuíram para a diminuição da carga negativa da caseína alfa-s1 de búfalo em relação à proteína homóloga em boi, pois uma carga negativa foi perdida para cada troca Glu29Gly e Glu163Gln, enquanto que para troca Ser130Leu foram perdidas duas cargas negativas. Já as trocas His19Gln, Lys57Thr, Lys120Asn e Arg134Gln ocasionaram a perda de quatro resíduos de aminoácidos carregados positivamente. Acredita-se que a ocorrência dessas últimas trocas de aminoácidos compensou a perda dos aminoácidos de carga negativa, de modo que o ponto isoelétrico da caseína alfa-s1 de búfalo não foi afetado, resultando em uma proteína com características similares à de boi (CHIANESE et al., 1996; FERRANTI et al., 1998).

Os resultados obtidos nessa análise comparativa podem melhorar a compreensão da evolução do gene CSN1S1 em bovídeos, bem como das consequências funcionais que resultaram dos diferentes aminoácidos que constituem as caseínas alfa-s1.

Como não foi possível obter as sequências codificantes a partir das sequências parciais dos genes CSN3 e CSN2 de búfalo, não foram realizadas análises comparativas das sequências codificantes e de aminoácidos desses genes com outros bovídeos.

5.7 Comparação da sequência de DNA dos genes da raça Murrah com aquelas correspondentes na raça Mediterrâneo

A sequência de DNA do gene CSN1S1 do animal da raça Murrah obtida no presente estudo foi comparada com a sequência correspondente da raça Mediterrâneo, disponível no banco de dados americano NCBI (nº de acesso no

GenBank: NW_005785429.1), com o auxílio do programa *Clustal Omega*. As análises comparativas revelaram a ocorrência de 53 alterações de nucleotídeos na sequência de DNA do gene CSN1S1 de Murrah em relação à sequência de Mediterrâneo, sendo elas: 43 substituições, quatro inserções e seis deleções. Estas alterações estão descritas no Quadro 2 de acordo com o padrão internacional de nomenclatura de variações no DNA.

Quadro 2. Alterações de nucleotídeos observadas na sequência do gene CSN1S1 de Murrah, gerada no presente trabalho, em relação à sequência correspondente em Mediterrâneo (UMD_CASPUR_WB_2.0) disponível no banco de dados americano NCBI.

Região do gene CSN1S1	Posição/Alteração	Região do gene CSN1S1	Posição/Alteração
Íntron 1	(g. 400G>A)	Íntron 14	(g.13105C>T)
	(g.523_524insG)	Íntron 15	(g.13364G>T)
	(g.1006G>C)		(g.13479T>G)
Íntron 2	(g.1968G>A)		(g.13667C>G)
	(g.3066delT)		(g.14132A>G)
	(g.3217delT)	Íntron 16	(g.14334T>C)
Íntron 3	(g.3567C>T)		(g.14620T>C)
	(g.3683_3684insA)		(g.14682A>G)
	(g.3686A>C)		(g.14686A>C)
Íntron 8	(g.6499T>C)		(g.14750A>T)
	(g.6534C>T)		(g.14942C>G)
	(g.6975G>C)		(g.14949C>T)
	(g.7423A>G)	Éxon 17	(g.15149T>C)
Íntron 11	(g.9172A>G)	Íntron 17	(g.15240A>G)
	(g.9298_9299insT)		(g.15313A>T)
	(g.9359A>G)		(g.15526T>C)
	(g.10261delA)		(g.15619C>T)
Íntron 12	(g.10807C>T)		(g.15624_15625delTT)
	(g.11076C>T)	Éxon 18	(g.15809T>C)
	(g.11092delT)	Íntron 18	(g.15900C>T)
Íntron 13	(g.11262G>A)		(g.15942G>T)
	(g.11449G>A)		(g.15944T>A)
	(g.11633A>G)		(g.16192T>C)
	(g.11809T>C)		(g.16207A>G)
	(g.11951T>A)		(g.16210_16211insT)
	(g.12076G>A)		(g.16335T>C)

Como pode ser observado no Quadro 2, a maioria das substituições de nucleotídeos foram encontradas em regiões de íntrons, com exceção somente daquelas identificadas nos éxons 17 e 18.

As análises das sequências mostraram que a substituição de nucleotídeo identificada no éxon 17 (g.15149T>C) do gene CSN1S1 de Murrah é uma

substituição não sinônima que corresponde ao nucleotídeo 578 da sequência codificante do gene CSN1S1 de búfalo. A Figura 13 apresenta o alinhamento entre as sequências codificantes do gene CSN1S1 de Murrah e Mediterrâneo, onde a substituição não sinônima do éxon 17 está destacada.

```

Murrah      ATGAAACTTCTCATCCTTACCTGTCTTGTGGCTGTTGCTCTTGCCAGGCCTAAACAGCCTATCAAGCACC
Mediterrâneo ATGAAACTTCTCATCCTTACCTGTCTTGTGGCTGTTGCTCTTGCCAGGCCTAAACAGCCTATCAAGCACC
*****

Murrah      AAGGACTCCCTCAAGGAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCGCCTTTTCCAGAAGTGTT
Mediterrâneo AAGGACTCCCTCAAGGAGTCCTCAATGAAAATTTACTCAGGTTTTTTGTGGCGCCTTTTCCAGAAGTGTT
*****

Murrah      TGGAAAGGAGAAGGTCAATGAACTGAGCACGGATATTGGGAGTGAATCAACTGAGGATCAAGCCATGGAA
Mediterrâneo TGGAAAGGAGAAGGTCAATGAACTGAGCACGGATATTGGGAGTGAATCAACTGAGGATCAAGCCATGGAA
*****

Murrah      GATATTAAGCAAATGGAAGCTGAAAGCATTTTCGTCAAGTGAGGAAATTGTTCCCATTAGTGTTGAGCAGA
Mediterrâneo GATATTAAGCAAATGGAAGCTGAAAGCATTTTCGTCAAGTGAGGAAATTGTTCCCATTAGTGTTGAGCAGA
*****

Murrah      AGCACATTCAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTTAGACTGAA
Mediterrâneo AGCACATTCAAAGGAAGATGTGCCCTCTGAGCGTTACCTGGGTTATCTGGAACAGCTTCTTAGACTGAA
*****

Murrah      AAAATACAACACTACCCAGCTGGAAATTGTTCCCAATTTGGCTGAGGAACAACCTTCACAGTATGAAAGAGG
Mediterrâneo AAAATACAACACTACCCAGCTGGAAATTGTTCCCAATTTGGCTGAGGAACAACCTTCACAGTATGAAAGAGG
*****

Murrah      AATCCATGCCCCAACAGAAAGAACCTATGATAGGAGTGAATCAGGAACCTGGCCTACTTCACCCCTCAGCTTT
Mediterrâneo AATCCATGCCCCAACAGAAAGAACCTATGATAGGAGTGAATCAGGAACCTGGCCTACTTCACCCCTCAGCTTT
*****

Murrah      TCAGACAATTCTACCAGCTGGATGCCTATCCATCTGGTGCCTGGTATACGTTCCACTAGGCACGCAATAC
Mediterrâneo TCAGACAATTCTACCAGCTGGATGCCTATCCATCTGGTGCCTGGTATACGTTCCACTAGGCACGCAATAC
*****

Murrah      CCTGATGCCCCATCATTCTCTGACATCCCTAATCCCATCGGCTCTGAGAACAGTGGAAGACTACTATGC
Mediterrâneo CCTGATGCCCCATATTCTCTGACATCCCTAATCCCATCGGCTCTGAGAACAGTGGAAGACTACTATGC
*****

Murrah      CACTGTGGTGA 645
Mediterrâneo CACTGTGGTGA 645
*****

```

Figura 13. Alinhamento entre as sequências codificantes do gene CSN1S1 de Murrah e Mediterrâneo, realizado com o auxílio da ferramenta *Clustal Omega*. A única substituição não sinônima (éxon 17) está destacada em cinza.

Na Figura 14 se encontra ilustrado o alinhamento entre as sequências de aminoácidos da caseína alfa-s1 de Murrah e Mediterrâneo (nº de acesso no *GenBank*: XP_006071187.1). Nessa figura é possível observar que a substituição não sinônima do éxon 17 resultou na troca do aminoácido leucina por serina na posição 193 da caseína alfa-s1 de Murrah em relação à de Mediterrâneo (Leu193Ser).

```

Murrah      MKLLILTCLVAVALARPKQPIKHQGLPQGVNENLLRFFVAPFPEVFGKEKVNELSTDIGSESTEDQAME
Mediterrâneo MKLLILTCLVAVALARPKQPIKHQGLPQGVNENLLRFFVAPFPEVFGKEKVNELSTDIGSESTEDQAME
*****

Murrah      DIKQMEAESISSSEEIVPISVEQKHQKEDVPSERYLGYLEQLLRLLKKYNVPQLEIVPNLAEEQLHSMKE
Mediterrâneo DIKQMEAESISSSEEIVPISVEQKHQKEDVPSERYLGYLEQLLRLLKKYNVPQLEIVPNLAEEQLHSMKE
*****

Murrah      GIHAQQKEPMIGVNQELAYFYPQLFRQFYQLDAYPSGAWYYVPLGTQYPDAPSFSDIPNPIGSENSGKTT
Mediterrâneo GIHAQQKEPMIGVNQELAYFYPQLFRQFYQLDAYPSGAWYYVPLGTQYPDAPLFSFSDIPNPIGSENSGKTT
*****

Murrah      MPLW 214
Mediterrâneo MPLW 214
*****

```

Figura 14. Alinhamento entre as sequências de aminoácidos da caseína alfa-s1 de Murrah e Mediterrâneo, realizado com o auxílio da ferramenta *Clustal Omega*. A única troca de aminoácido (Leu193Ser) está destacada em cinza.

Estudos de polimorfismos com populações bubalinas têm mostrado que a substituição não sinônima do éxon 17 (g.15149T>C) compreende a um SNP que resulta na troca do aminoácido Leu193Ser na cadeia polipeptídica da caseína alfa-s1 de búfalo (FERRANTI et al., 1998; CHIANESE et al., 2009; COSENZA et al., 2011; EL NAHAS et al., 2013). Nessa posição, a presença de timina (T) corresponde ao alelo A (TTA) e a de citosina (C) ao alelo B (TCA), o qual tem se mostrado mais frequente nas populações bubalinas genotipadas até o momento (CHIANESE et al., 2009; EL NAHAS et al., 2013). Até o momento, não há estudos de associação desse polimorfismo com características de interesse econômico em búfalo.

Com relação à substituição de nucleotídeo identificada no éxon 18 (g.15809T>C), observou-se que a mesma está presente na região 3'UTR do gene CSN1S1 de búfalo.

Assim, a substituição de nucleotídeo do éxon 18, somada às demais substituições identificadas nos íntrons do gene CSN1S1 de Murrah, totalizam 42 potenciais polimorfismos do tipo SNP, ao quais são interessantes para serem estudados em rebanhos bubalinos leiteiros, visando buscar associações com características de produção e qualidade do leite.

6. CONCLUSÃO

A partir dos resultados obtidos neste trabalho, pode-se concluir que:

- a varredura de 33.792 clones da biblioteca genômica de búfalo resultou na identificação de dois clones positivos para os genes das caseínas, sendo um para o gene CSN1S1 (clone A/2) e outro para o gene CSN3 (clone L/8);

- as análises das sequências de DNA obtidas a partir do sequenciamento dos clones, resultaram na caracterização do gene CSN1S1 inteiro e de sequências parciais correspondentes aos genes CSN2 (éxon 9 e parte do íntron 8) e CSN3 (parte dos íntrons 2 e 3 e a sequência completa do íntron 4 e dos éxons 3, 4 e 5) de búfalo;

- nas análises comparativas, os éxons e íntrons dos genes CSN1S1, CSN2 (parcial) e CSN3 (parcial) apresentaram conservação acima de 85%, 98% e 94%, respectivamente, entre búfalo e boi. Já as regiões intergênicas identificadas mostraram uma conservação acima de 70% entre essas espécies.

- na avaliação das sequências de DNA quanto à presença de sequências repetitivas, os elementos transponíveis do tipo retrotransposon foram os mais representativos, abrangendo 43,83% e 44,98% das sequências de DNA dos clones A/2 e L/8, respectivamente. As análises comparativas das sequências repetitivas do gene CSN1S1 de búfalo com aquelas contidas no gene correspondente em outros bovídeos (boi, iaque, cabra e ovelha), camundongo e humano revelaram duas sequências repetitivas búfalo específicas (uma repetição em “tandem” GT e uma repetição de baixa complexidade rica em adenina) e sete comuns em todas as espécies de bovídeos, sendo elas: quatro SINES (BOV-A2, MAR1c_Mdo, MIRb e MIR3), uma LINE (BovB) e duas repetições em “tandem” (AATATTA e ATGA). Com relação às sequências parciais do gene CSN3 de búfalo, as análises comparativas auxiliaram na identificação de três sequências repetitivas búfalo específicas, incluindo duas SINES (Bov-tA1 e MIRb) e uma repetição de baixa complexidade rica

em adenina, e sete comuns em todas as espécies de bovídeos, sendo elas: quatro LINEs (L1_Art, L1-1_Ttr e dois BovB) e três SINEs (Bov-tA2, BOV-A2 e SINE1_BB).

- nas análises com as sequências homólogas de outros bovídeos, a sequência codificante do gene CSN1S1 de búfalo apresentou identidade de 98% com boi, 93% com cabra e de 90% com ovelha. Já a sequência de aminoácidos da caseína alfa-s1 de búfalo mostrou 95%, 87% e 84% de identidade com as sequências correspondentes em boi, cabra e ovelha, respectivamente.

- a comparação entre as sequências do gene CSN1S1 de representantes das raças Murrah e Mediterrâneo resultaram na identificação de 42 potenciais polimorfismos do tipo SNP, dos quais um se encontra no éxon 18 e os demais em regiões de íntrons do gene. O número de potenciais SNPs identificados nos respectivos íntrons do gene CSN1S1 de Murrah foram: dois no íntron 1; um no íntron 2; dois no íntron 3; quatro no íntron 8; dois no íntron 11; dois no íntron 12; seis no íntron 13; um no íntron 14; quatro no íntron 15; sete no íntron 16; quatro no íntron 17 e seis no íntron 18.

7. REFERÊNCIAS BIBLIOGRÁFICAS*

ABCB. **Associação Brasileira de Criadores de Búfalos**. 2010. Disponível em: <<http://www.bufalo.com.br/abcb.html>>. Acesso em: 10 abr. 2014.

ADELSON, D. L.; RAISON, J. M.; EDGAR, R. C. Characterization and distribution of retrotransposons and simple repeats in the bovine genome. **Proceedings of the National Academy of Sciences of the United States of American**, v. 106, p. 12855-12860, 2009.

AFOLDI, J.; LINDOBLAD-TOH, K. Comparative genomics as a tool to understand evolution and disease. **Genome Research**, v. 23, p. 1063-1068, 2013.

ALMEIDA, L. M.; SILVA, I. T.; SILVA JR, W. A.; CASTRO, J. P.; RIGGS, P. K.; CARARETO, C. M.; AMARAL, M. E. K. The contribution of transposable elements to *Bos taurus* gene structure. **Gene**, v. 390, p. 180-189, 2007.

ALMEIDA, L. M.; AMARAL, M. E. J.; SILVA, I. T.; SILVA JR, W. A.; RIGGS, P. K.; CARARETO, C.M. Report of a chimeric origin of transposable elements in a bovine-coding gene. **Genetics Molecular Research**, v. 7, p. 107-116, 2008.

AMANO, T. Genetic differences between swamp and river buffaloes in biochemical and immunological characteristics. In: Proceedings of the Preconference Symposium of the 5th World Conference on Animal Production. **Japanese Society of Zootechnical Sciences**, Tsukuba, Japão, p. 131-135, 1983.

AMARAL, M. E. J.; GRANT, J. R.; RIGGS, P. K.; STAFUZZA, N. B; RODRIGUES-FILHO, E. A.; GOLDAMMER, T.; WEIKARD, R.; BRUNNER, R. M.; KOCHAN, K. J.; GRECO, A. J.; JEONG, J.; CAI, Z.; LIN, G.; PRASAD, A.; KUMAR, S.; SARADHI, G. P.; MATHEW, B.; KUMAR, M. A.; MIZIARA, M. N.; MARIANI, P.; CAETANO, A. R.; GALVÃO, S. R.; TANTIA, M. S.; VIJH, R. K.; MISHRA, B.; KUMAR, S. T. B.; PELAI, V. A.; SANTANA, A. M.; FORNITANO, L. C.; JONES, B. C.; TONHATI, H.; MOORE, S.; STOTHARD, P.; WOMACK, J. E. A first generation whole genome RH map of the river buffalo with comparison to domestic cattle. **BMC Genomics**, v. 9, 2008.

ANUALPEC 2009: anuário da pecuária brasileira. São Paulo: FNP Consultoria/Agros Comunicação, 2009. 400 p.

BARKER, J. S. F.; TAN, S. G.; SELVARAJ, O. S.; MUKHERJEE, T. K. Genetic variation within and relationships among populations of Asian water buffalo (*Bubalus bubalis*). **Animal Genetics**, v. 28, p. 1-13, 1997a.

BARKER, J. S. F.; MOORE, S. S.; HETZEL, D. J. S.; EVANS, D.; TAN, S. G.; BYRNE, K. Genetic diversity of Asian water buffalo (*Bubalus bubalis*): microsatellite variation and a comparison with protein coding loci. **Animal Genetics**, v. 28, p. 103-115, 1997b.

*Normas do Programa de Pós-Graduação da Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal - FCAV-UNESP

BARLOWSKA, J.; WOLANCIUK, A.; LITWIŃCZUK, Z.; KRÓL, J. Milk proteins' polymorphism in various species of animals associated with milk production utility. **Intech**, 2012. Disponível em: <<http://www.intechopen.com/>>. Acesso em: 15 out. 2014.

BERARDINO, D. D.; IANNUZZI, L. Chromosome banding homologies in Swamp and Murrah buffalo. **The Journal of Heredity**, v. 72, p. 183-188, 1981.

BERNARDES, O. Bubalinocultura no Brasil: situação e importância econômica. **Revista Brasileira de Reprodução Animal**, v. 31, p. 293-298, 2007.

BEST, S.; LE, T. P.; TOWERS, G.; STOYE, J. P. Positional cloning of the mouse retrovirus restriction gene Fv1. **Nature**, v. 382, p. 826-829, 1996.

BHORGUESE, A. Buffalo Breeds and Management Systems In: MOIOLI, B.; BORGHESE, A. **Buffalo Production and Research**. Roma: FAO Regional Office for Europe, 2005. p. 51-76.

BONFATTI, V.; GIANTIN, M.; GERVASO, M.; COLETTA, A.; DACASTO, M.; CARNIER, P. Effect of CSN1S1-CSN3 (α_{s1} -k-casein) composite genotype on milk production traits and milk coagulation properties in Mediterranean water buffalo. **Journal of Dairy Science**, v. 95, p. 3435-3443, 2012.

BOURQUE, G.; LEONG, B.; VEGA, V. B.; CHEN, X.; LEE, Y. L.; SRINIVASAN, K. G.; CHEW, J. L.; RUAN, Y.; WEI, C. L.; NG, H. H.; LIU, E. T. Evolution of the mammalian transcription factor binding repertoire via transposable elements. **Genome Research**, v. 18, p. 1752-1762, 2008.

BRANDT, J.; SCHRAUTH, S.; VEITH, A.; FROSCHAUER, A.; HANEKE, T.; SCHULTHEIS, C.; GESSLER, M.; LEIMEISTER, C.; VOLFF, J. N. Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. **Gene**, v. 345, p. 101-111, 2005.

BRITO, M. A.; BRITO, J. R.; ARCURI, E.; LANGE, C.; SILVA, M.; SOUZA, G. **Composição do leite**. Empresa Brasileira de Pesquisa Agropecuária, 2007. Disponível em: <http://www.agencia.cnptia.embrapa.br/Agencia8/AG01/arvore/AG01_128_2172003_9243.html>. Acesso em: 30 set. 2014.

BRITTEN, R. J. Coding sequences of functioning human genes derived entirely from mobile element sequences. **Proceedings of the National Academy of Sciences of the United States of American**, v. 101, p. 16825-16830, 2004.

BROPHY, B.; SMOLENSKI, G.; WHEELER, T.; WELLS, D.; L'HUILLIER, P.; LAIBLE, G. Cloned transgenic cattle produce milk with higher levels of β -casein and κ -casein. **Nature Biotechnology**, v. 21, p. 157-162, 2003.

CAROLI, A. M.; CHESSA, S.; ERHARDT, G. J. Milk protein polymorphism in cattle: effect on animal breeding and human nutrition. **Journal Dairy Science**, v. 92, p. 5335-5352, 2009.

CERIOTTI, G.; CHESSA, S.; BOLLA, P.; BUDELLI, E.; BIANCHI, L.; DURANTI, E.; CAROLI, A. Single nucleotide polymorphisms in the ovine casein genes detected by polymerase chain reaction-single strand conformation polymorphism. **Journal of Dairy Science**, v. 8, p. 2606-2613, 2004.

CHEN, R.; SODERGEN, E.; WEINSTOCK, G. M.; GIBBS, R. A. Dynamic building of a BAC clone tiling path for the Rat Genome Sequencing Project. **Genome Research**, v. 14, p. 679-684, 2004.

CHIANESE, L.; GARRO, G.; FERRANTI, P.; CAIRA, S.; ADDEO, F. Occurrence of casein polymorphism in water buffalo milk. **European Association for Animal Production Publication**, v. 82, p. 187-191, 1996.

CHIANESE, L.; QUARTO, M.; PIZZOLONGO, F.; CALABRESE, M. G.; CAIRO, S.; MAURIELLO, R.; DE PASCALE, S.; ADDEO, F. Occurrence of genetic polymorphism at the alpha s1-casein locus in Mediterranean water buffalo milk. **International Dairy Journal**, v. 19, p. 181-189, 2009.

CITEK, J.; PANICKE, L.; REHOUT, V.; PROCHAZKOVA, H. Study of genetic distances between cattle breeds of Central Europe. **Czech Journal Animal Science**, v. 51, p. 429-436, 2006.

CORDAUX, R.; BATZER, M. A. The impact of retrotransposons on human genome evolution. **Nature Reviews Genetics**, v. 10, p. 691-703, 2009.

COSENZA, G.; PAUCIULLO, A.; COLETTA, A.; DI FRANCIA, A.; FELIGINI, M.; GALLO, D.; DI BERARDINO, D.; RAMUNNO, L. Short communication: translational efficiency of casein transcripts in Mediterranean river buffalo. **Journal Dairy Science**, v. 94, p. 5691-5694, 2011.

DALGLEISH, D. G. On the structural models of bovine casein micelles - Review and possible improvements. **Soft Matter**, v. 7, p. 2265-2272, 2011.

DAVIS, M. B.; DIETZ, J.; STANDIFORD, D. M.; EMERSON JR, C. P. Transposable element insertions respecify alternative exon splicing in three *Drosophila* myosin heavy chain mutants. **Genetics**, v. 150, p. 1105-1114, 1998.

DI STASIO, L.; MARIANI, P. The role of protein polymorphism in the genetic improvement of milk production. **Zootecnica e Nutrizione Animale**, v. 26, p. 69-90, 2000.

DOMINGUES, P. F.; LANGONI, H. **Manejo sanit rio animal**. Rio de Janeiro: EPUB, 2001. p. 224.

DUNN, C. A.; ROMANISH, M.; GUTIERREZ, L. E.; VAN LAGEMAAT, L. N.; MAGER, D. L. Transcription of two genes from a bidirectional endogenous retrovirus promoter. **Gene**, v. 366, p. 335-342, 2006.

ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nature Reviews Genetics**, v. 5, p. 435-445, 2004.

EL NAHAS, S. M.; BIBARS, M. A.; TAHA, D. A.; EL-SAYYAD, H. I. Detection of two CSN1S1 variants in Egyptian buffalo. **Journal of Genetic Engineering and Biotechnology**, v. 11, p. 75-77, 2013.

FAO. **Food and Agriculture Organization of the United Nations**. Faostat agriculture data, 2013. Disponível em: <<http://faostat.fao.org/site/573/default.aspx#ancor>>. Acesso em: 10 set. 2014.

FARREL, H. M. J.; JIMENEZ-FLORES, R.; BLECK, G. T.; BROWN, E. M.; BUTLER, J. E.; CREAMER, L. K.; HICKS, C. L.; HOLLAR, C. M.; NG-KWAI-HANG, K. F.; SWAISGOOD, H. E. Nomenclature of the proteins of cows'milk - sixth revision. **Journal Dairy Science**, v. 87, 1641-1674, 2004.

FERRANTI, P.; SCALONI, A.; CAIRA, S.; CHIANESE, L.; MALORNI, A.; ADDEO, F. The primary structure of water buffalo α_{s1} - and β -Casein: Identification of phosphorylation sites and characterization of a novel β -Casein variant. **Journal of Protein Chemistry**, v. 17, p. 835-844, 1998.

FISCHER, H.; ULBRICH, F. Chromosomes of the Murrah buffalo and its crossbreds with the Asiatic swamp buffalo (*Bubalus bubalis*). **Zeitschrift für Tierzüchtung und Züchtungsbiologie**, v. 84, p. 110-114, 1968.

FOX, P. F.; MCSWEENEY, P. L. H. Milk proteins. In: _____. **Dairy Chemistry and biochemistry**. London: Blackie Academic, 1998. p. 147-238.

FRAZER, K.A.; PACHTER, L.; POLIAKOV, A.; RUBIN, E.M.; DUBCHAK, I. VISTA: computational tools for comparative genomics. **Nucleic Acids Research**, v. 32, p. 273-279, 2004.

GOUDA, E. M.; GALAL, M. K.; ABDELAZIZ, S. A. Genetic variants and allele frequencies of kappa casein in Egyptian cattle and buffalo using PCR-RFLP. **Journal of Agricultural Science**, v. 5, 197-203, 2013.

GRIFFITHS, A. J. F.; WESSLER, S. R.; LEWONTIN, R. C.; GELBART, W. G. RNA: tradução e processamento. In: _____. **Introdução à genética**. 8. ed. Rio de Janeiro: Guanabara Koogan, 2006. p. 245-261.

HALLÉN, E.; ALLMERE, T.; NASLUND, J.; ANDRÉN, A.; LUNDÉN, A. Effect of genetic polymorphism of milk proteins on rheology of chymosin-induced milk gels. **International Dairy Journal**, v. 17, p. 791-799, 2007.

HOFF, K. J.; STANKE, M. WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes. **Nucleic Acids Research**, v. 41, p. w123-w128, 2013.

HOLT, C.; CARVER, J. A.; ECROYD, H. W. Invited review: Caseins and the casein micelles: Their biological functions, structures, and behavior in foods. **Journal Dairy Science**, v. 96, p. 6127-6146, 2013.

HUPPERTZ, T.; FOX, P. F.; KRUIF, K. G.; KELLY, A. L. High pressure induced changes in bovine milk proteins: a review. **Biochimica et Biophysica Acta**, v. 1764, p. 593-598, 2006.

IANNUZZI, L. Standard karyotype of the river buffalo (*Bubalus bubalis* L., 2n = 50). Report of the committee for the standardization of banded karyotypes of the river buffalo. **Cytogenetics and Cell Genetics**, v. 67, p. 102-113, 1994.

IANNUZZI, L.; DI MEO, G. P.; PERUCATTI, A.; SCHIBLER, L.; INCARNATO, D.; GALLAGHER, D.; EGGEN, A.; FERRETI, L.; CRIBIU, E. P.; WOMACK, J. The river buffalo (*Bubalus bubalis*, 2n=50) cytogenetic map: assignment of 64 loci by fluorescence in situ hybridization and R-banding. **Cytogenetic and Genome Research**, v. 102, p. 65-75, 2003.

IBGE. **Instituto Brasileiro de Geografia e Estatística**. Censo Agropecuário 2010. Disponível em: <http://www.ibge.gov.br/home/estatistica/economia/ppm/2010/tabelas_pdf/tab11.pdf>. Acesso em: 30 out. 2014.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Initial sequencing and analysis of the human genome. **Nature**, v. 409, p. 860-921, 2001.

IWASHITA, S.; UENO, S.; NAKASHIMA, K.; SONG, S. Y.; OHSHIMA, K.; TANAKA, K.; ENDO, H.; KIMURA, J.; KUROMARU, M.; FUKUTA, K.; DAVID, L.; OSADA, N. A tandem gene duplication followed by recruitment of a retrotransposon created the paralogous bucentaur gene (bcntp97) in the ancestral ruminant. **Molecular Biology and Evolution**, v. 23, p. 798-806, 2006.

JURKA, J.; KAPITONOV, V. V.; PAVLICEK, A.; KLONOWSKI, P.; KOHANY, O.; WALICHIEWICZ, J. Repbase Update, a database of eukaryotic repetitive elements. **Cytogenetic and Genome Research**, v. 110, p. 462-467, 2005.

JURKA, J.; KAPITONOV, V. V.; KOHANY, O.; JURKA, M. V. Repetitive sequences in complex genomes: structure and evolution. **Annual Review of Genomics Human Genetics**, v. 8, p. 241-259, 2007.

KAMINSKI, S. Polymorphism of milk protein genes in coding and regulatory regions and their effects on gene expression and milk performance traits. **Animal Science Papers and Reports**, v. 22, p. 109-113, 2004.

KAMINSKI, S.; MALEWSKI, T.; AHMAN, A.; WÓJCIK, E.; RUŚĆ, A.; OLEŃSKI, K.; JAKUBCZAK, A.; SAZANOV, A. A. Towards and integrated approach to study SNPs and expression of candidate genes associated with milk proteins biosynthesis. **Animal Genetics**, v. 44, p. 459-465, 2008.

KIKKAWA, Y.; YONEKAWA, H.; SUZUKI, H.; AMANO, T. Analysis of genetic diversity of domestic water buffaloes and anoas based on variations in the mitochondrial gene by cytochrome b. **Animal Genetics**, v. 28, p. 195-201, 1997.

KIPLAGAT, S. K.; LIMO, M. K.; KOSGEY, I. S. Genetic improvement of livestock for milk production. **Intech**, 2012. Disponível em: <<http://www.intechopen.com/>>. Acesso em: 10 abr. 2014.

KITTS, D. D.; YUAN, Y. V.; NAGASAWA, T.; MORIYAMA, Y. Effect of casein, casein phosphopeptides and calcium intake on ileal ⁴⁵Ca disappearance and temporal systolic blood pressure in spontaneous hypertensive rats. **British Journal of Nutrition**, v. 68, p. 765-781, 1992.

KORHONEN, H.; PIHLANTO-LEPPALA, A.; RANTAMAKI, P.; TUPASELA, T. Impact of processing on bioactive proteins and peptides. **Trends in Food Science & Technology**, v. 9, p. 307-319, 1998.

KRISSANSEN, G. W. Emerging health properties of whey proteins and their clinical implications. **Journal of the American College of Nutrition**, v. 26, p. 713S-23S, 2007.

KRÜGER, C. C. H.; SILVA, C. A.; VEDANA, M. I. S.; TIENE, C.; CÂNDIDO, L. M. B. Atividade antimicrobiana de peptídeos obtidos de caseína bovina. **Alimentos e Nutrição**, v. 17, p. 7-12, 2006.

LAU, C. H.; DRINKWATER, R. D.; YUSOFF, K.; TAN, S. G.; HETZEL, D. J. S.; BARKER, J. S. F. Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and cytochrome b sequence variation. **Animal Genetics**, v. 29, p. 253-264, 1998.

LEMAY, D. G.; LYNN, D. J.; MARTIN, W. F.; NEVILLE, M. C.; CASEY, T. M.; RINCON, G.; KRIVENTSEVA, E. V.; BARRIS, W. C.; HINRICHS, A. S.; MOLENAAR, A. J.; POLLARD, K. S.; MAQBOOL, N. J.; SINGH, K.; MURNEY, R.; ZDOBNOV, E. M.; TELLAM, R. L.; MEDRANO, J. F.; GERMAN, J. B.; RIJNKELS, M. The bovine lactation genome: insights into the evolution of mammalian milk. **Genome Biology**, v. 10, p. R43.1-R43.18, 2009.

LENSTRA, J. A.; BRADLEY, D. G. Systematics and phylogeny of cattle. In: FRIES, R.; RUVINSKY, A. **The genetics of cattle**. New York: CAB International, 1999. cap 1, p. 1-14.

LINDBLAD-TOH, K.; WADE, C. M.; MIKKELSEN, T. S.; KARLSSON, E. K.; JAFFE, D. B.; KAMAL, M.; CLAMP, M.; CHANG, J. L.; KULBOKAS, E. J.; ZODY, M. C. et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. **Nature**, v. 438, p. 803-819, 2005.

LIU, L.; YIN, J.; LI, W.; LIU, K.; PENG, Y.; TAN, P.; MA, R. Construction of a bacterial artificial chromosome library for the Rongchang pig breed and its use for the identification of genes involved in intramuscular fat deposition. **Biochemical and Biophysical Research Communications**, v. 391, p. 1280-1284, 2010.

LIU, Z. J.; CORDES, J. F. DNA marker technologies and their applications in aquaculture genetics. **Aquaculture**, v. 238, p. 1-37, 2004.

MACEDO, M. P.; WECHSLER, F. S.; RAMOS, A. A.; AMARAL, J. B.; SOUZA, J. C.; RESENDE, F. D.; OLIVEIRA, J. V. Composição físico-química e produção do leite de búfalas da raça mediterrâneo no oeste do estado de São Paulo. **Revista Brasileira de Zootecnia**, v. 30, p. 1084-1088, 2001.

MACGREGOR, R. **The Domestic Buffalo**. Tese apresentada no Royal College of Veterinary Surgeons. The Royal College of Veterinary Surgeons, Londres, p. 1-47, 1939.

MAGER, D. L.; HUNTER, D. G.; SCHERTZER, M.; FREEMAN, J. D. Endogenous retroviruses provide the primary polyadenization signal for two human genes (HLA2 and HLA3). **Genomics**, v. 59, p. 255-263, 1999.

MARCONDES, C. R. Melhoramento genético de búfalos no Brasil: avanços, entraves e perspectivas. **Revista Brasileira de Zootecnia**, v. 40, p. 307-315, 2011.

MARLETA, D.; CRISCIONE, A.; BORDONARO, S.; GUASTELLA, A. M.; D'URSO, G. Casein polymorphism in goat's milk. **Le Lait**, v. 87, p. 491-504, 2007.

MEISEL, H.; BOCKELMANN, W. Bioactive peptides encrypted in milk proteins: proteolytic activation and tropho-functional properties. **Antonie Leeuwenhoek**, v. 76, p. 207-215, 1999.

MIKKELSEN, T. S.; WAKEFIELD, M. J.; AKEN, B.; AMEMIYA, C. T.; CHANG, J. L.; DUKE, S.; GARBER, M.; GENTLES, A. J.; GOODSTADT, L.; HEGER, A. et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. **Nature**, v. 447, p. 167-177, 2007.

MILLER, W. J.; MCDONALD, J. F.; PINSKER, W. Molecular domestication of mobile elements. **Genetica**, v. 100, p. 261-270, 1997.

MISHRA, B. P.; MUKESH, M.; PRAKASH, B.; SODHI, M.; KAPILA, R.; KISHORE, A.; KATARIA, R. R.; JOSHI, B. K.; BHASIN, V.; RASOOL, T. J.; BUJARBARUAH, K. M. Status of milk protein, β -casein variants among Indian milch animals. **Indian Journal of Animal Sciences**, v. 79, p. 722-725, 2009.

MOUSE GENOME SEQUENCING CONSORTIUM. Initial sequencing and comparative analysis of the mouse genome. **Nature**, v. 420, p. 520-562, 2002.

NAKAMURA, K.; HASHIZUME, E.; SHIBATA, T.; NAKAMURA, Y.; MALA, S.; YAMANE, K. Small cytoplasmatic RNA (scRNA) gene from *Clostridium perfringens* can replace the gene for the *Bacillus subtilis* scRNA in both growth and sporulation. **Microbiology**, v. 141, p. 2965-2975, 1995.

NARDI JUNIOR, G.; RIBEIRO, M. G.; PAULIN, L.; JORGE, A. M. Brucelose em bubalinos: uma revisão com ênfase ao sorodiagnóstico oficial. **Veterinária e Zootecnia**, v. 19, p. 142-156, 2012.

NARDI JUNIOR, G.; SANTOS, E.B. Evolução da produção leiteira do Brasil. **Veterinária e Zootecnia**, v. 20, p. 216-217, 2013.

NCBI. **National Center for Biotechnonoly Information**. Disponível em: <<http://www.ncbi.nlm.nih.gov>>. Acesso em: 25 set. 2014.

PATEL, A. K.; SINGH, M.; SURYANARAYANA, V. V. S. Buffalo alpha S1-casein gene 5'-flanking region and its interspecies comparison. **Journal of Applied Genetics**, v. 55, p. 75-87, 2014.

PÉREZ, M. J.; LEROUX, C.; BONASTRE, A. S.; MARTIN, P. Occurrence of a LINE sequence in the 3'UTR of the goat alpha s1-casein E-encoding allele associated with reduced protein synthesis level. **Gene**, v. 147, p. 179-187, 1994.

RASMUSSEN, L. K.; JOHNSEN, L. B.; TSIORA, A.; SORENSEN, E. S.; THOMSEN, J. K.; NIELSEN, N. C.; JAKOBSEN, H. J.; PETERSEN, T. E. Disulphide-linked caseins and casein micelles. **International Dairy Journal**, v. 9, p. 215-218, 1999.

REBOLLO, R.; ROMANISH, M. T.; MAGER, D. L. Transposable elements: an abundant and natural source of regulatory sequences for hot genes. **Annual Review of Genetics**, v. 46, p. 21-42, 2012.

RIJNKELS, M. Multispecie comparison of the casein gene *loci* and evolution of casein gene family. **Journal Mammary Gland Biology**, v. 7, p. 327-345, 2002.

ROMANISH, M. T.; LOCK, W. M.; VAN DE LAGEMAAT, L. N.; DUNN, C. A.; MAGER, D. L. 2007. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. **PLoS Genetics**, v. 3, 2007.

SAZANOV, A. A.; MALEWSKI, T.; KAMINSKI, S.; ZWIERZCHOWSKI, L. Characterization of the CHORI-240 BAC clones containing the bovine CSN1S1, CSN2, STATH, CSN1S2 and CSN3 genes. **Journal of Applied Genetics**, v.47, p. 243-245, 2006.

SENO, L. O.; CARDOSO, V. L.; TONHATI H. Valores econômicos para as características de produção de leite de búfala no estado de São Paulo. **Revista Brasileira de Zootecnia**, v. 36, p. 2016-2022, 2007.

STAFUZZA, N. B.; ABBEY, C. A.; GILL, C. A.; WOMACK, J. E.; AMARAL, M. E. J. Construction and preliminary characterization of a river buffalo bacterial artificial chromosome library. **Genetics and Molecular Research**, v. 11, p. 3013-3019, 2012.

SWAISGOOD, H. E. Chemistry of the caseins. In: FOX, P. F.; MCSWEENEY, P. L. H. **Advanced dairy chemistry**. 3. ed. New York: Kluwer Academic/Plenum Publishers, 2003. p. 139-202.

TEIXEIRA, L. V.; BASTIANETTO, E.; OLIVEIRA, D. A. A. Leite de búfala na indústria de produtos lácteos. **Revista Brasileira de Reprodução Animal**, v. 29, p. 96-100, 2005.

THOMAS, J. W.; TOUCHMAN, J. W. Vertebrate genome sequencing: building a backbone for comparative genomics. **Trends Genetics**, v. 18, p. 104-108, 2002.

VERRUDA, M. R.; SALGADO, J. M. Análise química do leite de búfala em comparação ao leite de vaca. **Scientia Agricola**, v. 51, p. 131-137, 1994.

ZUCHT, H. D.; RAID, M.; ADERMANN, K.; MAGERT, H. J.; FORSSMAN, W. G. Casocidin-I: a casein- α s2 derived peptide exhibits antibacterial activity. **FEBS Lett**, v. 372, p. 185-188, 1995.

WALSH, C. D.; GUINEE, T. P.; REVILLE, W. D.; HARRINGTON, D.; MURPHY, J. J.; O'KENNEDY, B. T.; FITZGERALD, R. J. Influence of κ -casein genetic variant on rennet gel microstructure, cheddar cheese-making properties and casein micelle size. **International Dairy Journal**, v. 8, p. 707-714, 1998.

Apêndice - Soluções citadas no tópico Material e Métodos, utilizadas na seleção dos clones positivos para genes das caseínas.

IPTG 20% (200mg/mL)

- 2 g de IPTG
- completar com água para 10 mL

Meio LB-ágar

- 5 g de bacto triptona
- 2,5 g de extrato de levedura
- 5 g de NaCl
- 7,5 g Agar
- completar com água para 500 mL

Meio LB

- 5 g de bacto triptona
- 2,5 de extrato de levedura
- 5 g de NaCl
- completa com água para 500 mL

Tampão TAE (Tris-Acetato-EDTA) – 1X

- 40 mM de Tris-base
- 20 mM de ácido acético glacial
- 2 mM de EDTA

Tampão TBE 0,5X

- 44,5 mM de Tris-base
- 44,5 mM de ácido bórico
- 1 mM de EDTA

Xgal 2% (20 mg/mL)

- 200 µL Xgal a 100 mg/mL
- 800 µL de dimetilformamida