



**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO
CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIAS VETERINÁRIAS DO
SEMIÁRIDO**

Fernanda Fernandes Pinheiro da Costa

**RELAÇÕES FILOGENÉTICAS E PREDIÇÃO
DAS CARACTERÍSTICAS BIOQUÍMICAS A PARTIR
DE SEQUÊNCIAS DE AMINOÁCIDOS DA ENZIMA LIPOGÊNICA
DIACILGLICEROL ACILTRANSFERASE-1 EM RUMINANTES**

Petrolina – PE

2016

**UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO
CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIAS VETERINÁRIAS
DO SEMIÁRIDO**

FOLHA DE APROVAÇÃO

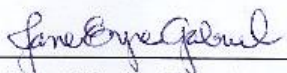
Fernanda Fernandes Pinheiro da Costa

**RELAÇÕES FILOGENÉTICAS E PREDIÇÃO
DAS CARACTERÍSTICAS BIOQUÍMICAS A PARTIR
DE SEQUÊNCIAS DE AMINOÁCIDOS DA ENZIMA LIPOGÊNICA
DIACILGLICEROL ACILTRANSFERASE-1 EM RUMINANTES**

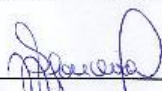
Dissertação apresentada como requisito para obtenção do título de Mestre em Ciências Veterinárias, pela Universidade Federal do Vale do São Francisco.

Aprovado em: 28 de Julho de 2016.

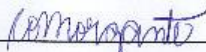
Banca Examinadora



Jane Eyre Gabriel, Doutorado em Ciências (Energia Nuclear na Agricultura), UNIVASF



João José de Simoni Gouveia, Doutorado Integrado em Zootecnia, UNIVASF



Carolina Vianna Morgante, Doutorado em Agronomia (Genética e Melhoramento de Plantas), EMBRAPA SEMIÁRIDO

FERNANDA FERNANDES PINHEIRO DA COSTA

**RELAÇÕES FILOGENÉTICAS E PREDIÇÃO
DAS CARACTERÍSTICAS BIOQUÍMICAS A PARTIR
DE SEQUÊNCIAS DE AMINOÁCIDOS DA ENZIMA LIPOGÊNICA
DIACILGLICEROL ACILTRANSFERASE-1 EM RUMINANTES**

Dissertação apresentada a Universidade Federal do Vale do São Francisco - UNIVASF, *Campus* de Ciências Agrárias, como requisito parcial para a obtenção do título de mestre.

Orientadora: Profa. Dra Jane Eyre Gabriel

Petrolina – PE

2016

Aos meus pais e a minha irmã,

DEDICO

AGRADECIMENTOS

À **Deus**, por me proporcionar forças e guiar o meu caminho permitindo dar esse passo.

À minha **família** por estar ao meu lado em todos os momentos da minha vida, apoiando e incentivando. Agradeço esse amor constante.

À minha orientadora **Profa. Dra. Jane Eyre Gabriel** pela oportunidade de desenvolver este trabalho, inestimável ajuda, paciência, direcionamento e estímulo.

Ao **Programa de Pós-Graduação em Ciências Veterinárias**, pela oferta do curso e as **secretárias** pela atenção.

Aos **professores da Pós-Graduação em Ciências Veterinárias** que contribuíram para minha formação.

À **Kárita Cláudia Freitas Lidani** pela preciosa ajuda na realização das análises.

Ao **Prof. Dr. João José de Simoni Gouveia e toda equipe** do Laboratório de Genética e Biotecnologia pela atenção prestada.

Ao **Sávio Luiz Pereira Nunes** que tanto contribuiu no encorajamento do curso.

À **Anete Ferraz Guzzi** pela amizade.

Ao **Dr. Madriano Christilis da Rocha Santos** pela ajuda em atividades durante o curso.

Ao **Francisco Peixoto de Luna** pelo apoio no transporte.

Ao **Augusto Henryque Costa de Souza, Valdenice Félix da Silva, Antônio Carlos Gomes** que mesmo indiretamente deram contribuição no curso.

Ao **Instituto Federal de Educação, Ciência e Tecnologia – Campus Petrolina Zona Rural** pela concessão da disponibilidade para a realização deste curso.

A todos que me ajudaram, de alguma forma, ao longo deste trabalho.

“Tudo posso naquele que me fortalece”.
Filipenses 4:13

RESUMO

A enzima diacilglicerol aciltransferase-1(DGAT-1) exerce um papel chave durante a biossíntese de gordura no leite. Neste estudo foram selecionadas sequências de aminoácidos da proteína DGAT-1 a partir de banco de dados de proteínas Swiss-Prot, das seguintes espécies *Bubalus bubalis* (tr|Q0ZR39|Q0ZR39, 489), *Bos indicus* (tr|A6YP72|A6YP72, 425), *Ovis aries* (tr|A8VJM4|A8VJM4, 489), *Capra hircus* (tr|A2T092|A2T092, 382), *Bos taurus* (sequência não polimórfica e polimórfica para o polimorfismo K232A) (sp|Q8MK44|1, 489). Tais sequências foram empregadas para inferir relações filogenéticas em diferentes espécies de ruminantes, bem como para prever características bioquímicas entre as sequências não-polimórfica e polimórfica para o polimorfismo K232A de DGAT-1 em bovino *Bos taurus*. Múltiplos alinhamentos a partir das sequências selecionadas foram realizados para identificar o grau de identidade desta enzima a partir dos programas Multalin e Swiss-Prot. Para inferência filogenética e avaliação da distância evolutiva entre as espécies construiu-se uma árvore evolutiva através do programa MEGA 6.0 utilizando o método Neighbor-Joining (NJ). A predição dos parâmetros bioquímicos, da estrutura secundária e das desordens da proteína DGAT-1 sem e com polimorfismo K232A foram realizados através de ferramentas dos programas ProtParam, CFSSP Chou e Fasman e DisEMBL, respectivamente. O alinhamento realizado entre as sequências das espécies de ruminantes revelou identidade de 77,50% com valor de E de 0.003. O agrupamento evolutivo entre as espécies revelou a presença de vários sub-ramos agrupados em um único ramo principal. A caracterização bioquímica da proteína DGAT-1 sem e com polimorfismo K232A em bovinos *Bos taurus* demonstrou uma alta estimativa de meia vida para ambas as sequências. Entretanto, a presença de um espiral na estrutura secundária e de uma desordem no critério Hot-Loops foi observada exclusivamente na proteína DGAT-1 não polimórfica. Os resultados revelaram elevado grau de conservação e estreita proximidade evolutiva entre diferentes espécies de ruminantes, bem como alterações na estrutura secundária e na predição de desordens da DGAT-1 sem o polimorfismo K232A, o que poderia estar comprometendo a ação enzimática e afetando o metabolismo biológico nos animais que não apresentam tal polimorfismo.

Palavras – chave: Bioinformática. Evolução. Polimorfismo. Proteína DGAT-1. Ruminantes.

ABSTRACT

The enzyme acyltransferase-1 diacylglycerol (DGAT-1) plays a key role during fat biosynthesis in milk. Amino acid sequences of the DGAT-1 protein were selected from protein database Swiss-Prot, such as: species *Bubalus bubalis* (tr | Q0ZR39 | Q0ZR39, 489), *Bos indicus* (tr | A6YP72 | A6YP72, 425), *Ovis aries* (tr | A8VJM4 | A8VJM4, 489), *Capra hircus* (tr | A2T092 | A2T092, 382), *Bos taurus* (non polymorphic sequence for the K232A polymorphism, sp | Q8MK44 | . 1, 489). Such sequences were employed for inferring phylogenetic relationships among different species of ruminants as well as for predicting biochemical characteristics between the sequences K232A non-polymorphic and polymorphism DGAT-1 in cattle *Bos taurus*. Multiple alignments from the selected sequences were performed to identify the degree of identity this enzyme from Multialin and Swiss-Prot programs. For phylogenetic inference and evaluation of the evolutionary distance between species, an evolutionary tree was constructed using MEGA 6.0 program using the Neighbor-Joining method (NJ). The prediction of biochemical parameters, the secondary structure and disorders of the protein DGAT-1 with and without polymorphism K232A were performed using tools of ProtParam, CFSSP Chou and Fasman and DisEMBL programs, respectively. Multiple alignments of the sequences of ruminant species showed a 77.50% identity and E value of 0.003. The evolutionary grouping of the species revealed the presence of several sub-branches grouped in an one main branch. Biochemical characterization of the 1 non-polymorphic and polymorphic DGAT- from K232A mutation in cattle showed a high estimated half-life for both sequences. Moreover, the presence of a spiral on the secondary structure and a disorder in the hot-loops criterion were observed exclusively in the non-polymorphic DGAT-1 protein. These results revealed a high degree of conservation and close evolutionary proximity between different species of ruminants, as well as changes in secondary structure and predicting disorders in the non-polymorphic DGAT-1, which could be compromising the enzymatic action and affecting the biological metabolism in animals that do not exhibit this polymorphism.

Key - words: Bioinformatics. Evolution. Polymorphism. DGAT-1 protein. Ruminants.

LISTA DE ILUSTRAÇÕES

Figura 1:	Estrutura tridimensional da lisina (A) e alanina (B)	17
Figura 2:	Síntese de triacilgliceróis	18
Figura 3:	Alinhamento comparativo das seis espécies de ruminantes utilizando o programa MultiAlin	32
Figura 4:	Alinhamento comparativo das sequências da DGAT-1 em espécies de ruminantes utilizando o programa Swiss – Prot	34
Figura 5:	Alinhamento comparativo sequências completas das espécies utilizando o programa Swiss – Prot	35
Figura 6:	Árvore filogenética representando a relação evolutiva entre espécies de ruminantes a partir das sequências da proteína DGAT-1	37
Figura 7:	Estimativas de divergência evolutiva entre as espécies selecionadas.....	38
Figura 8:	Estrutura secundária da proteína DGAT-1 de bovino <i>Bos taurus</i> : (A) sem e (B) com polimorfismo	39
Figura 9:	Representação da desordem da proteína DGAT-1 sem polimorfismo (A) e com polimorfismo (B) da espécie <i>Bos taurus</i>	42

LISTA DE TABELAS

Tabela 1: Parâmetros bioquímicos da proteína DGTA-1 sem polimorfismo e com polimorfismo de bovino <i>Bos taurus</i>	40
--	----

LISTA DE ABREVIações E SIGLAS

Amin -	Aminoácidos Carregados Negativamente
Amin +	Aminoácidos Carregados Positivamente
DGAT-1	Diacilglicerol Aciltransferase-1
DNA	Ácido Desoxirribonucleico
GMH	Grande Média de Hidropaticidade
NJ	Neighbor Joining
pH	Potencial Hidrogeniônico
pI	Ponto Isoelétrico
PM	Peso Molecular
QTL	Loco de Característica Quantitativa
SNP	Polimorfismo de Nucleotídeo Único
T _{1/2}	Tempo de Meia Vida

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO DE LITERATURA	15
	2.1 Produção de ruminantes no Brasil	15
	2.2 A proteína diacilglicerol-aciltransferase-1 (DGAT-1)	16
	2.3 Polimorfismos nos segmentos gênicos da proteína DGAT-1	19
	2.4 Bioinformática	21
3	HIPÓTESES	26
4	OBJETIVOS	27
	4.1 Objetivo geral	26
	4.2 Objetivos específicos	26
5	MATERIAIS E MÉTODOS	28
	5.1 Buscas das sequências de interesse em banco de dados biológicos	28
	5.2 Alinhamentos das múltiplas sequências da proteína DGAT-1	28
	5.3 Construção da árvore evolutiva	29
	5.4 Caracterização da estrutura secundária da proteína DGAT-1 sem e com polimorfismo K232A de bovino <i>Bos taurus</i>	30
	5.5 Parâmetros bioquímicos da proteína DGAT-1 sem e com polimorfismo K232A de bovino <i>Bos taurus</i>	30
	5.6 Desordens da proteína DGAT-1 sem e com polimorfismo K232A de bovino <i>Bos taurus</i>	30
6	RESULTADOS	32
	6.1 Análises de alinhamentos múltiplos das sequências da proteína DGAT-1	32
	6.2 Relações evolutivas entre ruminantes a partir de análises das sequências de aminoácidos da proteína DGAT-1	35
	6.3 Caracterização da estrutura secundária da proteína DGAT-1 em bovino <i>Bos taurus</i> sem e com polimorfismo K232A	38
	6.4 Parâmetros bioquímicos da proteína DGAT-1 sem e com polimorfismo K232A	39
	6.5 Desordens bioquímicas da proteína DGAT-1 sem e com polimorfismo K232A em bovino <i>Bos taurus</i>	40

7	DISCUSSÃO	43
8	CONCLUSÃO	49
	REFERÊNCIAS BIBLIOGRÁFICAS	50
	ANEXOS	57
	APÊNDICE	66

1 INTRODUÇÃO

A produção de ruminantes é destaque no cenário do agronegócio brasileiro, sendo uma importante atividade econômica e social. A população de ruminantes é vista como uma boa possibilidade de atividade rentável para a região semiárida brasileira, destacando-se como uma melhoria para a pecuária do Nordeste (ROBERTO, 2011).

O segmento gênico diacilglicerol-aciltransferase-1 (*dgat-1*) caracteriza-se como um potencial gene candidato para a síntese de gordura no leite. É codificante da enzima microsossomal diacilglicerol-aciltransferase-1 (DGAT-1), sendo esta constituída por 489 resíduos de aminoácidos, expressando-se em muitos tecidos, como o tecido epitelial, adiposo, glândulas mamárias e intestino, sendo estes os locais de maior expressão dessa proteína (CASES *et al.*, 1998; MILLAR; SMITH; KUNST, 2002). Em razão de sua função fisiológica, este gene tem despertado grande interesse para o melhoramento de rebanhos leiteiros, sendo necessária a realização de constantes estudos para elucidar sua estrutura e manifestação da expressão dessa proteína nas espécies leiteiras.

Um estudo detalhado deste gene associado a características de interesse econômico tem sido potencialmente incrementado graças ao advento da biologia molecular, permitindo grandes avanços em programas de melhoramento genético. Nesse contexto, a Bioinformática, uma importante sub-área da biologia molecular, tem permitido obter informações detalhada de biomoléculas através de análises computacionais. A aplicação desta tecnologia vem proporcionando um melhor entendimento e desenvolvimento de estudos biomoleculares, gerando relevantes informações de qualidade, com grande precisão e em reduzido tempo sobre biomoléculas de interesse.

Os marcadores moleculares são regiões no DNA em que se encontram polimorfismos, e estão fisicamente ligados a locos que determinam características de interesse, estando associados às características quantitativas e qualitativas de interesse econômico (MARIN *et al.*, 2005). O estudo de polimorfismos responsáveis por manifestações de características importantes, dentre elas a detecção de animais diferenciados quanto à produção de leite, assegura avanços em programas de melhoramento genético animal. Os polimorfismos são alterações no genoma dando

uma ou mais formas a molécula, podendo essas alterações trazerem benefícios para as espécies (MAKI, 2002). Polimorfismos detectados nas sequências gênicas que codificam a proteína DGAT-1 foram reconhecidos por apresentarem grandes efeitos sobre a composição da gordura no leite (GRISARD *et al.*, 2004).

O polimorfismo de uma única base (SNP) compreende alterações pontuais na sequência do DNA, capazes de modificar apenas uma base nitrogenada, sendo este o tipo mais encontrado de mutação na maioria das espécies e seu efeito pode estar relacionado a melhorias na produtividade. Os SNPs compreendem a substituição de um nucleotídeo por outro e para que a variação seja considerada um SNP, é necessário que o alelo menos frequente apresente frequência igual ou superior a 1% na população em estudo (BOREM; CAIXETA, 2006). Estudos realizados por Grisart *et al.* (2002) diagnosticaram mutações na posição 232, sendo tal alteração responsável pela substituição dos aminoácidos lisina (K) / alanina (A) e altamente desejada devido a seu efeito significativo na produção e composição do leite. De acordo com Lacorte *et al.* (2006), a substituição não conservadora de lisina por alanina (DGAT-1 K232A) pode ser observada como consequência da substituição das bases nitrogenadas adenina/adenina para guanina/citosina na posição 10.433 e 10.434 presentes na região codificante (éxons) VIII em bovinos europeus. Há fortes evidências de que o alelo K tem relação positiva sobre a funcionalidade da proteína DGAT-1, demonstrando que a mutação K232A aumenta a atividade da enzima em questão (GRISART *et al.*, 2004) Assim sendo, a proteína DGAT-1 apresenta grande importância econômica e análises investigativas visando à caracterização molecular de suas sequências devem proporcionar subsídios teóricos aos estudos de conservação de biomoléculas de interesse econômico.

2 REVISÃO DE LITERATURA

2.1 Produção de ruminantes no Brasil

A produção animal, notadamente de ruminantes, assume destaque no cenário brasileiro, visto que o país apresenta potencial de extensão territorial, o que contribui para bons resultados nessa prática. A criação de bovinos, búfalos, ovinos e caprinos é praticada ao longo de todo território brasileiro. O Brasil representa no mundo o segundo maior rebanho de bovinos, ficando atrás apenas da Índia (LIVESTOCK, 2013) e em se tratando de produção de leite desponta na quinta posição mundial (FAO, 2010).

Por estar em expansão no Brasil, a bubalinocultura é uma boa alternativa econômica, sendo destaque na região Norte, onde se concentra o maior rebanho nacional, seguida da região Sul, Sudeste, Nordeste e Centro-Oeste (IBGE, 2006). A exploração desses animais para a produção de leite destacou-se a partir dos anos 80 devido ao seu leite apresentar características físico-químicas de destaque em relação às outras espécies (BASTIANETTO, 2009).

Já a caprinovinocultura tem destaque na região Nordeste e Sul do Brasil. O rebanho de ovino teve o seu crescimento estimado em 21% entre 2002 e 2013, tendo maior representatividade nos estados da Bahia, Ceará, Piauí e Pernambuco (IBGE, 2013).

O Nordeste Brasileiro ocupa uma área de aproximadamente 1.640.000 km², correspondendo 19,9% do território nacional (ARAÚJO FILHO; CRISPIM, 2002). Nessa região, a produção de grandes e pequenos ruminantes caracteriza-se como atividade de grande importância econômica e social (COSTA, *et al.*, 2008).

Na região Nordeste a produção de caprinos e ovinos está distribuída nos estados da Bahia (38,28%), Piauí (16,03%), Ceará (15,28%), Pernambuco (13,72%), seguido dos demais estados nordestinos (CAMPOS; CAMPOS, 2013). A produção de ovinos foi destaque nos estados da Bahia, Maranhão, Paraíba, Sergipe. Os estados de Ceará, Pernambuco e Piauí registraram queda desta espécie, porém, o quantitativo desta espécie é insuficiente para impedir o crescimento regional (IBGE, 2013).

No estado de Pernambuco, a pecuária tem destaque na produção de caprinos, ovinos e bovinos, desempenhando um papel importante na economia (SAMPAIO *et al.*, 2009).

O panorama das espécies ruminantes na região semiárida tem demonstrado forte perspectiva de crescimento, uma vez que tais espécies apresentam excelente capacidade adaptativa às condições climáticas da região.

2.2 A proteína diacilglicerol-aciltransferase-1 (DGAT-1)

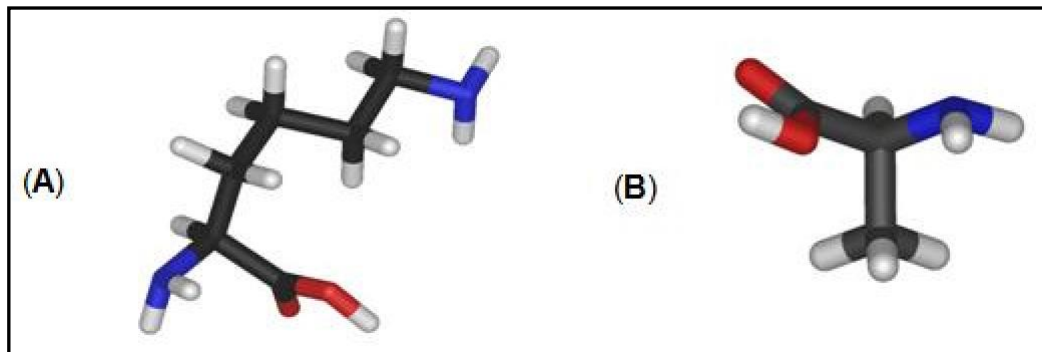
As proteínas são macromoléculas constituídas por polímeros de aminoácidos, essenciais e não essenciais. Por serem componentes cruciais na regulação metabólica dos seres vivos, estudos visando à compreensão da sua estrutura representam peça chave para decifrar suas funções. Dentre os 20 tipos de aminoácidos constituintes de proteínas, a lisina (K) e alanina (A) são destaques na síntese dos lipídios (VOET; VOET; PRATT, 2000).

A lisina é um aminoácido essencial com cadeia lateral polar, sendo altamente hidrofílica, cuja estrutura tridimensional está representada na (Figura 1A). A alanina é um aminoácido não essencial, hidrofóbico e apresenta cadeia lateral apolar, dobramento espacial por interações fracas, representada na (Figura 1B). Estes aminoácidos, por apresentarem características distintas, desempenham papel particular na determinação da estrutura tridimensional da proteína (NELSON; COX, 2011). Alterações na composição da proteína, por substituição de um desses aminoácidos, podem afetar a estrutura vital de seu dobramento tridimensional e conseqüentemente, afetar características estruturais e funcionais (EKERLJUNG, 2012). A estrutura primária de uma proteína é o elemento mais importante, pois esta estrutura determina sua forma e função. Se a estrutura primária de uma proteína for alterada por uma mutação, como por exemplo, a mudança de um aminoácido, a proteína poderá apresentar características estruturais e funcionais modificadas (SINHA; TREHAN, 2003). As interações moleculares de dobramento entre os aminoácidos da cadeia primária, formam um arranjo dando origem a estrutura secundária, que esta por sua vez, descreve um arranjo de interação dando origem a padrões estruturais recorrentes. As proteínas apresentam um mecanismo de dobramento específico e erros afetam sua atividade biológica. Para que elas funcionem corretamente é necessário que encontrem condições ambientais

favoráveis, tais como: temperatura, pH. Como uma proteína é formada por aminoácidos carregados de forma diferente, existe um valor de pH no qual a carga total de aminoácidos da proteína é equilibrada, denominado ponto isoelétrico. Neste pH, a proteína tem carga nula e a repulsão entre as moléculas diminui, propiciando suas interações (NELSON;COX, 2011).

A meia vida de uma proteína refere-se ao tempo em que ela desenvolve suas funções e adquire estabilidade, que é influenciada pela natureza dos seus aminoácidos constituintes (EISMANN, 2013). Qualquer alteração na sequência dos aminoácidos da proteína poderá gerar características bioquímicas desejáveis, como aumento na produtividade e melhoria no produto final.

Figura 1 - Estrutura tridimensional da lisina (A) e alanina (B).



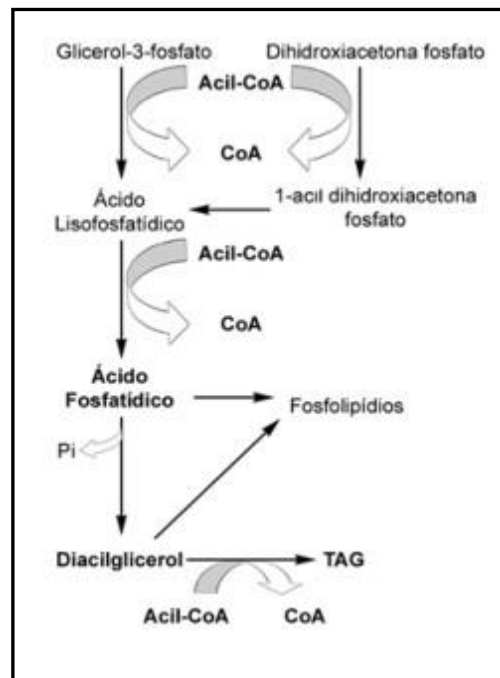
Fonte: adaptado de VOET; VOET; PRATT, 2000.

Em 1998 foi identificada por Cases *et al.* (1998) que o gene *dgat-1* codifica a enzima diacilglicerol aciltransferase -1, uma proteína ligada a uma membrana do retículo endoplasmático, capaz de catalisar o passo final na síntese dos lipídios, desempenhando papel fundamental no metabolismo dos lipídios, incluindo a produção de gordura e percentual do leite. Esta enzima é constituída por 489 aminoácidos, sendo expressa em diversos tecidos, dentre eles: fígado, tecido adiposo e epitelial, intestino, glândulas mamárias (Cases *et al.*, 2001).

Análises bioquímicas mostram que o triacilglicerol é sintetizado no retículo endoplasmático por duas vias (TANTIA *et al.*, 2006; SILVA *et al.*, 2011). A primeira envolve a transferência de acil-CoA independente do ácido graxo e de fosfolipídeos para a posição 3-diacilglicerol para a formação do triacilglicerol. Esta reação é catalisada pela enzima: diacilglicerol aciltransferase. A síntese do triacilglicerol consiste-se inicialmente na esterificação de ácidos graxos utilizando uma molécula

de glicerol. Ocorrem a transferência de três acilas e uma desfosforilação ao longo da via, empregando, três moléculas de ácidos graxos. A primeira acilação pode empregar como aceptor o glicerol-3-fosfato ou dihidroxiacetona fosfato, dando origem a ácido lisofosfatídico. Este sofrerá nova acilação gerando ácido fosfatídico. O ácido fosfatídico é desfosforilado, gerando diacilglicerol, substrato da última acilação da via. O diacilglicerol junto com acil CoA participa da conversão do triacilglicerol (Figura 2) (VOET; VOET; PRATT, 2000; SHOCKEY *et al.*, 2006). Sua atividade enzimática é controlada tanto na transcrição como na pós-transcrição da proteína (FURBASS *et al.*, 2006; KUHN *et al.*, 2004).

Figura 2 - Síntese de triacilgliceróis



Fonte: adaptado de LODISH *et al.*, 2014

Por aumentar a síntese de triacilglicerol na composição do leite, esta enzima desempenha papel essencial no controle quantitativo e qualitativo na deposição dos ácidos graxos, apresentando um grande interesse para os criadores de rebanho leiteiro.

2.3 Polimorfismos nos segmentos gênicos da proteína DGAT-1

Polimorfismos são alterações na sequência de alelos, na sequência de bases ou na estrutura cromossômica, com frequência superior a 1% na espécie (STRACHAN; READ, 2002). O estudo de polimorfismos é realizado com o intuito de associá-los indivíduos geneticamente superiores, pois eles podem estar relacionados às diferenças positivas nas propriedades físico-químicas de um determinado produto, apresentando um impacto positivo no melhoramento animal, principalmente em características econômicas para os pecuaristas.

Os polimorfismos de uma única base (SNPs) são modificações na sequência do DNA que alteram apenas uma base, ou seja, são polimorfismos de um único nucleotídeo, em que a base nitrogenada de um nucleotídeo é trocado por outra qualquer (G/T, A/T, G/C, A/G, A/C) e vice versa, em uma porção significativa de mais de 1% de uma população (PENNISI, 1998). As formas mais comuns de variações na sequência do DNA normalmente são bialélicas, encontrados em apenas duas variantes em uma espécie. De acordo com Brookes (1999), os SNPs são os polimorfismos mais frequentes no DNA, ocorrendo a cada 300 – 600 nucleotídeos, e sendo encontrados nas regiões codificantes (éxons) e não codificantes (íntrons) do genoma.

Por apresentar frequência elevada no genoma, os SNPs são de grande utilidade para as pesquisas, especialmente quando encontradas em genes relacionados às características de interesse econômico. Os SNPs em uma dada sequência podem alterar a formação de proteínas e possivelmente essa alteração pode provocar variações de características individuais (ARBEX, 2007). São caracterizados como sendo causadores de características de interesse na produção do leite, em particular aos níveis de gordura, sendo este constituinte de grande valor comercial. Por ser apontado como causador desta característica, são ótimos auxiliares para o desenvolvimento de programas de melhoramento genético, tendo como sua característica de grande importância econômica na espécie.

Um gene candidato apresenta sequência e ação biológica conhecida para uma determinada característica de interesse (FERREIRA; GRATTAPAGLIA, 1998). Um gene, para ser candidato funcional desempenha efeito sobre uma função metabólica e manifesta alguma característica de interesse (GLAZIER *et al.*, 2002). O gene *dgat-1* é conhecido como candidato funcional e posicional para o conteúdo de

gordura no leite, sendo o primeiro gene com efeito comprovado cientificamente para a produção de gordura no leite. A primeira evidência de localizações cromossômicas responsáveis por características de importância econômica foi relatada por Gelderman (1975). Estudos demonstram que na região cromossômica 14, existe grande impacto para a produção de leite, em particular aos níveis de gordura, sendo de grande interesse para o melhoramento de rebanhos leiteiros. Esta evidência é relacionada devido à posição próxima a um QTL (Locus de Característica Quantitativa) (SPELMAN *et al.*, 2002, BENNEWITZ *et al.*, 2004).

No gene *dgat-1* foi identificado polimorfismo nas raças Sindi, Gir, Holandês e nos cruzamentos Holandês x Gir (LARCORTE *et al.*, 2006), a frequência do alelo A foi encontrado em 73% em animais da raça Holandesa, seguida por Holandês x Gir. Polimorfismos encontrados no gene *dgat1* alteram a eficácia da enzima quando converte diglicerídeos em triglicerídeos na deposição de gordura em animais. De acordo com Smith *et al.* (2000), a ausência do gene *dgat-1* em algumas fêmeas, afeta a síntese de triacilglicerídeos, contribuindo para redução e/ou extinção da produção de leite pelo animal.

Winter *et al.* (2004) observaram a presença de 19 polimorfismos por análises de sequenciamento do gene *dgat-1*. Foram encontrados dois polimorfismos na região promotora do gene responsável pela deposição de gordura no leite, sendo um deles o número variável de repetições em tandem (VNTR) na posição 1.465. Ainda, uma sequência de 18 nucleotídeos (AGGCCCGCCCTCCCCGG) foi encontrada em determinados rebanhos bovinos, o que ocasiona grande influência na concentração da gordura no leite (WINTER *et al.*, 2004; THALLER *et al.* 2003; KUHN *et al.*, 2004.). Ao estudar esse gene em caprinos, Sousa *et al.* (2008) encontraram a mesma sequência de 18 nucleotídeos. Angiolillo *et al.* (2007) sequenciaram um fragmento genômico de 1 kb (do exon 12 ao 17) em cabra e revelaram a existência de um polimorfismo de nucleotídeo único, envolvendo a substituição da base nitrogenada T por C no íntron 16, e concluíram que estudos com associação para detectar polimorfismos adicionais na DGAT-1 podem influenciar o conteúdo de gordura do leite e de outros traços de interesse econômico. Yuan *et al.* (2007), estudando polimorfismos em búfalos encontraram quatro repetições, porém o polimorfismo não estava associado a características de produção. Amorim *et al.* (2014), verificando variantes no gene *dgat-1* em búfalas da raça Murrah, encontraram dois polimorfismos na região do íntron 8 nas posições

g.276C>T e g.297G>T, porém futuros estudos devem ser conduzidos com um maior número de animais a fim de que se comprove uma efetiva associação com características de produção de gordura.

Liefers *et al.* (2002) em estudos com novilhas observaram alterações nos aminoácidos alanina por valina, na posição 140 do éxon 3, cujas características foram associadas com aumentos na produção do leite e porcentagem de gordura.

Grisart *et al.* (2002) identificaram mutações em diferentes posições do gene, sendo a substituição dos aminoácidos lisina/alanina na posição 232 da proteína (K232A) responsáveis por um efeito significativo na produção e composição do leite. O alelo A, fixado na maioria das raças zebuínas, está associado com o aumento na produção de leite e proteínas e um decréscimo na produção de gordura. O alelo K, com alta frequência em raças européias, está associado à diminuição da produção de proteína e associado à composição e a um alto teor de gordura no leite (VALENTE 2001). Estudos indicaram forte associação entre K232A e aumento na porcentagem de gordura do leite e redução da produção de leite (WINTER *et al.*, 2002; SPELMAN *et al.*, 2002; HORI-OSHIMA; BARRERAS-SERRANO, 2003).

Assim, torna-se evidente que o estudo dos polimorfismos em genes candidatos que codificam enzimas da biossíntese de gordura do leite permite obter melhores parâmetros de produtividade e qualidade em ruminantes leiteiros.

2.4 Bioinformática

A Bioinformática é uma área do conhecimento que surgiu em meados da década de 1990, utilizando a tecnologia da informação para o gerenciamento de dados biológicos, ou seja, é a ciência que utiliza as ferramentas da computação para resolver os problemas da biologia. É uma área multidisciplinar, em constante crescimento, que envolve conhecimento da engenharia de softwares, estatística, matemática, ciência da computação e biologia molecular (ESPINDOLA *et al.*, 2010). Surgiu da necessidade de se compreender aspectos estruturais e funcionais dos genes e de seus produtos finais. Tendo a biologia molecular como uma grande área de relação, compreende o armazenamento e dados biológicos com auxílio dos métodos computacionais, a fim de reconhecer padrões que provavelmente seriam impossíveis de serem analisados sem tal ferramenta. É uma área que foi impulsionada pelo grande aumento da quantidade de dados biológicos depositados

nos bancos públicos de dados de informação biológica. A utilização desta ferramenta permite o desenvolvimento de pesquisa de qualidade e precisão em reduzido tempo.

Um dos campos de maior destaque na utilização da Bioinformática vem sendo os estudos na área da genômica e proteômica, com o intuito de identificar genes e proteínas a partir da investigação de sequências de biomoléculas depositadas nos bancos de dados (PALLEN, 1999). Uma ferramenta muito utilizada na bioinformática são os bancos de dados. Nestes estão organizadas e estruturadas as informações de modo a facilitar consulta a sequências de interesse (PROSDOCIMI *et al.*, 2002). Dentre os bancos de dados disponíveis, GenBank/NCBI, PDB-Protein Data Bank, EMBL, BLAST, Gene Ontology, Swiss-prot. Clustal W, Multialin são ferramentas disponíveis para promover o alinhamento comparativo de múltiplas sequências.

Um dos maiores bancos de dados de sequências de proteína é o Swiss-Prot (BOECKMANN *et al.*, 2003), desenvolvido pelo Swiss Institute of Bioinformatics (LESK, 2008). Trata-se de um banco curado, no qual permite-se obter sequências de proteínas, e informações adicionais quanto à sua estrutura e similaridades com outras proteínas. Apresenta como vantagem a organização e definição da proteína, permitindo uma maior confiança na pesquisa (PROSDOCIMI, 2007). O Swiss-Prot provê registros para famílias e grupos de proteínas, sendo regularmente revisado para atualização dos registros de descobrimento de dados científicos.

Outra ferramenta de livre acesso é o MultiAlin, programa computacional bastante utilizado no alinhamento de sequências. Neste programa permite-se realizar uma análise comparativa com múltiplas sequências simultaneamente, de forma a se observar os níveis de identidade e o grau de similaridade (FILHO; ALENCASTRO, 2003), para que se possa obter relações funcionais, estruturais ou evolutivas.

Outro programa que permite alinhamento de sequências é o COBALT. Este permite realizar o alinhamento simultâneo de múltiplas sequências de proteínas. Esta ferramenta permitir a obtenção do valor de E gerado a partir do alinhamento das sequências em bancos de dados (PAPADOPOULOS; AGARWALA, 2007).

Atualmente, a bioinformática é bastante utilizada para a construção de bibliotecas genômicas e bancos de dados, relação entre genes, análises de sequências, construções de árvores filogenéticas e predição da configuração tridimensional de proteínas.

A filogenia é definida como uma indicação das supostas relações de ancestralidade em um conjunto de espécies (MIYAKI *et al.*, 2001). O estudo das relações entre os organismos utilizando dados moleculares, tais como: sequências de aminoácidos de proteína, parte de um pressuposto de homologia entre as espécies, presente em um ancestral comum a elas (RUSSO, 2001).

As árvores filogenéticas são construídas a partir de uma matriz que contenha dados de informações sobre a morfologia, bioquímica, genes ou sequências de proteínas disponíveis de determinada espécie, inferindo relações evolutivas entre os organismos. Tendo como base o ancestral comum, é adotada como um esquema representativo da relação evolutiva, baseada na semelhança das sequências, expressando similaridade, ancestralidade ou parentesco evolutivo (MATSUDA *et al.*, 1999; YANG, 2006).

A árvore filogenética consiste basicamente na separação das espécies em grupos correspondentes aos sub-ramos que compartilham características em comum. As ramificações das árvores filogenéticas representam a distância evolutiva e os comprimentos dos ramos podem representar as estimativas de tempo evolutivo, baseada nas similaridades das sequências. Assim, as sequências mais próximas apresentam mais características em comum do que as mais distantes.

Dentre outra potencialidade da bioinformática destaca-se a utilização de programas computacionais para predição de estruturas secundárias das proteínas. O CFSSP: Chou & Fasman Secondary Structure Prediction Server é um servidor online e de livre acesso para a predição de estrutura secundária de proteína (KUMAR, 2013). Este servidor prevê regiões de estrutura secundária como alfa hélice, folhas beta e espiral a partir de sequências de aminoácidos de proteína. O método de CFSSP foi um dos primeiros algoritmos elaborados para a predição da estrutura secundária e baseia-se predominantemente em parâmetros de probabilidade a partir de frequências de aminoácidos em cada tipo de estrutura secundária (CHOU; FASMAN, 1974).

O ProtParam é uma ferramenta da bioinformática que calcula características bioquímicas para uma dada proteína armazenada em banco de dados, tais como: Swiss/Prot ou TrEMBL, sendo acessível através do servidor ExPASy (<http://web.expasy.org/protparam/>). A proteína pode ser especificada com o número de acesso Swiss – Prot / TrEMBL ou ID, ou na forma de uma sequência de aminoácidos. Dentre os parâmetros calculados pelo ProtParam incluem: a) PONTO

ISOELÉTRICO, onde o número de cargas positivas e negativas da proteína iguale-se e a molécula passar a ter uma carga líquida neutra (KNEZEVIC *et al.*, 2004); b) PESO MOLECULAR calculado com base no somatório das massas moleculares de cada um dos aminoácidos componentes da sequência analisada, sendo medidos em Dalton (Da); c) ESTIMATIVA DE MEIA-VIDA, que corresponde ao tempo em que metade da concentração de uma proteína será desfeita por processo bioquímico (EISMANN, 2013); d) ÍNDICE DE INSTABILIDADE que fornece uma estimativa da estabilidade da proteína, onde valores menores que 40 indicam estabilidade e superiores a 40 sugerem instabilidade da proteína (GASTEIGER *et al.*, 2005); e) ÍNDICE ALIFÁTICO, determinação do volume relativo ocupado pelas cadeias laterais alifáticas (alanina, valina, isoleucina e leucina). Proteínas com índice alifático elevado tendem a ser mais termoestáveis (GURUPRASAD *et al.*, 1990); f) GRANDE MÉDIA DE HIDROPATICIDADE que corresponde a análise quantitativa do grau de hidrofobicidade ou hidrofiliicidade dos aminoácidos de uma proteína, sendo usado para caracterizar ou identificar uma possível estrutura de domínio ou de uma proteína (KYTE; DOOLITTLE, 1982).

O DisEMBL é outra ferramenta da bioinformática acessível ao público utilizada para prever regiões de desordens em proteínas. Esta ferramenta computacional permite obter informações quanto às regiões desordenadas/não estruturadas dentro de uma sequência linear proteica, sendo útil para muitos estudos bioquímicos, em especial na biologia estrutural e projetos genômicos (LINDING *et al.*, 2003). A predição utilizando o DisEMBL baseia-se em três critérios característicos para a definição da desordem, dentre eles destacam-se: a) LOOPS/COILI, onde neste preditor, resíduos são classificados como pertencentes a algum tipo de região ordenada, caso estejam localizados em uma alfa hélice ou folha beta (TORRIERI, 2010). De acordo com Linding *et al.* (2003), coils (“espirais”) não são necessariamente desordenados; b) REMARK-465, onde resíduos podem estar dentro de uma região desordenada (TORRIERI, 2010). De acordo com Linding *et al.* (2003), os resíduos pertencentes a esta região podem não ter relação com a desordem estrutural, classificando este preditor como tendo uma alta taxa de falsos positivos; e finalmente c) HOT-LOOPS, onde resíduos são classificados com um alto grau de mobilidade determinado pelo fator de temperatura do carbono alfa, fornecendo informações importantes acerca da dinâmica da proteína. Este preditor

pode ser considerado como regiões estruturalmente desordenadas (TORRIERI, 2010).

3 HIPÓTESES

- As sequências de aminoácidos da proteína DGAT-1 de ruminantes são altamente conservadas.
- As espécies de ruminantes possuem proximidade em sua evolução.
- Polimorfismo pontual K232A na proteína DGAT-1 afeta a estrutura secundária e características bioquímicas dessa enzima em bovino *Bos taurus*.

4 OBJETIVOS

4.1 Objetivo geral

Avaliar sequências da proteína DGAT-1 de diferentes espécies de ruminantes para inferir relações filogenéticas e predizer características bioquímicas na proteína sem e com polimorfismo K232A em bovino *Bos taurus*, utilizando ferramentas de bioinformática.

4.2 Objetivos específicos

- Selecionar sequências de aminoácidos da proteína DGAT-1 de espécies de ruminantes em banco de dados;
- Identificar a identidade/similaridade entre espécies de ruminantes através de alinhamento de sequências de aminoácidos da proteína DGAT-1;
- Estabelecer relações filogenéticas entre diferentes ruminantes a partir de alinhamentos de sequências de aminoácidos da proteína DGAT-1;
- Prever parâmetros bioquímicos da proteína DGAT-1 sem e com polimorfismo K232A em bovino *Bos taurus*;
- Avaliar a estrutura secundária da proteína DGAT-1 sem e com polimorfismo K232A em bovino *Bos taurus*;
- Analisar possíveis desordens na proteína DGAT-1 sem e com polimorfismo K232A em bovino *Bos taurus*;

5 MATERIAIS E MÉTODOS

5.1 Buscas das sequências de interesse em banco de dados biológicos

As sequências de aminoácidos da proteína DGAT-1 nas espécies de ruminantes foram obtidas por meio de seleção e busca a partir de um banco de dados de informação biológica, denominado Swiss-Prot (<http://www.uniprot.org/uniprot/?query=DGAT1&sort=score>). Este banco permite obter dados apenas de sequências de proteínas que foram depositadas ao longo dos tempos, estando virtualmente disponível para o acesso público, além de apresentar alta integração com outros bancos. Armazena dados a partir do EMBL, tais como sequências de proteínas com grau alto de anotação, destacando função, estrutura e modificações pós-tradução (PROSDOCIMI, 2007).

Tal busca foi realizada informando o nome da proteína de interesse apenas em ruminantes domésticos na caixa de texto da página principal do programa Swiss-Prot (ANEXO 1). As sequências selecionadas representaram segmentos proteicos parciais ou totais da proteína DGAT-1 dos seguintes animais ruminantes disponíveis no banco de dados Swiss-Prot: búfalo *Bubalus bubalis* (tr|Q0ZR39|Q0ZR39, 489), boi *Bos indicus* (tr|A6YP72|A6YP72, 425), ovelha *Ovis aries* (tr|A8VJM4|A8VJM4, 489), cabra *Capra hircus* (tr|A2T092|A2T092, 382), boi *Bos taurus* (sequência não polimórfica para o polimorfismo K232A non-polymorphy (sp|Q8MK44|1, 489) (ANEXO 2).

5.2 Alinhamentos das múltiplas sequências da proteína DGAT-1

Um alinhamento representa a disposição de duas ou mais sequências, com demonstração gráfica através do seu pareamento, tornando de forma organizada as sequências de proteínas para identificação de regiões similares de relações funcionais ou estruturais. A realização de alinhamentos múltiplos é útil para analisar regiões que sofreram mutações, no estudo de estruturas secundárias das proteínas e geração de informações para construção de árvores filogenéticas (VIANA; MOURA, 2010).

Nas análises de alinhamento foram empregadas ferramentas do programa MultiAlin (<http://multalin.toulouse.inra.fr/multalin/>) e Swiss-Prot

(<http://www.uniprot.org/uniprot/?query=DGAT1&sort=score>). Este programa permite comparar duas ou mais sequências e estimar os valores de identidade e similaridade entre as sequências da proteína das diferentes espécies. O programa MultiAlin realiza um alinhamento heurístico, múltiplo e global entre sequências de proteínas (CORPET, 1988).

Tais alinhamentos foram possíveis através da inserção das seis sequências de interesse (búfalo, boi *Bos indicus*, ovelha, cabra, boi *Bos taurus* com e sem polimorfismo) na caixa de texto da página principal do programa MultiAlin (ANEXO 4), gerando uma interface ilustrativa e colorida referente ao alinhamento das sequências. No programa Swiss-Prot foi realizado o alinhamento apenas de cinco sequências (búfalo, boi *Bos indicus*, ovelha, cabra, boi *Bos taurus* (sem polimorfismo K232A), cujas sequências foram selecionadas no próprio programa.

O programa COBALT foi empregado com o objetivo de realizar alinhamentos simultâneos de sequências de proteínas a fim de gerar dados avaliativos, tais como o valor de E, indicando a significância do alinhamento. As sequências de interesse foram adicionadas na caixa de texto da página principal do programa (ANEXO 5). Baseado nos resultados gerados no alinhamento, valores de identidade e de similaridade foram estabelecidos entre as espécies.

5.3 Construção da árvore evolutiva

Uma árvore filogenética foi construída a partir das seis sequências de aminoácidos da proteína DGAT-1 das espécies de ruminantes por análises de agrupamentos pelo software MEGA versão 6.0 (TAMURA *et al.*, 2011). Nessas análises foram empregados apenas os segmentos comuns de aminoácidos, localizados entre os sítios 108 e 489 nas seis sequências selecionadas (ANEXO 3). Este programa permite realizar análises moleculares relacionadas com a filogenia, proporcionando ferramentas para explorar e analisar as sequências. O método utilizado para a construção da árvore filogenética foi o de Neighbor-Joining (NJ) representado pelas relações entre as sequências baseadas em sua distância, avaliando a matriz da distância produzida (TAMURA *et al.*, 2011).

Inicialmente, foi realizada a identificação das sequências fazendo-se o download dessas sequências para o programa MEGA, em seguida foi realizado o alinhamento das múltiplas sequências, utilizando o algoritmo Clustal W

disponibilizado pelo próprio programa, a fim de construir uma árvore filogenética, sendo escolhido o método NJ, que utiliza uma matriz de distâncias emparelhadas estimado no modelo de Jones-Thornton-Taylor (JTT) para sequências de aminoácidos (HALL, 2013).

5.4 Caracterização da estrutura secundária da proteína DGAT-1 sem e com polimorfismo K232A de bovino *Bos taurus*

Para obtenção da estrutura secundária da proteína DGAT-1 sem e com o polimorfismo K232A utilizou-se o programa CFSSP: Chou & Fasman Secondary Structure Prediction Server. Este programa prevê regiões de estrutura secundária como hélice, folha e espiral a partir de sequências de aminoácidos de uma proteína de interesse (KUMAR, 2013). Foram adicionadas as sequências de interesse no quadro da página principal do programa CFSSP Chou & Fasman (ANEXO 6) para obtenção do resultado.

5.5 Parâmetros bioquímicos da proteína DGAT-1 sem e com polimorfismo K232A de bovino *Bos taurus*

Para obtenção dos parâmetros bioquímicos da proteína DGAT-1 da espécie *Bos taurus* sem e com o polimorfismo K232A utilizou-se o programa ProtParam (<http://web.expasy.org/protparam/>). Foram obtidas informações quanto às características bioquímicas das proteínas, tais como: peso molecular, ponto isoelétrico (pI), composição dos aminoácidos, número de aminoácidos carregados negativamente e positivamente, estimativa de meia vida, índice de instabilidade e índice alifático (GASTEIGER *et al.*, 2005). Para obtenção destes dados adicionou-se a sequência de interesse no quadro da página principal do programa (ANEXO 7), obtendo-se, desta forma, os parâmetros desejados.

5.6 Desordens da proteína DGAT-1 sem e com polimorfismo K232A de bovino *Bos taurus*

A predição de desordens da proteína DGAT-1 em bovino *Bos taurus* sem e com o polimorfismo K232A foi realizada utilizando-se o programa DisEMBL™

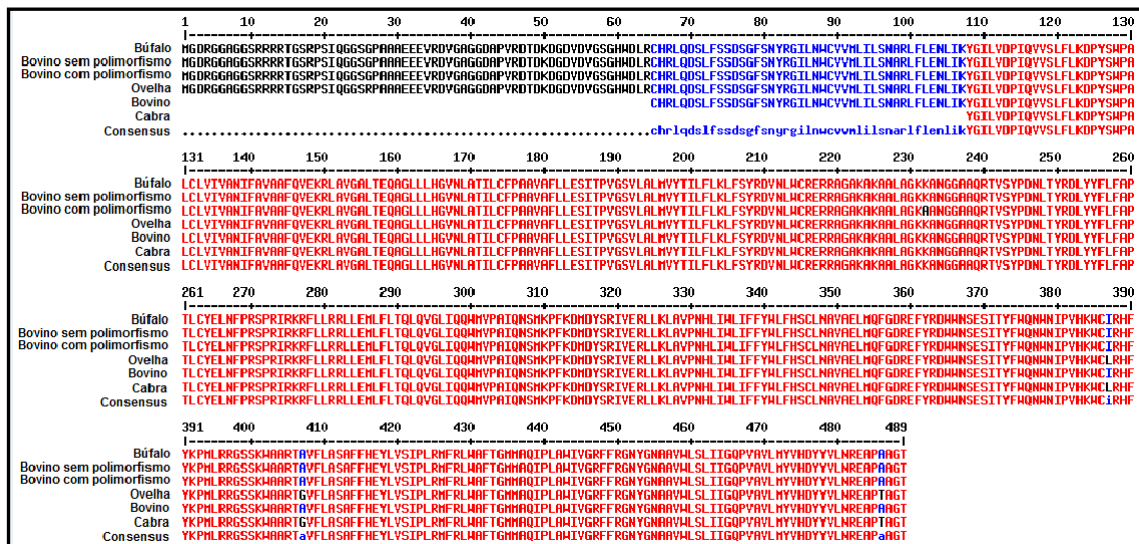
(DARNEL, 2003). Para obtenção dos dados referentes à desordem na proteína DGAT-1 sem e com o polimorfismo K232A em *Bos taurus* as sequências de interesse foram adicionadas individualmente na caixa de texto da página principal do programa (ANEXO 8), clicando no item “DisEMBL protein” e obtendo-se desta forma o resultado gerado através de gráfico, representando três possíveis regiões de desordens (Loops/Coil, Remark 465 e Hot-loops).

6 RESULTADOS

6.1 Análises de alinhamentos múltiplos das sequências da proteína DGAT-1

Os resultados gerados pelo alinhamento entre as seis sequências da proteína DGAT-1 das espécies de ruminantes estão representados na Figura 2, para o qual utilizou-se o programa MultiAlin. O resultado comparativo das seis sequências da proteína DGAT-1, sendo duas delas parciais (bovino *Bos indicus* e cabra), revelou uma identidade de 77,30% de um total de 489 aminoácidos, onde 378 aminoácidos foram representados como idênticos e quatro como similares (representados pela cor azul e preta) entre as espécies. Em tais análises, o segmento da sequência situado entre as posições 1 a 64 representado em preto, está presente apenas nas espécies de búfalo, bovino *Bos taurus* e ovelha (Figura 3). O segmento entre as posições 65 a 107, representado em azul corresponde a domínios da proteína presentes nas espécies de búfalo, bovino *Bos taurus*, ovelha e bovino *Bos indicus* (Figura 3). A região da proteína representada em vermelho (posições 108 a 489) é comum às seis espécies de ruminantes analisadas.

Figura 3 - Alinhamento comparativo da sequência de DGAT-1 das seis espécies de ruminantes.



Fonte: Dados gerados pelo programa MultiAlin.
<http://multalin.toulouse.inra.fr/multalin/cgi-bin/multalin.pl>

Na Figura 4 é apresentado o alinhamento comparativo das cinco sequências nas espécies de ruminantes, gerado automaticamente pelo programa Swiss-Prot, de onde foram selecionadas as sequências. Os resultados demonstrados na quarta e quinta linhas representam alinhamentos das sequências parciais de bovino *Bos indicus* e cabra respectivamente, sendo as demais linhas representadas por sequências totais da proteína DGAT-1 dos demais ruminantes. Nas análises, o símbolo asterisco (*), apresentado na linha inferior do alinhamento, indica que os aminoácidos presentes nestas espécies possuem máxima identidade entre as sequências selecionadas. Na Figura 4 pode ser observada ainda a presença de símbolos gerados automaticamente durante o alinhamento: dois pontos (:) e ponto (.), que correspondem a alterações de dois ou um na natureza do aminoácido nessa posição, respectivamente.

Com este resultado, pode-se observar uma alta identidade entre as sequências das diferentes espécies de ruminantes (Figura 4). No resultado comparativo das sequências, foi revelada uma identidade de 77,50% em um total de 489 aminoácidos, onde 379 aminoácidos estão representados como idênticos e três como similares entre as espécies.

Figura 4 - Alinhamento comparativo das seqüências da DGAT-1 em espécies de ruminantes utilizando o programa Swiss – Prot.

Búfalo	1	MGDRGGAGGSRRRRTGSRPSIQGGSGPAAAE E EVRDVVGAGGDAPVRDTRDKDGDVDVGS	60
Ovelha	1	MGDRGGAGGSRRRRTGSRPSIQGGSGPAAAE E EVRDVVGAGGDAPVRDTRDKDGDVDVGS	60
Bovino sem polimorfismo	1	MGDRGGAGGSRRRRTGSRPSIQGGSGPAAAE E EVRDVVGAGGDAPVRDTRDKDGDVDVGS	60
Bovino bos indicus	1	-----	0
Cabra	1	-----	0
Búfalo	61	WDLRCHRLQDSLFSDDSGFSNYRGILNWCVMHILSNARLFLENLIKYGILVDP IQVVS	120
Ovelha	61	WDLRCHRLQDSLFSDDSGFSNYRGILNWCVMHILSNARLFLENLIKYGILVDP IQVVS	120
Bovino sem polimorfismo	61	WDLRCHRLQDSLFSDDSGFSNYRGILNWCVMHILSNARLFLENLIKYGILVDP IQVVS	120
Bovino bos indicus	1	----CHRLQDSLFSDDSGFSNYRGILNWCVMHILSNARLFLENLIKYGILVDP IQVVS	56
Cabra	1	-----YGILVDP IQVVS	13
Búfalo	121	FLKDPYSWPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLHGVNLATI LCFPAAVA	180
Ovelha	121	FLKDPYSWPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLHGVNLATI LCFPAAVA	180
Bovino sem polimorfismo	121	FLKDPYSWPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLHGVNLATI LCFPAAVA	180
Bovino bos indicus	57	FLKDPYSWPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLHGVNLATI LCFPAAVA	116
Cabra	14	FLKDPYSWPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLHGVNLATI LCFPAAVA	73
Búfalo	181	FLLESITPVGSVLALMVYITILFLKLF SYRDVNLWCRERRAGAKAKAALAGK KANGGAAQR	240
Ovelha	181	FLLESITPVGSVLALMVYITILFLKLF SYRDVNLWCRERRAGAKAKAALAGK KANGGAAQR	240
Bovino sem polimorfismo	181	FLLESITPVGSVLALMVYITILFLKLF SYRDVNLWCRERRAGAKAKAALAGK KANGGAAQR	240
Bovino bos indicus	117	FLLESITPVGSVLALMVYITILFLKLF SYRDVNLWCRERRAGAKAKAALAGK KANGGAAQR	176
Cabra	74	FLLESITPVGSVLALMVYITILFLKLF SYRDVNLWCRERRAGAKAKAALAGK KANGGAAQR	133
Búfalo	241	TVSYPDNLT YRDLYYFLFAP T LCEYELNFRSPRIKRFLRRRLLEMLF LTQLQVGLIQQW	300
Ovelha	241	TVSYPDNLT YRDLYYFLFAP T LCEYELNFRSPRIKRFLRRRLLEMLF LTQLQVGLIQQW	300
Bovino sem polimorfismo	241	TVSYPDNLT YRDLYYFLFAP T LCEYELNFRSPRIKRFLRRRLLEMLF LTQLQVGLIQQW	300
Bovino bos indicus	177	TVSYPDNLT YRDLYYFLFAP T LCEYELNFRSPRIKRFLRRRLLEMLF LTQLQVGLIQQW	236
Cabra	134	TVSYPDNLT YRDLYYFLFAP T LCEYELNFRSPRIKRFLRRRLLEMLF LTQLQVGLIQQW	193
Búfalo	301	MVPAIQNSMKPFKDMYSRIVERLLK LAVPNHLIWL IFFYWL FHSCLNAVAELMQGDRE	360
Ovelha	301	MVPAIQNSMKPFKDMYSRIVERLLK LAVPNHLIWL IFFYWL FHSCLNAVAELMQGDRE	360
Bovino sem polimorfismo	301	MVPAIQNSMKPFKDMYSRIVERLLK LAVPNHLIWL IFFYWL FHSCLNAVAELMQGDRE	360
Bovino bos indicus	237	MVPAIQNSMKPFKDMYSRIVERLLK LAVPNHLIWL IFFYWL FHSCLNAVAELMQGDRE	296
Cabra	194	MVPAIQNSMKPFKDMYSRIVERLLK LAVPNHLIWL IFFYWL FHSCLNAVAELMQGDRE	253
Búfalo	361	FYRDWVNSE SITYFWQNNIPVHKWCI RHFYK PMLRRGSSKWAARTAVFLASAFFHEYLV	420
Ovelha	361	FYRDWVNSE SITYFWQNNIPVHKWCI RHFYK PMLRRGSSKWAARTAVFLASAFFHEYLV	420
Bovino sem polimorfismo	361	FYRDWVNSE SITYFWQNNIPVHKWCI RHFYK PMLRRGSSKWAARTAVFLASAFFHEYLV	420
Bovino bos indicus	297	FYRDWVNSE SITYFWQNNIPVHKWCI RHFYK PMLRRGSSKWAARTAVFLASAFFHEYLV	356
Cabra	254	FYRDWVNSE SITYFWQNNIPVHKWCI RHFYK PMLRRGSSKWAARTAVFLASAFFHEYLV	313
Búfalo	421	SIPLRMFRLWAF TGMMQIPLAWIVGR FFRGN YGNAAVVLS L IIGQPVAVLMYVHDYVVL	480
Ovelha	421	SIPLRMFRLWAF TGMMQIPLAWIVGR FFRGN YGNAAVVLS L IIGQPVAVLMYVHDYVVL	480
Bovino sem polimorfismo	421	SIPLRMFRLWAF TGMMQIPLAWIVGR FFRGN YGNAAVVLS L IIGQPVAVLMYVHDYVVL	480
Bovino bos indicus	357	SIPLRMFRLWAF TGMMQIPLAWIVGR FFRGN YGNAAVVLS L IIGQPVAVLMYVHDYVVL	416
Cabra	314	SIPLRMFRLWAF TGMMQIPLAWIVGR FFRGN YGNAAVVLS L IIGQPVAVLMYVHDYVVL	373
Búfalo	481	NREAPAAGT	489
Ovelha	481	NREAPTAGT	489
Bovino sem polimorfismo	481	NREAPAAGT	489
Bovino bos indicus	417	NREAPAAGT	425
Cabra	374	NREAPTAGT	382

Fonte: Dados gerados pelo programa Swiss - Prot.

(<http://www.uniprot.org/align/A2016060514483A1C7ED25EE8374758DF3FD545FD0A1F441>)

Entretanto, análise dos resultados encontrados a partir do alinhamento realizado exclusivamente a partir das seqüências completas da proteína DGAT-1 de búfalo, ovelha e bovino *Bos taurus*, depositada no banco de dados revelou um valor elevado de identidade (99,18%), com 485 aminoácidos em posições idênticas e três aminoácidos em posições similares e alteração de um ou dois aminoácidos, conforme pode ser observado na Figura 5.

Figura 5 - Alinhamento comparativo seqüências completas das espécies utilizando o programa Swiss – Prot.

Búfalo	1	MGDRGGAGGSRRRRTGSRPSIQGGSGPAAAE E E V R D V G A G G D A P V R D T D K D G D V D V G S G H	60
Ovelha	1	MGDRGGAGGSRRRRTGSRPSIQGGSGPAAAE E E V R D V G A G G D A P V R D T D K D G D V D V G S G H	60
Bovino sem polimorfismo	1	MGDRGGAGGSRRRRTGSRPSIQGGSRPAAAE E E V R D V G A G G D A P V R D T D K D G D V D V G S G H *****	60
Búfalo	61	WDLRCHRLQDSL FSSDSGFSNYRGILNWCVVMLILSNARLFLENLIKYGILVDPIQVVSL	120
Ovelha	61	WDLRCHRLQDSL FSSDSGFSNYRGILNWCVVMLILSNARLFLENLIKYGILVDPIQVVSL	120
Bovino sem polimorfismo	61	WDLRCHRLQDSL FSSDSGFSNYRGILNWCVVMLILSNARLFLENLIKYGILVDPIQVVSL *****	120
Búfalo	121	FLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGVNLTILCFPAAVA	180
Ovelha	121	FLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGVNLTILCFPAAVA	180
Bovino sem polimorfismo	121	FLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGVNLTILCFPAAVA *****	180
Búfalo	181	FLLESITPVGSLALMVYITILFLKLFYSYRDVNLWCERRRAGAKAKAALAGKKANGGAAQR	240
Ovelha	181	FLLESITPVGSLALMVYITILFLKLFYSYRDVNLWCERRRAGAKAKAALAGKKANGGAAQR	240
Bovino sem polimorfismo	181	FLLESITPVGSLALMVYITILFLKLFYSYRDVNLWCERRRAGAKAKAALAGKKANGGAAQR *****	240
Búfalo	241	TVSYDNLTYRDLYYFLFAPLTCYELNFPSPRIRKRFLLRRLLEMLFLTQLQVGLIQQW	300
Ovelha	241	TVSYDNLTYRDLYYFLFAPLTCYELNFPSPRIRKRFLLRRLLEMLFLTQLQVGLIQQW	300
Bovino sem polimorfismo	241	TVSYDNLTYRDLYYFLFAPLTCYELNFPSPRIRKRFLLRRLLEMLFLTQLQVGLIQQW *****	300
Búfalo	301	MVPAIQNSMKPFKMDYSRIVERLLKLA VPNHLIWLIFFYWLFHSCLNVAELMQFGDRE	360
Ovelha	301	MVPAIQNSMKPFKMDYSRIVERLLKLA VPNHLIWLIFFYWLFHSCLNVAELMQFGDRE	360
Bovino sem polimorfismo	301	MVPAIQNSMKPFKMDYSRIVERLLKLA VPNHLIWLIFFYWLFHSCLNVAELMQFGDRE *****	360
Búfalo	361	FYRDWNSSEITYFWQNWNI PVHKWCI RHFYKPLRRGSSKWAARTAVFLASAFFHEYLV	420
Ovelha	361	FYRDWNSSEITYFWQNWNI PVHKWCI RHFYKPLRRGSSKWAARTAVFLASAFFHEYLV	420
Bovino sem polimorfismo	361	FYRDWNSSEITYFWQNWNI PVHKWCI RHFYKPLRRGSSKWAARTAVFLASAFFHEYLV *****	420
Búfalo	421	SIPLRMFRLWAF TGMMQIPLAWIVGRFFRGNYGNAAVWLSLII GQPVAVLMYVHDYYVL	480
Ovelha	421	SIPLRMFRLWAF TGMMQIPLAWIVGRFFRGNYGNAAVWLSLII GQPVAVLMYVHDYYVL	480
Bovino sem polimorfismo	421	SIPLRMFRLWAF TGMMQIPLAWIVGRFFRGNYGNAAVWLSLII GQPVAVLMYVHDYYVL *****	480
Búfalo	481	NREAPAAGT	489
Ovelha	481	NREAPAAGT	489
Bovino sem polimorfismo	481	NREAPTAGT *****	489

Fonte: Dados gerados pelo programa Swiss - Prot.

(<http://www.uniprot.org/align/A2016060514483A1C7ED25EE8374758DF3FD545FD0A1F50A>)

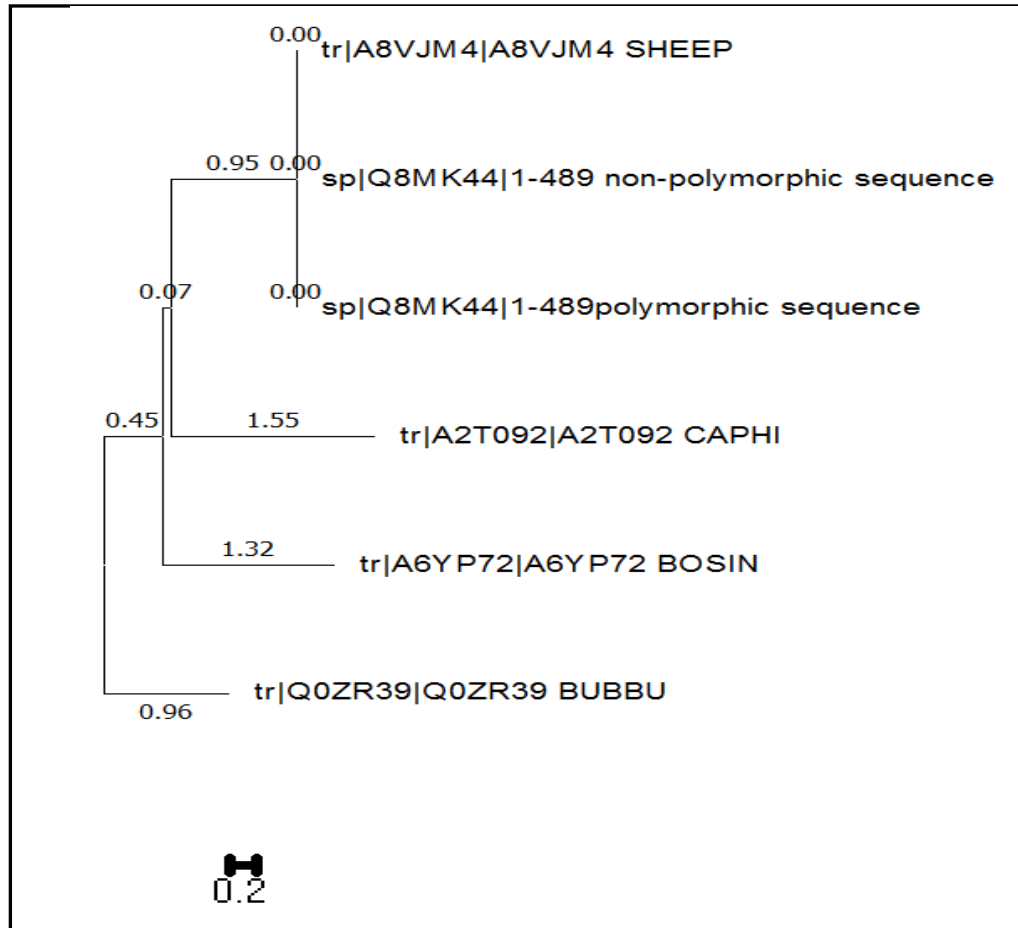
Em todas essas avaliações comparativas foram obtidos valores de E de 0.003, o que indica que os dados obtidos no alinhamento foram significativos. De acordo com LESK (2008), valores de $E \leq 0.02$ sugerem, com alta significância, que tais seqüências são homólogas.

6.2 Relações evolutivas entre ruminantes a partir de análises das seqüências de aminoácidos da proteína DGAT-1

A Figura 6 representa a árvore filogenética proveniente das seis seqüências de aminoácidos da proteína DGAT-1 em ruminantes. A árvore foi obtida após o alinhamento múltiplo das seqüências da proteína DGAT-1 dos animais ruminantes, empregando o método de Neighbor-Joining para representação dos agrupamentos

entre as espécies. Na Figura 6 é possível detectar a presença de um ramo principal, constituído por quatro sub-ramos: o primeiro agrupou as espécies de ovelha (0,0), bovino *Bos taurus* com (0,0) e sem polimorfismos (0,0). Um segundo constituído pela sequência da espécie de cabra (1,55), um terceiro pela sequência de bovino *Bos indicus* (1,32) e um quarto sub-ramo proveniente da sequência da espécie de búfalo (0,96), provenientes dos segmentos comuns das seis sequências da proteína DGAT-1 das espécies de ruminantes. As espécies ovelha, bovino *Bos taurus* sem e com polimorfismo, por estarem agrupadas no mesmo sub-ramo, são espécies que possuem um alto grau de identidade e as demais, mesmo estando em sub-ramos diferentes, também apresentam alta identidade em relação às demais, visto que foram provenientes da mesma ramificação principal. Na Figura 6, observa-se a formação de ramos e nós terminais que representam a estimativa do tempo evolutivo e a unidade taxonômica, respectivamente.

Figura 6 - Árvore filogenética representando a relação evolutiva entre seis espécies de ruminantes a partir das sequências da proteína DGAT-1.



Fonte: Figura gerada pelo programa MEGA 6.0.

As estimativas de divergência evolutiva entre as seis sequências da proteína DGAT-1 estão representadas na Figura 7. As sequências de búfalo e bovino *Bos indicus* foram as mais diferenciadas com uma taxa de 2,729, enquanto que as sequências da proteína de búfalo e bovino *Bos taurus* com polimorfismo foram as menos diferenciadas evolutivamente com taxa de 0,004.

Figura 7 - Estimativas de divergência evolutiva entre as espécies selecionadas.

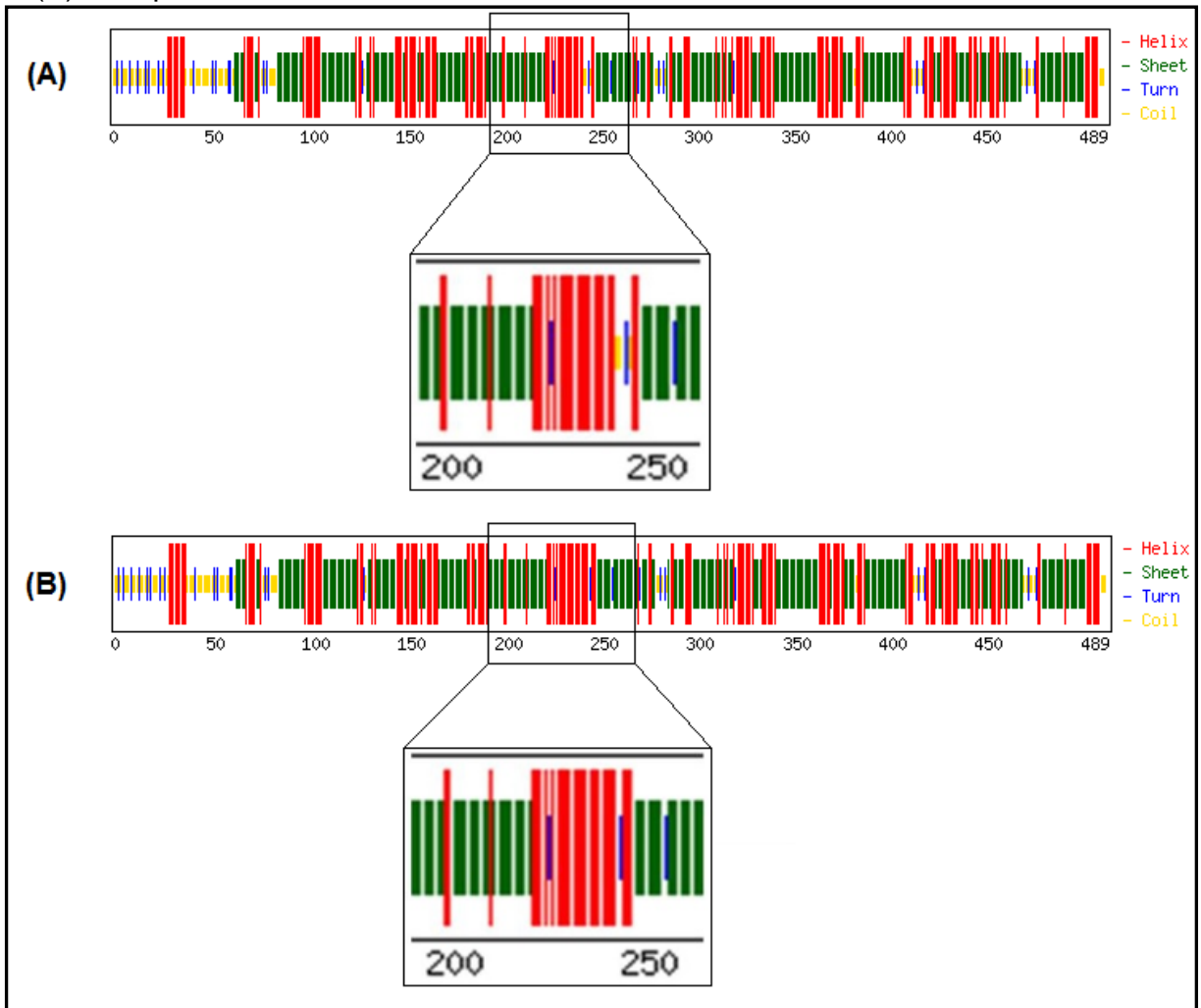
Búfalo Q0ZR39					
Bovino <i>Bos indicus</i> A6YP72	2.729				
Ovelha A8VJM4	2.410	2.366			
Cabra A2T092	3.103	2.862	2.506		
Bovino <i>Bos taurus</i> sem polimorfismo Q8MK44	2.410	2.366	0.000	2.506	
Bovino <i>Bos taurus</i> com polimorfismo Q8MK44	2.410	2.366	0.004	2.506	0.004

Fonte: Dados gerados pelo programa MEGA 6.

6.3 Caracterização da estrutura secundária da proteína DGAT-1 em bovino *Bos taurus* sem e com polimorfismo K232A

As Figuras 8A e 8B representam a estrutura secundária da proteína DGAT-1 de bovino *Bos taurus* sem e com polimorfismo, respectivamente. A barra azul representa curva (“Turn”), a amarela, espiral (“Coil”), a verde, folha (“Sheet”) e a vermelha representa a hélice (“Helix”). Nesse resultado foi observada a presença de uma espiral (barra amarela na Figura 8A) entre as posições 200 e 250 na proteína não polimórfica, não observada naquela com polimorfismo K232A (Figura 8B). Análise da estrutura secundária dos demais segmentos da proteína DGAT-1 sem e com polimorfismo não revelou a existência de outras diferenças.

Figura 8 - Estrutura secundária da proteína DGAT-1 de bovino *Bos taurus*: (A) sem e (B) com polimorfismo.



Fonte: Figura gerada pelo programa CFSSP Chou e Fasman.
(<http://cho-fas.sourceforge.net/index.php>)

6.4 Parâmetros bioquímicos da proteína DGAT-1 sem e com polimorfismo K232A

Os parâmetros bioquímicos da proteína DGAT-1 de bovino *Bos taurus* sem e com polimorfismo K232A foram obtidos utilizando o programa ProtParam (Tabela 2).

Tabela 1 - Parâmetros bioquímicos da proteína DGTA-1 sem polimorfismo e com polimorfismo de bovino *Bos taurus*

	pl	PM	Amin. -	Amin. +	T _{1/2}	Índice instabilidade	Índice alifático	GMH
Bovino <i>Bos taurus</i> sem polimorfismo	9.56	55.60	37	54	30	37.94	99.75	0.163
Bovino <i>Bos taurus</i> com polimorfismo K232A	9.52	55.54	37	53	30	37.94	99.96	0.175

pl: Ponto isoelétrico; PM: Peso molecular; Amin. -: Aminoácidos negativos; Amin. +: Aminoácidos positivos; T_{1/2}: Tempo de meia vida; GMH: Grande média de hidropaticidade.

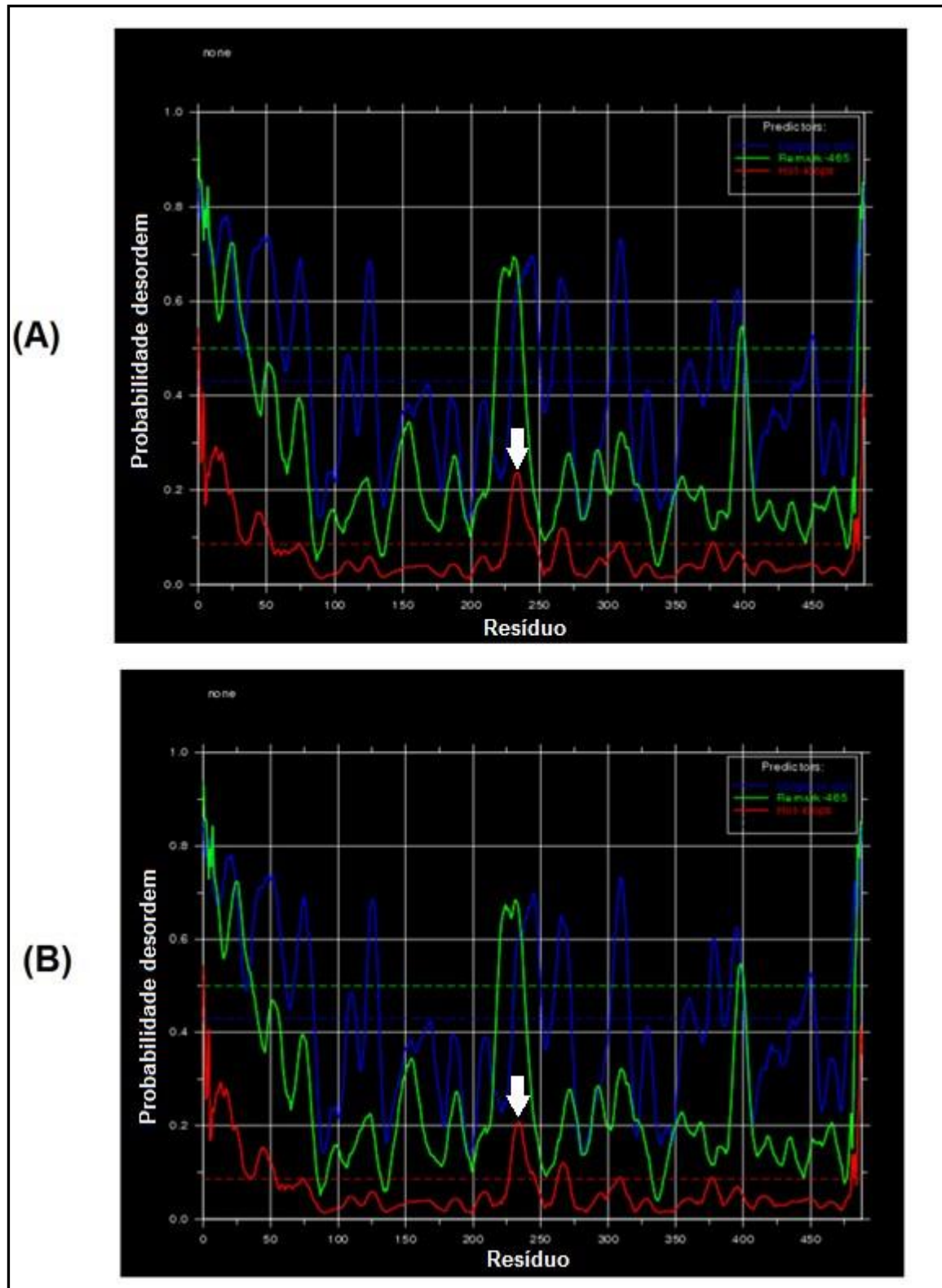
Os resultados encontrados para os parâmetros bioquímicos revelaram que a proteína não polimórfica e polimórfica apresentaram valores diferenciados para pl, peso molecular, índice alifático, GMH. A espécie não polimórfica por apresentar uma quantidade maior de aminoácidos carregados positivamente tende a possuir características hidrofílicas. Os resultados revelam que ambas as proteínas não polimórfica e polimórfica são caracterizadas como um peptídeo estável e possuem valores estimados de meia vida e quantidade de aminoácidos carregados negativamente idênticos.

6.5 Desordens bioquímicas da proteína DGAT-1 sem e com polimorfismo K232A em bovino *Bos taurus*

Os parâmetros referentes às análises de desordens da proteína DGAT-1 sem e com polimorfismo K232A da espécie *Bos taurus* utilizando o programa DisEMBL são apresentados na Figura 9A e 9B. Os três tipos caracterizados de desordens por meio dessas análises estão representados por: a) “Loops ou Coil”, nos quais os resíduos são classificados como pertencentes a algum tipo de região ordenada, não estando necessariamente desordenados, representado pela região azul, b) “Remark 465”, quando os resíduos podem estar dentro de uma região desordenada, porém, pode ser atribuído a um resíduo que não tem relação com a desordem estrutural de fato, representado em verde e c) “Hot-Loops”, apresentam alto grau de mobilidade determinado pelo fator temperatura, podendo ser consideradas regiões

estruturalmente desordenadas, fornecendo informações acerca da dinâmica da proteína, representado em vermelho. O eixo das abcissas representados pelos resíduos dos aminoácidos e o eixo das ordenadas pela probabilidade da desordem (Figura 9A e 9B). De acordo com os resultados obtidos, na proteína sem polimorfismo encontrou-se um pico ligeiramente maior representada pela cor vermelha (“Hot-Loops”) entre os resíduos de aminoácidos 200 e 250.

Figura 9 - Representação da desordem da proteína DGAT-1 sem polimorfismo (A) e com polimorfismo (B) da espécie *Bos taurus*. A seta na figura representa o sítio de alteração entre as duas sequências analisadas.



Fonte: Figura gerada pelo programa DisEMBL.
(<http://dis.embl.de/cgiDict.py>)

7 DISCUSSÃO

A enzima diacilglicerol aciltransferase-1, expressa pelo gene *dgat-1*, destaca-se como uma enzima crucial atuante no metabolismo dos triglicerídeos (SANDERS *et al.*, 2006; KUEHN *et al.*, 2007). Esta enzima atua na etapa final da biossíntese dos triglicerídeos e estudos revelam que apresenta ação na velocidade de acumulação dos lipídeos (TURCHETTO-ZOLET, 2011). Para estabelecer o grau de conservação da proteína DGAT-1 entre as espécies de búfalo, bovino *Bos taurus* e *Bos indicus*, cabra e ovelha foram empregadas ferramentas de Bioinformática. Comparando suas sequências de aminoácidos, verificou-se que há um alinhamento significativo (77,30% e 77,50% respectivamente) entre as sequências (Figura 3 e 4), provavelmente por serem espécies que evolutivamente são muito próximas, indicando um alto grau de conservação entre os aminoácidos presentes nesta proteína. Esses resultados são semelhantes aos encontrados por Venkatachalapathy *et al.* (2008) que, ao estudarem a similaridade entre sequências de DGAT1 de búfalo, gado, porco, macaco, humanos e ratos, observaram bom alinhamento nas sequências, revelando uma alta identidade entre as espécies. Nas análises comparativas entre as sequências de aminoácidos presentes na proteína DGAT-1 foi observada alta similaridade, revelando 77,50% a identidade entre as sequências, o valor de E encontrado nos parâmetros de alinhamento das sequências foi de 0.003, indicando um alinhamento significativo (Figura 4). Apesar de as sequências dos aminoácidos das espécies bovino *Bos indicus* e cabra não representarem sequências completas encontradas no banco de dados, as sequências parciais apresentaram boa identidade. Esta alta identidade entre as sequências analisadas pode estar associada à função da proteína DGAT-1 nos processos fisiológicos no metabolismo do triacilglicerol, uma vez desempenha papel fundamental no metabolismo dos lipídeos, como descrito por Cases *et al.* (1998). De acordo com Lesk (2008), se duas proteínas tiverem mais de 45% de resíduos idênticos em seu alinhamento, elas terão estruturas bastante similares e provavelmente funções comuns. Ainda este mesmo autor diz que se elas tiverem 25% de resíduos idênticos, provavelmente terão um padrão de enovelamento similar.

Foi a partir do alinhamento das seis sequências de aminoácidos das espécies de ruminantes que se pôde construir a árvore filogenética (Figura 6), devendo esta

expressar ancestralidade e/ou parentescos evolutivos entre as sequências (MIYAKI, 2001). Por meio da árvore filogenética construída pelo programa MEGA 6.0, verificou-se a formação de um único ramo principal, subdivido em quatro sub-ramos, sendo o primeiro formado pelas sequências de aminoácidos das espécies de ovelha, bovino *bos taurus* com e sem polimorfismo, um segundo formado pela sequência de cabra, um terceiro pela sequência de bovino *bos indicus* e um quarto pela sequência de búfalo (Figura 6). Pode-se concluir que as espécies são provenientes de um ancestral comum, provavelmente representando um grupo monofilético. Comparando-se as seis sequências da proteína DGAT-1 selecionadas nesse estudo, elas refletem proximidade na evolução e tais evidências denotam um elevado grau de conservação entre as sequências da proteína DGAT-1 nestes ruminantes por estarem agrupadas em um único ramo principal.

Alguns estudos têm sido realizados a partir de análises de alinhamentos múltiplos de sequências da proteína DGAT-1 a fim de inferir relações evolutivas entre diferentes espécies. Os autores Venkatachalapathy *et al.* (2008) construíram uma árvore filogenética partindo de sequências de aminoácidos da proteína DGAT-1 de 15 espécies diferentes e verificaram um único ramo agrupando mamíferos, mosca, nematoide e plantas. De acordo com este mesmo autor, os mamíferos bovinos e búfalos estavam intimamente agrupados em um ramo principal, enquanto suíno estava ligado a outro sub-ramo agrupado a este ramo principal.

Ainda, Turchetto-Zolet *et al.* (2011) aproveitando ampla disponibilidade de dados sobre o genoma de várias espécies, conduziu um estudo sobre a caracterização da origem e história evolutiva das famílias dos genes *dgat-1* e *dgat-2* em organismos eucarióticos. Em sua análise filogenética foi confirmada a existência de distintos clados para os genes *dgat-1* e *dgat-2* em eucariotos, indicando que eles devem ter evoluído de forma assimétrica e revelando que esses genes são muito divergentes nos aspectos moleculares e estruturais, embora apresentem um padrão de convergência funcional (TURCHETTO-ZOLET *et al.*, 2011).

Segmentos gênicos que codificam a enzima lipogênica DGAT-1 apresentam um potencial sobre características de produção de leite em bovinos, sendo considerado como um forte candidato para maximizar o teor de gordura no leite (GRISART *et al.*, 2002; WINTER *et al.*, 2002). Smith *et al.* (2000) encontraram que a lactação estava ausente em ratos que não expressavam cópias desse gene, caracterizando-se também como um forte gene candidato funcional. Já está bem

esclarecido que a detecção de uma substituição do aminoácido lisina por alanina (K232A) fornece um efeito importante sobre a composição e produção do leite. De acordo com Kuape *et al.* (2004), *dgat-1k* parece ser o alelo ancestral e a substituição K232A provavelmente ocorreu após a divergência da linhagem *Bos indicus* e *Bos taurus*. A conservação evolutiva do resíduo lisina entre os mamíferos indica a importância funcional de um resíduo carregado positivamente, hidrofílico, na posição em que a sua substituição por um resíduo alanina, hidrofóbico, pode de certa forma, ser previsto por alterar a funcionalidade da enzima. Estes resultados revelam que alterações na composição na sequência de aminoácido da proteína, proveniente por uma substituição de aminoácido na constituição da proteína, pode gerar peptídeos com características estruturais modificadas, podendo alterar as propriedades da proteína e afetar diretamente a sua função biológica (VAN HOLDE *et al.*, 2006).

Sob esse aspecto, a estrutura secundária de uma proteína corresponde ao primeiro nível de dobramento da proteína e é obtido pelo arranjo espacial de aminoácidos que formam padrões de estruturas regulares do tipo hélice e folha. Na estrutura secundária de uma proteína as regiões irregulares correspondem à curva e espiral, sendo a região espiral denominada de desorganizada. Na bioquímica e biologia estrutural, a estrutura secundária é a forma tridimensional de segmentos de uma proteína, podendo ser definida por ligações de hidrogênio formadas entre os átomos de hidrogênio e oxigênio contido nas ligações peptídicas da proteína (BOYLE, 2005). Os resultados apresentados nesse estudo demonstram a presença de uma espiral situada entre os aminoácidos 200 e 250 na proteína sem polimorfismo, enquanto que a presença do espiral não foi revelada na proteína com o polimorfismo K232A (Figura 8). O aparecimento da espiral na proteína sem o polimorfismo caracteriza-se como uma estrutura anormal, evidenciando uma falta de organização estrutural na molécula (BOYLE, 2005). As proteínas, para desempenharem bem suas funções biológicas, necessitam de uma estrutura conformacional bem definida. Mudanças na sua conformação podem ser muito significativas, afetando, conseqüentemente, o comprometimento e desempenho de sua ação. As enzimas representam um caso especial na função da proteína, elas ligam-se a moléculas-alvo catalisando reações químicas através de ligação com o seu substrato em um mecanismo bioquimicamente denominado como Chave-Fechadura. Em 1984 estudos sobre a especificidade das enzimas realizados por Fischer estabeleceram que as enzimas são estruturalmente complementares aos

seus substratos de modo a se encaixarem como uma chave em uma fechadura (VOET; VOET; PRATT, 2000). Esta ideia estabeleceu que cada enzima possui um sítio ativo específico a uma determinada estrutura que se associa ao seu substrato para desempenhar o seu papel funcional. Se o sítio ativo da enzima possuir variedade na interação com o seu substrato a enzima não será capaz de interagir com a mesma intensidade, ou em alguns casos, contrariamente poderá promover melhoria nessa interação. Baseado nesta evidência pode-se especular que a ausência da espiral posicionada entre os aminoácidos 200 e 250 na estrutura da proteína com polimorfismo pode ser o responsável por modificações espaciais na forma geométrica do sítio ativo da enzima, o qual poderia afetar fortemente a ligação do substrato no local do sítio ativo da enzima. Este domínio diferenciado na estrutura secundária da proteína DGAT-1 com polimorfismo em *Bos taurus* poderia ser o responsável por exercer uma influência positiva sobre a ação biológica durante a biossíntese de lipídios presentes no leite.

Dentre os parâmetros bioquímicos analisados a partir da estrutura linear da proteína DGAT-1 sem e com polimorfismo K232A em bovino *Bos taurus*, utilizando ferramentas do programa ProtParam, foi constatado que as proteínas DGAT-1 sem e com polimorfismo apresentam um peso molecular de 55.60 e 55.54 kDa, ponto isoelétrico de 9.56 e 9.52 (Tabela 1). Este valor encontrado indica que ambas as proteínas possuem propriedades alcalinas ($\text{pH} > 7$). A proteína não polimórfica apresenta características hidrofílicas, uma vez que a quantidade de aminoácidos carregados positivamente é ligeiramente superior ao encontrado na DGAT-1 polimórfica, bem como seu índice alifático e sua grande média de hidropaticidade apresentaram valores ligeiramente menores (Tabela 1). Ambas as proteínas revelam ser estáveis, visto que, os valores encontrados em seu índice de instabilidade são menores que 40 (GASTEIGER *et al.*, 2005). O tempo de meia vida biológica entre as duas espécies foi de 30 horas. Este fato reside na “expectativa de vida” das proteínas, sendo este tempo período em que a proteína exibe estabilidade e pode desenvolver suas funções. De acordo com Yu *et al.*, (2002), a proteína DGAT-1 na presença de um inibidor de protease tem uma meia vida relativamente longa de 18 horas, e esta meia vida não foi afetada pela expressão de DGAT-1, apoiando a noção de que mecanismos pós-traducionais de degradação dessa proteína exercem um papel significativo no controle dos níveis do estado estacionário da proteína DGAT-1. Assim, o valor de meia vida encontrado em ambas as proteínas DGAT-1

sem e com polimorfismo para a espécie em estudo parece apoiar uma forte capacidade de este peptídeo exercer sua função biológica de forma mais extensiva na biossíntese de lipídeos. Embora, a presença da espiral na espécie sem o polimorfismo possa ser um fator limitante para a completa atuação dessa enzima e ocasionar a diferença no conteúdo de triglicerídeos na composição e produção do leite em bovinos *Bos taurus*.

O estudo das desordens na proteína é importante para a compreensão da sua função. As regiões consideradas desordenadas por apresentar elevada flexibilidade na conformação estrutural da proteína afetam o mecanismo de funcionamento da proteína e, conseqüentemente, a sua funcionalidade biológica (VUCETIC *et al.*, 2003). Regiões desordenadas apresentam ausência de interação com as cadeias laterais. De acordo com Romero *et al.* (2001), geralmente as desordens estão associadas a regiões de baixa complexidade, tipicamente por apresentar níveis mais elevados de subconjuntos específicos da proteína, sendo essas regiões ricas em aminoácidos hidrofílicos, que são capazes de romper a estrutura, quando comparados com regiões que são estruturadas. De acordo com Tompa (2009), dois atributos são responsáveis pela principal causa da desordem em uma proteína: baixo valor de hidrofobicidade, impedindo a formação de um núcleo globular estável; e uma carga líquida (função das somas das cargas positivas e negativas), favorecendo os estados estendidos da proteína devido à repulsão eletrostática.

Analisando as predições obtidas por cada um dos três critérios de desordens empregados neste estudo, observou-se que para a proteína DGAT-1 sem o polimorfismo, obteve-se um percentual de desordem ligeiramente superior ao comparado com a proteína DGAT-1 com o polimorfismo. Essa desordem pode ter sido ocasionada pela presença do aminoácido lisina (hidrofílico) na posição compreendida entre os aminoácidos 200 e 250 (Figura 9A), mais especificamente na posição 232. A desordem ocasionada no estudo foi observada no critério “Hot-Loops”, onde os resíduos de aminoácidos apresentam um alto grau de mobilidade (DARNEL, 2003), proporcionando um comportamento menos conservador. Dentre os demais preditores “Loops/Coil” e “Remark465”, por serem classificados como não necessariamente desordenados, mesmo estando em uma região desordenada, não foi observada uma relação com a desordem estrutural (LINDING *et al.*, 2003). Sendo assim, a análise destes critérios no estudo da proteína DGAT-1 em bovino *Bos taurus* é desconsiderada.

A falta de ordenamento na proteína DGAT-1 sem polimorfismo em bovino *Bos taurus* compreendido na posição entre 200 e 250 dos aminoácidos pode estar conferindo desvantagens em relação à proteína DGAT-1 com polimorfismo na mesma espécie (Figura 9B). Isso porque a falta de uma estrutura ordenada nesta posição poderá comprometer o mecanismo biológico exercido por essa enzima, ocasionando características físico-químicas diferenciadas no leite desses animais.

8 CONCLUSÃO

De acordo com os resultados deste trabalho podem ser concluídos que:

- A proteína DGAT-1 apresentou elevado grau de conservação entre seus aminoácidos nas diferentes espécies de ruminantes.
- As diferentes espécies de ruminantes analisadas revelaram agrupamento filogenético em um único ramo principal, demonstrando estreita proximidade evolutiva.
- A caracterização bioquímica dessa proteína demonstrando a presença da espiral e desordens “Hot-Loops” exclusivamente na sequência da DGAT-1 não polimórfica em bovino *Bos taurus* pode estar afetando características qualitativas e quantitativas do leite.

REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, D. B.; FREITAS, A. C.; CARDOSO, D. F.; VENTURINI, G. C.; SOUZA, G. F. P.; TONHATI, H. Varreduras por polimorfismos no gene *dgat1* em búfalos da raça Murrah (*Bubalus bubalis*). **Revista Brasileira de Higiene e Sanidade Animal**, v. 8, p. 190 - 199, 2014.

ANGIOLILLO, A.; B.; DOMENECH, A.; SASTRE, Y.; BADAOU, B.; JORDANA, J. Identification of a single nucleotide polymorphism at intron 16 of the caprine acyl-coenzyme A: diacylglycerol acyltransferase 1 (DGAT1) gene. **Journal of Dairy Research**, v. 74, p. 47 - 51, 2007.

ARAÚJO FILHO, J. A.; CRISPIM, S. M. A. **Pastoreio combinado de bovinos, caprinos e ovinos em áreas de caatinga no Nordeste do Brasil**. In: Conferência Virtual Global sobre Produção Orgânica de Bovinos de Corte, 1., 2002, Via internet. Anais Eletrônicos... Concórdia: Universidade do Contestado, 2002. Disponível em <<http://sbmaonline.org.br/anais/v/trabalhos/pdfs/bm014.pdf>>. Acesso em: 15 mar. 2016.

ARBEX, W. **Modelos computacionais para identificação de informação genômica associada a resistência ao carrapato bovino**, Instituto Alberto Luiz Coimbra de Pós-graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2007.

BASTIANETTO, E. Criação de búfalos no Brasil: situação e perspectiva. **Revista Brasileira de Reprodução Animal**, n.6, p. 98 – 103, 2009.

BENNEWITZ, J.; REINSCH, N.; PAUL, S.; LOOFT, C.; KAUPE, B.; WEIMANN, C.; ERHARDT, G.; THALLER, G.; KÜHN, C.; SCHWERIN, M.; THOMSEN, H.; REINHARDT, F.; REENTS, R.; KALM, E. The *DGAT1* K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. **Am. Dairy Science Association**, v. 87, p. 431 – 442, 2004.

BOECKMANN, B.; BAIROCH, A.; APWEILER, R.; BLATTER, M. C.; ESTREICHER, A.; GASTEIGER, E.; MARTIN, M. J.; MICHOD, K.; O'DONOVAN, C.; PHAN, I.; PILBOUT, S.; SCHNEIDER, M. The swiss-prot protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Research**, v. 31, p. 365 – 370, 2003.

BORÉM, A.; CAIXETA, E.T. **Marcadores moleculares**. Viçosa: UFV, 2006, 532p.

BOYLE, J. Lehninger principles of biochemistry: Nelson D, and Cox M. **Biochemistry and Molecular Biology Education**, 4ª edição, 1119 p, 2005.

BROOKES, A. J. The Essence of SNPs. **Gene**, v. 234, p. 177 – 186, 1999.

CAMPOS, R. T.; CAMPOS, K. C. Diagnóstico técnico-econômico da ovinocaprinocultura no estado do ceará. **Revista Teoria e Evidência Econômica**, n. 40, p. 126 – 152, 2013.

CASES, S.; STONE, S. J.; ZHOU, P.; YEN, E.; TOW, B.; LARDIZABAL, K. D.; VOELKER, T.; FARESE JR., R. V. Cloning of *DGAT2*, a second mammalian

diacylglycerol acyltransferase, and related family members. **Journal of Biological Chemistry**, v. 276, p. 38870 - 38876, 2001.

CASES, S.; SMITH, S. J.; ZHENG, Y. W.; MYERS, H. M.; LEAR, S. R.; SANDE, E.; NOVAK, S.; COLLINS, C.; WELCH, C. B.; LUSIS, A. J.; ERICKSON, S. K.; FARESE JR, R. V. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. **Proceedings of the National Academy of Sciences**, v. 95, p. 13018 - 13023, 1998.

CHOU, P. Y.; FASMAN, G. D. Prediction of protein conformation. **Biochemistry**, v. 13, n. 2, p. 222 - 245, 1974.

CORPET, F. Multiple sequence alignment with hierarchical clustering. **Nucleic Acids Research**, v. 16, n. 22, p. 10881 - 10890, 1988.

COSTA, R. G.; ALMEIDA, E. C.; PIMENTA FILHO, E. C.; HOLANDA JUNIOR, E. V.; SANTOS, N. M. Caracterização do sistema de produção caprino e ovino na região semi-árida do estado da Paraíba, Brasil. **Archivos de zootecnia**, v. 57, n. 218, p. 195 – 205, 2008.

COUTINHO, M. J. F.; CARNEIRO, M. S. S.; EDVAN, R. L.; PINTO, A. P. A pecuária como atividade estabilizadora no semiárido brasileiro. **Veterinária e Zootecnia**, v. 20, n. 3, p. 9 – 17, 2013.

DARNELL, J. C.; JENSEN, K. B.; JIN, P.; BROWN, V.; WARREN, S. T.; DARNELL, R. B. Order, disorder, and flexibility: Prediction from protein sequence. **Structure**, v. 11, p. 1316 – 1317, 2003.

EISMANN, A. M. **Bioquímica: Introdução ao metabolismo**. 1. ed. Joinville: Clube de Autores, 2013, 189p.

EKERLJUNG, L.; LENNARTSSON, J.; GEDDA, L. The HER2-binding affibody molecule ($Z_{HER2:342}$)₂ increases radiosensitivity in SKBR-3 Cells. *Plos One*, v. 7, 2012.

ESPINDOLA, F. S.; CALABRIA, L. K.; REZENDE, A. A. A.; PEREIRA, B. B.; SANTANA, F. A.; AMARAL, I. M. R.; LOBATO, J.; FRANÇA, J. L.; MARIO, J. L.; FIGUEIREDO, L. B.; SANTOS-LOPES, L. P.; GOUVEIA, N. M.; NASCIMENTO, R.; TEIXEIRA, R. R.; REIS, T. A.; ARAUJO, T. G. Recursos de Bioinformática Aplicados às Ciências Ômicas como Genômica, Transcriptômica, Proteômica, Interatômica e Metabolômica. **Bioscience Journal**, v. 26, p. 463 – 477, 2010.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 2. ed. Brasília: EMBRAPA-CENAGEM, 1995, 220p.

FILHO, O. A. S.; ALENCASTRO, R. B. Modelagem de proteínas por homologia. **Química Nova**, v. 26, nº2, p. 253-259, Rio de Janeiro, 2003.

Food and Agriculture Organization of the United Nations – FAO. 2010. Disponível em: <www.fao.org>.

FURBASS, R.; WINTER, A.; FRIES, R. Alleles of the bovine DGAT1 variable number of tandem repeat associated with a milk fat QTL at chromosome 14 can stimulate gene expression. **Physiological Genomics**, v. 25, p. 116 – 120, 2006.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M. R., APPEL, R. D., BAIROCH, A. Protein Identification and Analysis Tools on the ExPASy Server. **Protocol**, The Proteomics Protocols Handbook, p. 571- 607, 2005.

GELDERMANN H. Investigation on inheritance of quantitative characters in animal by gene markers. **Theoretical and Applied Genetics**, v. 46, p. 319 – 330, 1975.

GLAZIER, A. M.; NADEAU, J. H.; AITMAN, T. J. Finding genes that underlie complex traits. **Science**, v. 298, p. 2345 - 2349, 2002.

GRISART, B.; COPPIETERS, W.; FARNIR, F.; KARIM, L.; FORD, C.; BERZI, P.; CAMBISANO, N.; MNI, M.; REID, S.; SIMON, P.; SPELMAN, R.; GEORGES, M.; SNELL, R. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. **Genome Research**, v.12, n. 2, p. 222 – 231, 2002.

GRISART, B.; FARNIR, F.; KARIM, L.; CAMBISANO, N.; KIM, J.J.; KVASZ, A.; MNI, M.; SIMON, P.; FRÈRE, J. M.; COPPIETERS, W.; GEORGES, M. Genetics and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. **Proceedings of the National Academy of Sciences**, v. 101, n. 8, p. 2398 - 2403, 2004.

GURUPRASAD, K.; REDDY, B.V.B.; PANDIT, M.W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. **Protein Engineering**, v. 4, n. 2, p. 155-161, 1990.

HALL, B. G. Building phylogenetic trees from molecular data with MEGA. **Molecular biology and evolution**, v. 30, n. 5, p. 1229 – 1235, 2013.

HORI-OSHIMA, S.; BARRERAS-SERRANO, A. Relationship between DGAT1 and Pit-1 genes polymorphism and Milk yield in holstein cattle. Proceedings, Western Section, **American Society of Animal Science**, 2003.

Instituto Brasileiro de Geografia e Estatística – IBGE. 2006. **Censo Agropecuário**. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/periodicos/51/agro_2006.pdf>. Acesso em: 13 abr. 2016.

Instituto Brasileiro de Geografia e Estatística – IBGE. 2013. **Produção da Pecuária Municipal**. Disponível em: <ftp://ftp.ibge.gov.br/Producao_Pecuaria/Producao_da_Pecuaria_Municipal/2013/ppm2013.pdf> Acesso em: 12 abr. 2016.

KAUPE, B.; ERHARDT, G.; BENNEWITZ, J.; SCHWERIN, M. F. R. Evidence for multiple alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on milk fat content in cattle. **Genetics**, v. 167, 1873 - 1881, 2004.

KNEZEVIC, Z. D.; SILER-MARINKOVIC, S. S.; MOJOVIC, L. V. Immobilized lipases as practical catalysts. **APTEFF**, v. 35, p. 1-280, 2004.

KUEHN, C.; EDEL, C.; WEIKARD, R. T. G. Dominance and parent of origin effects of coding and non-coding alleles at the acylCoA-diacylglycerol-acyltransferase (DGAT1)

gene on milk production traits in German Holstein cows. **Biomed Central Genetics**, v. 8, p. 1 – 9, 2007.

KUHN, C.; THALLER, G.; WINTER, A.; BININDA-EMONDS, O. R. P.; KAUPE, B.; ERHARDT, G.; BENNEWITZ, J.; SCHWERIN, M. F. R. Evidence for multiple alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on milk fat content in cattle. **Genetics**, v. 167, 1873 - 1881, 2004.

KUMAR, T. A. CFSSP: Chou and Fasman Secondary Structure Prediction Server. **Wide Spectrum**, v.1, n. 9, p. 15 - 19, 2013.

KYTE, J.; DOOLITTLE, R. A Simple Method for Displaying the Hydrophobic Character of a Protein. **Journal of Molecular Biology**, v. 157, p. 105 – 132, 1982.

LACORTE, G. A.; MACHADO, M. A.; MARTINEZ, M. L. DGAT1 K232A polymorphism in Brazilian cattle breeds. **Genetics and Molecular Research**, v. 5, n. 3, p. 475 – 482, 2006.

LESK, A. M. **Introdução a Bioinformática**. 2ª Edição. Porto Alegre: Artmed, 2008. 381p.

LIEFERS, S. C.; PAS, M. F. W.; VEERKAMP, R. F. Associations between leptin gene polymorphisms on production, live weight energy balance, feed intake and fertility in Holstein heifers. **Journal of Dairy Science**, v. 85, p. 1633 – 1638, 2002.

LINDING, R.; JENSEN, L. J.; DIELLA, F.; BORK, P.; GIBSON, T. J.; RUSSELL, R. B. Protein Disorder Prediction: Implications for Structural Proteomics. **Structure**, v.11, p. 1453 - 1459, 2003.

LIVESTOCK. In: Estados Unidos. **Department of agriculture**. PSD: production, supply and distribution online. Reports. Foreign Agricultural Service – USDA – 2013. Disponível em: <<http://www.fas.usda.gov/psdonline>>. Acesso em: 12 abr. 2016.

LODISH, H.; BERK, A.; KAISER, C. A.; KRIEGER, M.; BRETSCHER, A.; PLOEGH, H.; AMON, A.; SCOTT, M, P. **Biologia celular e molecular**, 7ª edição, editora artmed, 2014.

MAKI, H. Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. **Annual Reviews of Genetics**, v. 36, p. 279 – 303, 2002.

MARIN, A.L.A.; CERVIGNI, G. D. L.; MOREIRA, M. A.; BARROS, E. G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, v.30, p. 333 – 342, 2005.

MATSUDA, H.; YOSHIKAWA, T.; TABE, T.; KISHINAMI, R.; HASHIMOTO, A. On the Implementation of a Phylogenetic Tree Database. **Department of Mathematical Information Science**, p. 42 – 45, 1999.

MILLAR, A. A.; SMITH, M. A.; KUNST, L. All fatty acids are not equal: discrimination in plant membrane lipids. **Trends in Plant Science**, v. 5, n. 3, p. 95 – 101, 2000.

MIYAKI, CY.; RUSSO, C.; PEREIRA, S. L. **Reconstrução filogenética**: Introdução e o método da máxima parcimônia. *Biologia molecular e evolução*. 2. ed. Ribeirão Preto: Holos Editora, 2001, 257p.

NELSON, D. L.; COX, M. M. **Princípios da bioquímica de Lehninger**. 5. ed. Porto Alegre: Artmed, 2011, 1273p.

PALLEN, M. J. Microbial genomes. **Molecular Microbiology**, v. 32, n. 5, p. 907 -912, 1999.

PAPADOPOULOS, J. S.; AGARWALA, R. COBALT: constraint-based alignment tool for multiple protein sequences. **Bioinformatics**, v. 23, n.9, p. 1073 – 1079, 2007.

PENNISI, E. Genome data snake tree of life. **Science**, v. 284, p. 1305 - 1307, 1998.

PROSDOCIMI, F.; CERQUEIRA, G. C.; BINNECK, E.; SILVA, A. F.; REIS, A. N.; JUNQUEIRA, A. C.; SANTOS, A. C. F.; NBANI JUNIOR, A.; WUST, C. I.; CAMARGO FILHO, F.; KESSEDJIAN, J. L.; PETRETSKI, J. H.; CAMARGO, L. P.; FERREIRA, R de G, M.; LIMA, R. P.; PEREIRA, R. M.; JARDIM, S.; SAMPAIO, V. de S.; FLATSCHART, A. V. F. *Bioinformática: manual do usuário*. **Revista Biotecnologia Ciência e Desenvolvimento**, v. 5, n. 29, 2002.

PROSDOCIMI, F. **Introdução a Bioinformática**. *Bio Tecnologia Ciência & Desenvolvimento*, 2007.

ROBERTO, J.V.B. Fatores ambientais, nutricionais e de manejo e índices de conforto térmico na produção de ruminantes no semiárido. **Revista Verde de Agroecologia e Desenvolvimento sustentável**, v. 6, n. 2, p. 8 - 13, 2011.

ROMERO, P.; OBRADOVIC, Z.; GARNER, E. C.; BROWN, C. J.; DUNKER, A. K. Sequence complexity of disordered protein. **Proteins**, v.42, p.38 – 48, 2001.

RUSSO, C.A.M. **Como escolher genes para problemas filogenéticos específicos**. *Biologia molecular e evolução*. 2. ed. Ribeirão Preto: Holos Editora, 2001, 257p.

SAMPAIO, B.; SAMPAIO, Y.; LIMA, R. C.; AIRES, A.; SAMPAIO, G. A economia da caprinocultura em Pernambuco: problemas e perspectivas. **Revista de Economia**, v. 35, n. 2, p. 137 – 159, 2009.

SANDERS, K.; BENNEWITZ, J.; REINSCH, N.; THALLER, G.; PRINZENBERG, E. M.; KUHN, C.; KALM, E. Characterization of the DGAT1 mutations and the CSN1S1 promoter in the German Angeln dairy cattle population. **Journal Dairy Science**, v. 89, n. 8, p. 3164 – 3174, 2006.

SHOCKEY, J. M.; GIDDA, S. K.; CHAPITAL, D. C.; KUAN, J. C.; DHANOA, P. K.; BLAND, J. M.; ROTHSTEIN, S. J.; MULLEN, R. T.; DYER, J. M. Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are

localized to different subdomains of the endoplasmic reticulum. **Plant Cell**, v. 18, n. 9, p. 2294 – 2313, 2006.

SILVA, S. C. C.; VESCO, A. P.; LUIZZETI, F.; GASPARINO, E.; BOGATOLI, A.; MARQUES, L. A.; VOLTOLINI, D. S. M.; OLIVEIRA, D. P.; SOUZA, K. R. S. Frequência alélica e genotípica do gene *dgat1* em uma população de bovinos de leite. **Archivos de Zootecnia**, v. 60, n. 232, p.1343 - 1346, 2011.

SINHA, V. R.; TREHAN, A. Biodegradable microspheres for protein delivery. **Journal of controlled release**, v. 90, p. 261 - 280, 2003.

SMITH, J. A.; LEWIS, A. M.; WIENER, P. W. Genetic variation in the bovine myostation gene in UK beef cattle: allele frequencies and haplotype analysis in the South Devon. **Animal Genetics**, v. 31, n. 5, p. 306 – 309, 2000.

SMITH, S. J.; CASES, S.; JANSEN, D. R.; CHEN, H. C.; SANDE, E.; TOWL, B.; SANAN, D. A.; RABER, J.; EKCEL, R. H.; FARESE JR, R. V. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking *dgat*. **Nature Genetics**, v. 25, p. 87 - 90, 2000.

SOUSA, A. C.; SOARES, M. A. S.; RODRIGUES, M. T. Análises preliminares da região promotora do gene *DGAT1* em caprinos leiteiros. **Revista Brasileira de Biociências**, v. 6, p. 55 – 56, 2008.

SPELMAN, R. J.; FORD, C. A.; MCELHINNEY, P.; GREGORY, G. C.; SNELL, R. G. Characterization of the *DGAT1* gene in the New Zealand dairy population. **Journal Dairy Science**, v. 85, n. 12, p. 3514 – 3417, 2002.

STRACHAN, T.; READ, A.P. **Genética molecular humana**. 2ª ed. Porto Alegre: Art Med, 2002.

TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M.; KUMAR, S. Mega 5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. **Molecular Biology and Evolution**, v. 28, n. 10, p. 2731 – 2739, 2011.

TANTIA, M. S.; VIJH, R. K.; MISHRA, B.; KUMAR, S. T.; SODHI, M. *DGAT1* and *ABCG2* polymorphism in Indian cattle (*Bos indicus*) and buffalo (*Buballus bubalis*) breeds. **Veterinary Research**, v. 2, p. 1 - 5, 2006.

THALLER, G.; KÜHN, C.; WINTER, A. *DGAT1*, a new positional and functional candidate gene for intramuscular fat deposition in cattle. **Animal Genetics**, v. 34, n. 5, p. 354 – 357, 2003.

TOMPA, P. Bioinformatics approaches to the structure and function in intrinsically disordered proteins. In: **From protein structure to function with bioinformatics**, p.113-142, 2009.

TORRIERI, R. **Identificação e caracterização computacional de proteínas do tipo IUP no proteoma predito de *Schistosoma mansoni***. 2010. 136p. Dissertação (Mestrado em Ciências) – Centro de Pesquisa René Rachou, Belo Horizonte.

TURCHETTO-ZOLET, A. C.; MARASCHIN, F. S.; MORAIS, G. L.; CAGLIARI, A.; ANDRADE, M. B.; PINHEIRO, M. M.; MARGIS, R. Evolutionary view of acyl-CoA

diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. **BMC Evolutionary Biology**, v. 20, n. 11, p. 263, 2011.

VALENTE, J.; DURÃES, M. C.; MARTINEZ, M. L.; TEIXEIRA, N. M. **Melhoramento genético de bovinos de leite**. Juiz de Fora: Embrapa Gado de Leite, p. 87-92, 2001.

VAN HOLDE, K. E.; JOHNSON, W. C.; HO, P. S. **Principles of Physical Biochemistry**. 2. ed. New Jersey: Prentice Hall, 2006, 752p.

VENKATACHALAPATHY, R. T.; SHARMA, A.; SUKLA, S.; BHATTACHARYA, T. K. Cloning and characterization of DGAT1 gene of Riverine buffalo. **DNA Sequence**, v. 19, n. 3, p. 177 – 84, 2008.

VENKATACHALAPATHY, R.T.; SHARMA, A.; RADHA, K. Polymorphism at DGAT1 locus in Indian buffalo, zebuandBos indicus x Bos taurus cattle breeds. **Veterinary Science Research Journal**, v. 5, p. 13 – 17, 2014.

VIANA, G.V.R.; MOURA, H. A. S. Algoritmos para alinhamento de seqüências. **Revista Científica da Faculdade Lourenço Filho**, v.7, p. 67 – 83, 2010.

VOET, D.; VOET, J. G.; PRATT, C. W. **Fundamentos de bioquímica**. Porto Alegre: Editora Artmed, 2000, 1752p.

VUCETIC, S.; BROWN, C. J.; DUNKER, A. K.; OBRADOVIC, Z. Flavors of protein disorder. **Proteins**, v. 52 n. 4, p. 573-84, 2003.

WERNER, M. E. **Bioquímica: Fundamentos para Medicina y Ciencias de La Vida**. 1ª Edição. Editorial Reverte. 680p, 2008. WINTER, A.; ALZINGER, A.; FRIES, R. Assessment of the gene content of the chromosomal regions flanking bovine DGAT1. **Genomics**, v.83, p.172-180, 2004.

WINTER, A.; KRÄMER, W.; WERNER, F. A. O.; KOLLERS, S.; KATA, S.; DURSTEWITZ, G.; BUITKAMP, J.; WOMACK, J. E.; THALLER, G.; FRIES, R. Association of the lysine-232/alanine polymorphisms in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. **Proceedings of the National Academy of Sciences**, v. 99, n. 14, p. 9300 – 9305, 2002.

YANG, Z. **Computational molecular evolution**. New York: OUP Oxford, 2006, 374p.

YU, Y. H.; ZHANG, Y.; OELKERS, P.; STURLEY, S. L.; RADER, D. J.; GINSBERG, H. N. Posttranscriptional control of the expression and function of diacylglycerol acyltransferase-1 in mouse adipocytes. **Journal of Biological Chemistry**, v. 277, n. 52, p. 50876 – 50884, 2002.

YUAN, J.; ZHOU, J.; DENG, X. Molecular Cloning and Single Nucleotide Polymorphism Detection of Buffalo *DGAT1* Gene. **Biochemical Genetics**, v. 45, n. 7 – 8, p. 611 – 621, 2007.

ANEXOS

ANEXO 1 - PÁGINA PRINCIPAL DO PROGRAMA Swiss – Prot

<http://www.uniprot.org/uniprot/?query=reviewed%3Ayes>

www.uniprot.org/uniprot/?query=DGAT1+bos+taurus&sort=score

UniProtKB dgat1 bos taurus Advanced Search

BLAST Align Retrieve/ID mapping Help Contact

UniProtKB results

About UniProtKB Basket 5

Filter by BLAST Align Download Add to basket Columns 1 to 25 of 45 Show 25

Quote terms: "bos taurus" ×

Entry	Entry name	Protein names	Gene names	Organism	Length
<input type="checkbox"/> Q8MK44	DGAT1_BOVIN	Diacylglycerol O-acyltransferase 1	DGAT1	Bos taurus (Bovine)	489
<input type="checkbox"/> F1MIW3	F1MIW3_BOVIN	O-acyltransferase	DGAT1	Bos taurus (Bovine)	485
<input type="checkbox"/> Q7YRA3	EXOS4_BOVIN	Exosome complex component RRP41	EXOSC4 RRP41	Bos taurus (Bovine)	245
<input type="checkbox"/> Q7YR89	CYHR1_BOVIN	Cysteine and histidine-rich protein...	CYHR1	Bos taurus (Bovine)	311
<input type="checkbox"/> M9NJ94	M9NJ94_BOVIN	Diacylglycerol O-acyltransferase 1	DGAT1	Bos taurus (Bovine)	71
<input type="checkbox"/> M9NGP9	M9NGP9_BOVIN	Diacylglycerol O-acyltransferase 1	DGAT1	Bos taurus (Bovine)	71
<input type="checkbox"/> M9NJU3	M9NJU3_BOVIN	Diacylglycerol O-acyltransferase 1	DGAT1	Bos taurus (Bovine)	71
<input type="checkbox"/> Q7YR99	Q7YR99_BOVIN	Heat shock transcription factor 1	HSF1	Bos taurus (Bovine)	141
<input type="checkbox"/> Q1H629	Q1H629_BOVIN	Leptin	OBES2 ob	Bos taurus (Bovine)	48
<input type="checkbox"/> Q1H8S4	Q1H8S4_BOVIN	Diacylglycerol acyltransferase	DGAT1	Bos taurus (Bovine)	80

Filter by
 Reviewed (4) Swiss-Prot
 Unreviewed (41) TrEMBL
Popular organisms
 Bovine (43)
 Bos taurus x Bos indicus (2)
Search terms
 Filter "dgat1" as:
 gene name (11)
View by
 Taxonomy
 Keywords
 Gene Ontology
 Enzyme class
 Pathway

ANEXO 2 - SEQUÊNCIAS SELECIONADAS NO PROGRAMA Swiss – Prot.

>tr|Q0ZR39|Q0ZR39_BUBBU *Bubalus bubalis* GN=DGAT1 PE=3 SV=2
 MGDREGGAGGSRRRTGSRPSIQGGSGPAAAEVEVRDVGAGGDAPVRDTPDKDGDVDVVGSGH
 WDLRCHRLQDSLFSDDSGFSNYRGILNWCVVMLILSNARLFLENLIKYGILVDPIQVVSL
 FLKDPYSPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLLHGVLNATILCFPAAVA
 FLLESITPVGSLALMVYTIILFLKLFSDYRDVNLWCRERRAGAKAKAALAGKKANGGAAQR
 TVSYPDNLTYRDLYYFLFAPTLCYELNFRSPRIRKRFLRLRLEMLFLTQLQVGLIQQW
 MVPAIQNSMKPFKDMYSRIVERLLKLA VPNHLIWLIFFYWLFHSCNAVAELMQFGDRE
 FYRDWWNSE SITYFWQWNIPVHKWCIRHFYKPLRRGSSKWAARTAVFLASAFFHEYLV
 SIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQPVAVLMYVHDYVVL
 NREAPAAGT

>tr|A6YP72|A6YP72_BOSIN *Bos indicus* GN=DGAT PE=4 SV=1
 CHRLQDSLFSDDSGFSNYRGI LNWCVVMLILSNARLFLENLIKYGILVDPIQVVSLFLKD
 PYSWPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLLHGVLNATILCFPAAVAFLE
 SITPVGSLALMVYTIILFLKLFSDYRDVNLWCRERRAGAKAKAALAGKKANGGAAQRTVSY
 PDNLTYRDLYYFLFAPTLCYELNFRSPRIRKRFLRLRLEMLFLTQLQVGLIQQWMPVA
 IQNSMKPFKDMYSRIVERLLKLA VPNHLIWLIFFYWLFHSCNAVAELMQFGDREFRYRD
 WNSE SITYFWQWNIPVHKWCIRHFYKPLRRGSSKWAARTAVFLASAFFHEYLVSIPL
 RMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQPVAVLMYVHDYVVLNREA
 PAAGT

>tr|A8VJM4|A8VJM4_SHEEP *Ovis aries* GN=DGAT1 PE=3 SV=1
 MGDREGGAGGSRRRTGSRPSIQGGSRPAAAEVEVRDVGAGGDAPVRDTPDKDGDVDVVGSGH
 WDLRCHRLQDSLFSDDSGFSNYRGILNWCVVMLILSNARLFLENLIKYGILVDPIQVVSL
 FLKDPYSPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLLHGVLNATILCFPAAVA
 FLLESITPVGSLALMVYTIILFLKLFSDYRDVNLWCRERRAGAKAKAALAGKKANGGAAQR
 TVSYPDNLTYRDLYYFLFAPTLCYELNFRSPRIRKRFLRLRLEMLFLTQLQVGLIQQW
 MVPAIQNSMKPFKDMYSRIVERLLKLA VPNHLIWLIFFYWLFHSCNAVAELMQFGDRE
 FYRDWWNSE SITYFWQWNIPVHKWCLRHFYKPLRRGSSKWAARTGVFLASAFFHEYLV
 SIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQPVAVLMYVHDYVVL
 NREAPTAGT

>tr|A2T092|A2T092_CAPHI *Capra hircus* GN=DGAT1 PE=2 SV=1
 YGILVDPIQVVSLFLKDPYSPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLLHGVL
 NLATILCFPAAVAFLE SITPVGSLALMVYTIILFLKLFSDYRDVNLWCRERRAGAKAKAA
 LAGKKANGGAAQRTVSYPDNLTYRDLYYFLFAPTLCYELNFRSPRIRKRFLRLRLEML
 FLTQLQVGLIQQWMPVAIQNSMKPFKDMYSRIVERLLKLA VPNHLIWLIFFYWLFHSCN
 NAVAELMQFGDREFRYRDWWNSE SITYFWQWNIPVHKWCLRHFYKPLRRGSSKWAARTG
 VFLASAFFHEYLVSIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQP
 VAVLMYVHDYVVLNREAPTAGT

>sp|Q8MK44|1-489, non-polymorphic sequence
 MGDREGGAGGSRRRTGSRPSIQGGSGPAAAEVEVRDVGAGGDAPVRDTPDKDGDVDVVGSGH
 WDLRCHRLQDSLFSDDSGFSNYRGILNWCVVMLILSNARLFLENLIKYGILVDPIQVVSL
 FLKDPYSPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLLHGVLNATILCFPAAVA
 FLLESITPVGSLALMVYTIILFLKLFSDYRDVNLWCRERRAGAKAKAALAGKKANGGAAQR
 TVSYPDNLTYRDLYYFLFAPTLCYELNFRSPRIRKRFLRLRLEMLFLTQLQVGLIQQW
 MVPAIQNSMKPFKDMYSRIVERLLKLA VPNHLIWLIFFYWLFHSCNAVAELMQFGDRE
 FYRDWWNSE SITYFWQWNIPVHKWCIRHFYKPLRRGSSKWAARTAVFLASAFFHEYLV
 SIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQPVAVLMYVHDYVVL
 NREAPAAGT

>sp|Q8MK44|1-489, polymorphic sequence
 MGDREGGAGGSRRRTGSRPSIQGGSGPAAAEVEVRDVGAGGDAPVRDTPDKDGDVDVVGSGH
 WDLRCHRLQDSLFSDDSGFSNYRGILNWCVVMLILSNARLFLENLIKYGILVDPIQVVSL
 FLKDPYSPALCLVIVANIFAVAAAFQVEKRLAVGALTEQAGLLLHGVLNATILCFPAAVA
 FLLESITPVGSLALMVYTIILFLKLFSDYRDVNLWCRERRAGAKAKAALAGKAANGGAAQR
 TVSYPDNLTYRDLYYFLFAPTLCYELNFRSPRIRKRFLRLRLEMLFLTQLQVGLIQQW
 MVPAIQNSMKPFKDMYSRIVERLLKLA VPNHLIWLIFFYWLFHSCNAVAELMQFGDRE
 FYRDWWNSE SITYFWQWNIPVHKWCIRHFYKPLRRGSSKWAARTAVFLASAFFHEYLV
 SIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQPVAVLMYVHDYVVL
 NREAPAAGT

ANEXO 3 – SEGMENTOS COMUNS DAS SEQUÊNCIAS SELECIONADAS

>tr|Q0ZR39|Q0ZR39_BUBBU

YGILVDPIQVVSLFLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGV
 NLATILCFPAAVAFLLSITPVGSVLALMVYITILFLKLFYSYRDVNLWCRERRAGAKAKAA
 LAGKKANGGAAQRTVSYDPNLTyrDLYYFLFAPTLCYELNFPRSPRIRKRFLLRRLLEML
 FLTQLQVGLIQQWMPAIONSMKPFKMDYSRIVERLLKLAVPNHLIWLIFFYWLFHSC
 NAVAEMLQFGDREFYRDWWNSEsITYFQWNWNI PVHKWCIRHFYKPLRRGSSKWAARTA
 VFLASAFFHEYLVSIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQP
 VAVLMYVHDYVVLNREAPAAGT

>tr|A6YP72|A6YP72_BOSIN

YGILVDPIQVVSLFLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGV
 NLATILCFPAAVAFLLSITPVGSVLALMVYITILFLKLFYSYRDVNLWCRERRAGAKAKAA
 LAGKKANGGAAQRTVSYDPNLTyrDLYYFLFAPTLCYELNFPRSPRIRKRFLLRRLLEML
 FLTQLQVGLIQQWMPAIONSMKPFKMDYSRIVERLLKLAVPNHLIWLIFFYWLFHSC
 NAVAEMLQFGDREFYRDWWNSEsITYFQWNWNI PVHKWCIRHFYKPLRRGSSKWAARTA
 VFLASAFFHEYLVSIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQP
 VAVLMYVHDYVVLNREAPAAGT

>tr|A8VJM4|A8VJM4_SHEEP

YGILVDPIQVVSLFLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGV
 NLATILCFPAAVAFLLSITPVGSVLALMVYITILFLKLFYSYRDVNLWCRERRAGAKAKAA
 LAGKKANGGAAQRTVSYDPNLTyrDLYYFLFAPTLCYELNFPRSPRIRKRFLLRRLLEML
 FLTQLQVGLIQQWMPAIONSMKPFKMDYSRIVERLLKLAVPNHLIWLIFFYWLFHSC
 NAVAEMLQFGDREFYRDWWNSEsITYFQWNWNI PVHKWCLRHFYKPLRRGSSKWAARTG
 VFLASAFFHEYLVSIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQP
 VAVLMYVHDYVVLNREAPTAGT

>tr|A2T092|A2T092_CAPHI

YGILVDPIQVVSLFLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGV
 NLATILCFPAAVAFLLSITPVGSVLALMVYITILFLKLFYSYRDVNLWCRERRAGAKAKAA
 LAGKKANGGAAQRTVSYDPNLTyrDLYYFLFAPTLCYELNFPRSPRIRKRFLLRRLLEML
 FLTQLQVGLIQQWMPAIONSMKPFKMDYSRIVERLLKLAVPNHLIWLIFFYWLFHSC
 NAVAEMLQFGDREFYRDWWNSEsITYFQWNWNI PVHKWCLRHFYKPLRRGSSKWAARTG
 VFLASAFFHEYLVSIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQP
 VAVLMYVHDYVVLNREAPTAGT

>sp|Q8MK44|1-489 non-polymorphic sequence

YGILVDPIQVVSLFLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGV
 NLATILCFPAAVAFLLSITPVGSVLALMVYITILFLKLFYSYRDVNLWCRERRAGAKAKAA
 LAGKKANGGAAQRTVSYDPNLTyrDLYYFLFAPTLCYELNFPRSPRIRKRFLLRRLLEML
 FLTQLQVGLIQQWMPAIONSMKPFKMDYSRIVERLLKLAVPNHLIWLIFFYWLFHSC
 NAVAEMLQFGDREFYRDWWNSEsITYFQWNWNI PVHKWCIRHFYKPLRRGSSKWAARTA
 VFLASAFFHEYLVSIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQP
 VAVLMYVHDYVVLNREAPAAGT

>sp|Q8MK44|1-489 polymorphic sequence

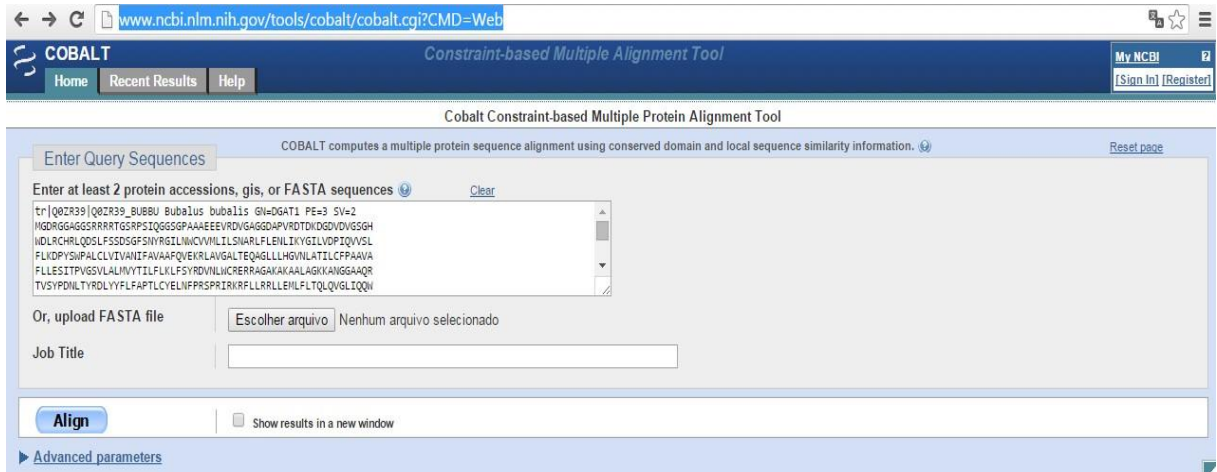
YGILVDPIQVVSLFLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGV
 NLATILCFPAAVAFLLSITPVGSVLALMVYITILFLKLFYSYRDVNLWCRERRAGAKAKAA
 LAGKANGGAAQRTVSYDPNLTyrDLYYFLFAPTLCYELNFPRSPRIRKRFLLRRLLEML
 FLTQLQVGLIQQWMPAIONSMKPFKMDYSRIVERLLKLAVPNHLIWLIFFYWLFHSC
 NAVAEMLQFGDREFYRDWWNSEsITYFQWNWNI PVHKWCIRHFYKPLRRGSSKWAARTA
 VFLASAFFHEYLVSIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVWLSLIIGQP
 VAVLMYVHDYVVLNREAPAAGT

ANEXO 4 - PÁGINA PRINCIPAL DO PROGRAMA MultAlin

<http://multalin.toulouse.inra.fr/multalin/>

ANEXO 5 - PÁGINA PRINCIPAL DO PROGRAMA COBALT

<http://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web>



The screenshot displays the COBALT web interface. At the top, the browser address bar shows the URL www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web. The page header includes the COBALT logo, navigation links for Home, Recent Results, and Help, and user options for My NCBI, Sign In, and Register.

The main content area is titled "Cobalt Constraint-based Multiple Protein Alignment Tool" and includes a sub-header: "COBALT computes a multiple protein sequence alignment using conserved domain and local sequence similarity information." A "Reset page" link is located in the top right of this section.

The "Enter Query Sequences" section contains a text input field with the following FASTA-style sequences:


```
tr|Q02739|Q02739_BUBBU Bubalus bubalis GH-DGATI PE-3 SV-2
MGDRGGAGGSRRRRTGSRPSIQGSGPAAHEEVVDVAGGDAVVDTDKGGVIVGSGH
NDLRCHLQDQLFSSDSGFSINVRGILMNCVWILISNARLFLEHLIYVGLVDPVQVYSL
FLKDPYSNPALCLVIVANIFVAARFQVEKRLAVGALTEQGLLLHGVNLTILCPAAVA
FLLESITPVGSLALMIVTILFLKLFYRDVNLICRERRAGAKAKAALAGKANGGAQR
TVSYPDNLTYRDLYYFLFAPTLCELEINFRSPRIKRFLLRLEHLFLTQVGLTQQN
```

Below the text area, there are options to "Or, upload FASTA file" (with a file selection button labeled "Escolher arquivo" and the text "Nenhum arquivo selecionado") and a "Job Title" input field.

At the bottom of the form, there is an "Align" button and a checkbox labeled "Show results in a new window". A link for "Advanced parameters" is located at the very bottom left.

ANEXO 6 - PÁGINA PRINCIPAL DO PROGRAMA CFSSP CHOU E FASMAN

<http://cho-fas.sourceforge.net/>



CFSSP: Chou & Fasman Secondary Structure Prediction Server

[Home](#) [Blog](#) [Tools](#) [Academic](#) [Contact](#) [Mail](#)

This server predicts secondary structure of protein from the amino acid sequence. In this server, Chou & Fasman algorithm has been implemented.

— Enter the protein sequence (in fasta format) —

```
>sp|Q8MK44|1-489, non-polymorphic sequence
MGDRGGAGGSRRTGSRPSIQGGSGPAAAAEEVRDVGAGGDAPVRDTRDKDGDVDVGS
GHWDLRCHRLQDSL FSSDSGFSNYRGILNWCVVMLILSNARLFLENLIKYGILVDP
IQVVSLFLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGVN
LATILCFPAAVAFLLESITPVGSVLALMVYITILFLKLF SYRDVNLWCRERRAGAKA
KAAALAGKKANGGAAQRTVSYPDNLT YRDLYYFLFAP
TLCYELNFRSPRIRKRFLRRRLEMLFLTQLQVGLIQQW
MVPAIQNSMKPFKDMYSRIVERLLKLA VPNHLIWLIF
FYWLFHSCLNVAELMQFGDRE
```


ANEXO 7 - PÁGINA PRINCIPAL DO PROGRAMA ProtParam

<http://web.expasy.org/protparam/>

← → ↻ web.expasy.org/protparam/ 🔍 ☆ ☰

ExPASy Bioinformatics Resource Portal ProtParam Home | Contact

ProtParam tool

ProtParam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Disclaimer).

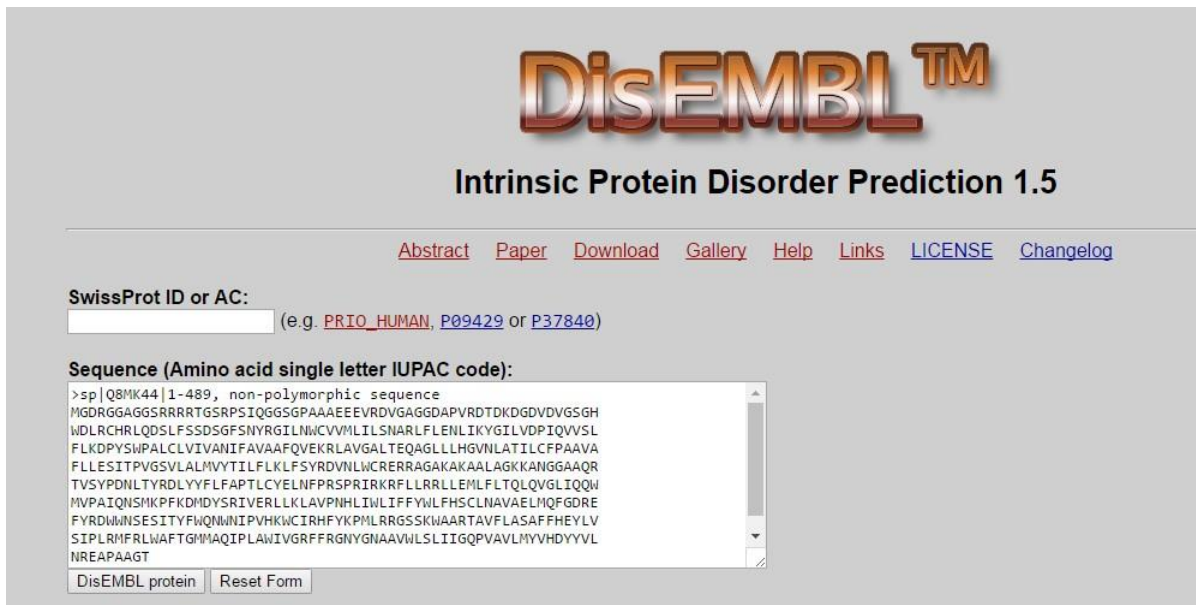
Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1_DROME**):

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

```

MGDRGGAGGSRRRRTGSRPSIQGGSGPAAAEVEEVRDVGAGGDAPVRDTDKDGDVDVG
SGHIDLRCRRLQDSLFSQSGFSNYRGLNWCVHILSNARLFLENLIKYGILLVDP
IQVSLFLKDPYSHPALCLVIVANIFAVAFQVEKRLAVGALTEQAGLLHGVNLAT
ILCFPAVAWFLLESITPVGSLALWVYTLFLKLFYRDVNLKCRERRAGAKAKAAL
AGKANGGAAQRTVSYVDNLTYRDLVYFLFAPTLCEVLFNFRSPRIRKFLRLRLE
MLFLTQLQVGLIQQWVPAIQNSMKPFKMDYSRIVERLLKLVAVNHLIHLFFVWL
FHSLNAVAELIQFGDREFYRDWVNSSEITFYQNMNIPVHKWICIRHFYKPMRLRGS
SKWAARTAVFLASAFFHEVLYLSIPLRFRLWFTGMMAQIPLAMIVGRFFRGNVYNA
  
```

ANEXO 8 - PÁGINA PRINCIPAL DO PROGRAMA DisEMBL™<http://dis.embl.de>

The screenshot shows the main interface of the DisEMBL™ website. At the top, the logo "DisEMBL™" is displayed in a large, stylized font. Below it, the title "Intrinsic Protein Disorder Prediction 1.5" is centered. A navigation bar contains links for "Abstract", "Paper", "Download", "Gallery", "Help", "Links", "LICENSE", and "Changelog". The main form area includes a "SwissProt ID or AC:" label and a text input field with a placeholder "(e.g. PRIO_HUMAN, P09429 or P37840)". Below this is a "Sequence (Amino acid single letter IUPAC code):" label and a text area containing a protein sequence. At the bottom of the form, there are two buttons: "DisEMBL protein" and "Reset Form".

DisEMBL™

Intrinsic Protein Disorder Prediction 1.5

[Abstract](#) [Paper](#) [Download](#) [Gallery](#) [Help](#) [Links](#) [LICENSE](#) [Changelog](#)

SwissProt ID or AC:
 (e.g. [PRIO_HUMAN](#), [P09429](#) or [P37840](#))

Sequence (Amino acid single letter IUPAC code):
>sp|Q8MK44|1-489, non-polymorphic sequence
MGDRGGAGGSRRRRTGSRPSIQG65GPAAAEVVDVGAGGDAPVRDTRDKDGDVDVGS
NDLRCHRQDLSLFSSD5GFSNYRGILNMCVVMILSNARLFLENLIKYGILVDP
FLKDPYSWPALCLVIVANIFAVAAFQVEKRLAVGALTEQAGLLHGVNLTILCF
FLLSITPVGSVLALMVYIILFLKLFYRDVNLWCRERRAGAKAKALAGKANGGAAQR
TVSYPDNLTYRDLYYFLFAPTLCELNFPSPRIRKRFLLRRLLEMLFTQLQVGLIQ
MVPAINSMKPFKMDYSRIVERLLKLVAPHLIWLIFFYWL FHSCLNAVAELMQF
FYRDWVNSISITYFWQNMNIPVHKWICIRHFYKPHLRGSSKWAARTAVFLASAF
SIPLRMFRLWAFTGMMAQIPLAWIVGRFFRGNYGNAAVNLSLIIGQPVAVLHYV
NREAPAAGT

DisEMBL protein Reset Form

APÊNDICE

APÊNDICE 1 - ARTIGO PUBLICADO

Animal Biology Journal - Volume 4, Number 4 - ISSN: 1949-498X

Phylogenetical Clustering Among Domestic Ruminants from Diacylglycerol-Acyltransferase-1 Lipogenic Enzyme and Biochemical Characterization of Its Single Nucleotide Polymorphism K232A in Bovine *Bos taurus*

*Fernanda Fernandes Pinheiro da Costa*¹,
*Kárita Cláudia Freitas Lidani*², *Sávio Luiz Nunes*³,
*João José de Simoni Gouveia*⁴ and *Jane Eyre Gabriel*^{1,3,*}

¹Colegiado de Pós-Graduação em Ciências Veterinárias no Semiárido, Centro de Ciências Agrárias, Universidade Federal do Vale do São Francisco UNIVASF, Petrolina, PE, Brasil

²Departamento de Patologia Médica, Hospital de Clínicas, Universidade Federal do Paraná UFPR, Curitiba, PR, Brasil

³Colegiado de Ciências Biológicas, Centro de Ciências Agrárias, Universidade Federal do Vale do São Francisco UNIVASF, Petrolina, PE, Brasil

⁴Colegiado de Zootecnia, Centro de Ciências Agrárias, Universidade Federal do Vale do São Francisco UNIVASF, Petrolina, PE, Brasil

ABSTRACT

This study aimed to assess multiple sequences of the diacylglycerol-acyltransferase-1 (DGAT1) lipogenic enzyme from different species of suborder Ruminantia to infer the evolutionary clustering among domestic ruminants and to further predict biochemical aspects of the non-polymorphic and polymorphic harboring K232A DGAT1 proteins of bovine *Bos taurus* using computational biology tools. First, amino acid sequences of the DGAT1 protein from domestic water buffalo *Bubalus bubalis* (Q0ZR39), Zebu bovine *Bos indicus* (A6YP72), sheep *Ovis aries* (A8VJM4), goat *Capra hircus* (A2T092) and bovine *Bos taurus* (Q8MK44) were randomly searched in the non-redundant protein sequence database UniProtKB/Swiss-Prot. A dendrogram showing the evolutionary relationships among specimens was constructed from the sequences of interest using Mega 6.0 software with the neighbor-joining method. Next, the prediction of biochemical characteristics associated with primary and secondary structures of the non-polymorphic and polymorphic harboring K232A DGAT1 protein sequences of bovine *Bos taurus* was performed using ProtParam program. The evolutionary clustering among specimens inferred from amino acid sequences of the DGAT1 protein in ruminants demonstrated that animals selected were sub-branched with all specimens intimately grouped in a main unique cluster. Biochemical characterization of the non-polymorphic and polymorphic DGAT1 proteins exhibited high estimates of half-life inferred from *in vitro* mammalian reticulocytes with identical values of 30 hours. In conclusion, such findings provide insights into the evolutionary biology of domestic ruminants and prediction of primary and secondary structures of the non-polymorphic and polymorphic harboring K232A DGAT1 proteins by the comparing multiple sequence alignments of biological molecules using Bioinformatics' tools.

Keywords: bioinformatics, conservation degree, DGAT1 protein, evolutionary biology, ruminant

INTRODUCTION

The diacylglycerol-acyltransferase (DGAT) proteins act as a key metabolic enzyme that catalyze the conversion of diacylglycerol and fatty acyl CoA to triacylglycerol, exerting crucial biological roles in the final step of lipogenesis during neutral lipid biosynthesis (Cases et al., 2001, Shi and Cheng, 2009, Liu et al., 2012, Yen et al., 2015). These proteins are expressed in nearly all tissues, such as liver, adipose tissue, skeletal muscle, including the mammary gland (Liu et al., 2012). Considerable attention has been expended to clarify the functions of the DGAT proteins and it is well established that the triacylglycerol metabolism mediated by DGAT activities comprehends intestinal fat absorption, adipose tissue formation, lactation and lipoprotein assembly (Cases et al., 1998, Yen et al., 2008, Yen et al., 2015). In excess energy situations, glucose and insulin-mediated signaling result in increasing the transcription of genes, and the excess energy will consequently activate the fatty acid synthesis and the triglyceride formation to store energy as an adipose tissue (Meegalla et al., 2002). Regarding to DGAT1 protein, some biological effects of this lipogenic protein are associated with obesity and other metabolic diseases, since mice lacking DGAT1 exhibiting decreased triglyceride content in mammary tissue were resistant to diet-induced obesity and diabetes mellitus (Lewin et al., 2008). Recently, Lu et al., (2012) found that knockdown of the DGAT1 expression in liposome by RNA interference provoked significant reductions in the triacylglycerol contents in the bovine mammary epithelial cells, reinforcing the importance of this peptide in the triglyceride synthesis in bovine mammary tissue. Over the past decade, the gene encoding the DGAT1 microsomal enzyme has been identified as a strong candidate gene due to its appropriate chromosomal location and that knockout transgenic mice lacking both functional allele of DGAT1 were totally devoid of lactation (Smith et al., 2000).

According to Winter et al., (2002), a non-conservative lysine (K) to alanine (A) substitution at amino acid position 232 in the mature polypeptide chain (DGAT1 K232A) could be directly associated with at least 50% of the quantitative trait locus variation for some milk production traits. This non-conservative substitution of lysine by alanine as a consequence of an adenine/adenine to guanine/cytosine replacement at positions 10433 and 10434 of exon VIII was found in European bovine populations (Spelman et al., 2002, Thaller et al., 2003). In fact, the presence of the K allele assures higher fat yield and both fat and protein percentages, whilst the A-allele was responsible for higher milk yield (Kaupe et al., 2004, Miltiadou et al., 2010). Recently, Tăbăran et al., (2015) showed that in Romanian Holstein cattle the K allele is associated with increased fat percentage and higher levels of C16:0 and C18:0 fatty acids and the DGAT1 protein harboring the polymorphism K232A revealed a high genetic variance for fat percentage, unsaturated C18 and C16:0. Thus, several lines of evidence strongly support the fact that K232A is indeed the causal mutation or quantitative trait nucleotide due to its known role in fat metabolism and knockout effect (Cases et al., 1998; Smith et al., 2000, Mackay 2001, Grisart et al., 2004, Tăbăran et al., 2015). Despite a number of previous studies on the DGAT1 protein, which have emphasized its importance as potential biomolecule targets to increase triacylglycerol on the milk composition and yield in domestic ruminants, little is known about its evolutionary timeline in eukaryotes.

With the advent of the Bioinformatics, the biological fields have tremendously progressed in the ability to process and interpret large volumes of biological data using computational biology analyses, advancing molecular and structural characterization of several bioactive molecules (Schatz et al., 2010; Gabriel et al., 2013, Oliani et al., 2015). Thus, the purpose of this study was to assess comparatively the conservation degree of the multiple amino acid sequences of the DGAT1 protein inferring phylogenetical relationships among distinct animals of the suborder Ruminantia. Moreover, this study aimed to characterize biochemically non-polymorphic and polymorphic harboring K232A DGAT1 sequences of bovine *Bos taurus* by predicting protein primary and secondary structure aspects using computational biology tools.

MATERIAL AND METHODS

Multiple Alignments of Amino Acid Sequences of the Ruminant DGAT1 Protein

The amino acid sequences of the DGAT1 protein from domestic water buffalo *Bubalus bubalis* (Q0ZR39, 489 amino acids), Zebu bovine *Bos indicus* (A6YP72, 425 amino acids), sheep *Ovis aries* (A8VJM4, 489 amino acids), goat *Capra hircus* (A2T092, 382 amino acids) and bovine *Bos taurus* (Q8MK44, 489 amino acids) were randomly selected from the high-quality annotated and non-redundant protein sequence database UniProtKB/Swiss-Prot. This database allows the functional annotation of multiple protein sequences and accurate classification of whole proteins into a network structure reflecting their evolutionary relationships (Boutet et al., 2007). The DGAT1 protein sequences selected were submitted to multiple alignments using the constraint-based multiple protein alignment tool COBALT (Papadopoulos and Agarwala, 2007). Such computational algorithm provides the annotation of multiple protein sequences from conserved domain database, protein motif, and local information of the sequence similarity, including additional data to improve the quality score and significance of the matches, such as expected value (e-value). E-values from multiple alignment analyses represent the probability of the alignment occurring by chance and are calculated based on the quality of alignment (the score) and the size of the database (Baxevanis and Ouellette, 2005). Moreover, statistical evaluative parameters concerning identity and similarity values were automatically validated in these comparative analyses. Secondly, a dendrogram demonstrating the evolutionary relationships among specimens was constructed from sequences of interest using Mega 6.0 software (Tamura et al., 2013), as described in details by Gabriel et al., (2015). Divergence times for all branching points in the topology were calculated using the RelTime method (Tamura et al., 2012).

Biochemical Characterization of Non-Polymorphic and Polymorphic Harboring K232A DGAT1 Sequences

Two sequences of the DGAT1 protein of bovine *Bos taurus* were biochemically characterized under these experimental conditions: a) non-polymorphic DGAT1 protein containing the lysine (K) amino acid at position 232 (Q8MK44) and b) DGAT1 protein sequence harboring the causal polymorphism K232A with the substitution by alanine (A) amino acid at same position. Biochemical parameters of the amino acid residues of non-polymorphic and polymorphic harboring K232A DGAT1 sequences were generated using the ProtParam tools that allow the validation of various physical and chemical parameters from prediction of the primary structure of the bovine DGAT1 protein sequence (Gasteiger et al., 2005). These computed parameters include the theoretical isoelectric point (pI), estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Finally, the prediction of the secondary structure of the non-polymorphic and polymorphic harboring K232A DGAT1 proteins of bovine *Bos taurus* was carried out using the Chou & Fasman Protein Secondary Structure Prediction CFSSP algorithm (Chou and Fasman, 1974a, Chou and Fasman, 1974b).

RESULTS

Comparative alignments of multiple sequences of the DGAT1 protein using CLUSTAL W method revealed an identity value of 77.505% by Swiss-Prot analyses, with 279 residues detected as identical positions in a total content of 489 amino acids (Figure 1). Therefore, only 03 amino acid residues were detected as similar positions along the sequences analyzed (Figure 1). Under these molecular evolutionary analyses, all comparative assessments generated e-values of 0.003, suggesting that the data generated from these alignments were significant. Moreover, comparative alignments of the DGAT1 protein were exclusively performed from bovine, buffalo and ovine sequences, specimens that have complete segment of the protein of interest deposited on database set (489 amino acid residues), excluding the partial fragments of the DGAT1 protein of goat and bovine *Bos indicus*. A higher identity value of 99.182% was revealed under these analyses, identifying 3 and 485 amino acid residues in similar and identical positions, respectively (data not shown).

Phylogenetic groupings were established to examine the overall evolutionary relationship among specimens inferred from amino acid sequences of the DGAT1 protein in animals of the suborder Ruminantia, as shown in Figure 2. The NJ tree obtained from multiple alignments of amino acid sequences of the ruminant DGAT1 protein revealed the presence of a unique cluster containing buffalo (0.00), bovine *Bos taurus* (non-polymorphic and polymorphic K232A sequences) (0.00) and sheep (0.00) species grouped intimately in a same branch (Figure 2). DGAT1 sequences of bovine *Bos indicus* (1.3) and goat (1.5) were linked to main cluster in distinct sub-branches (Figure 2). Estimates concerning evolutionary divergence were inferred from amino acid sequences of the DGAT1 protein. Buffalo, goat and non-polymorphic bovine DGAT1 sequences were the most differentiated (2.901), whereas the sequences of the protein of interest were the least differentiated in buffalo and bovine *Bos taurus* harboring the polymorphism K232A (0.003).

Q0ZR39	Q0ZR39	BUBBU	61	WDLRCHRLQDLSLFSSDSGFSNYRGIILNMCVVHLLSNARLFLENLIKYGILVDPPIQVVS	120
A6YP72	A6YP72	BOSIN	1	----CHRLQDLSLFSSDSGFSNYRGIILNMCVVHLLSNARLFLENLIKYGILVDPPIQVVS	56
A8VJM4	A8VJM4	SHEEP	61	WDLRCHRLQDLSLFSSDSGFSNYRGIILNMCVVHLLSNARLFLENLIKYGILVDPPIQVVS	120
A2T092	A2T092	CAPHI	1	-----YGIIVDPPIQVVS	13
Q8MK44	DGAT1_BOVIN		61	WDLRCHRLQDLSLFSSDSGFSNYRGIILNMCVVHLLSNARLFLENLIKYGILVDPPIQVVS	120

Q0ZR39	Q0ZR39	BUBBU	121	FLKDFYSWPALCLIVVANI FAVAAAFQVEKRLAVGALTEQAGLLHGVNLTATILCFPAAVA	180
A6YP72	A6YP72	BOSIN	57	FLKDFYSWPALCLIVVANI FAVAAAFQVEKRLAVGALTEQAGLLHGVNLTATILCFPAAVA	116
A8VJM4	A8VJM4	SHEEP	121	FLKDFYSWPALCLIVVANI FAVAAAFQVEKRLAVGALTEQAGLLHGVNLTATILCFPAAVA	180
A2T092	A2T092	CAPHI	14	FLKDFYSWPALCLIVVANI FAVAAAFQVEKRLAVGALTEQAGLLHGVNLTATILCFPAAVA	73
Q8MK44	DGAT1_BOVIN		121	FLKDFYSWPALCLIVVANI FAVAAAFQVEKRLAVGALTEQAGLLHGVNLTATILCFPAAVA	180

Q0ZR39	Q0ZR39	BUBBU	181	FLLESITPVGSVLAIMVYITILFLKLFYSYRDVNLNCRERRAGAKAKAALAGKKGAAQR	240
A6YP72	A6YP72	BOSIN	117	FLLESITPVGSVLAIMVYITILFLKLFYSYRDVNLNCRERRAGAKAKAALAGKKGAAQR	176
A8VJM4	A8VJM4	SHEEP	181	FLLESITPVGSVLAIMVYITILFLKLFYSYRDVNLNCRERRAGAKAKAALAGKKGAAQR	240
A2T092	A2T092	CAPHI	74	FLLESITPVGSVLAIMVYITILFLKLFYSYRDVNLNCRERRAGAKAKAALAGKKGAAQR	133
Q8MK44	DGAT1_BOVIN		181	FLLESITPVGSVLAIMVYITILFLKLFYSYRDVNLNCRERRAGAKAKAALAGKKGAAQR	240

Q0ZR39	Q0ZR39	BUBBU	241	TVSYFDNLTIRDLVYVFLFAPLTCYELNFFRSPRIKRFLLRRLLEMLFLIQLGVGLIQQW	300
A6YP72	A6YP72	BOSIN	177	TVSYFDNLTIRDLVYVFLFAPLTCYELNFFRSPRIKRFLLRRLLEMLFLIQLGVGLIQQW	236
A8VJM4	A8VJM4	SHEEP	241	TVSYFDNLTIRDLVYVFLFAPLTCYELNFFRSPRIKRFLLRRLLEMLFLIQLGVGLIQQW	300
A2T092	A2T092	CAPHI	134	TVSYFDNLTIRDLVYVFLFAPLTCYELNFFRSPRIKRFLLRRLLEMLFLIQLGVGLIQQW	193
Q8MK44	DGAT1_BOVIN		241	TVSYFDNLTIRDLVYVFLFAPLTCYELNFFRSPRIKRFLLRRLLEMLFLIQLGVGLIQQW	300

Q0ZR39	Q0ZR39	BUBBU	301	MVFAIQSMKPFKMDYSRIVERLLKLVFNHILNLIFFYWLPHSCLNVAEALMQFGDRE	360
A6YP72	A6YP72	BOSIN	237	MVFAIQSMKPFKMDYSRIVERLLKLVFNHILNLIFFYWLPHSCLNVAEALMQFGDRE	296
A8VJM4	A8VJM4	SHEEP	301	MVFAIQSMKPFKMDYSRIVERLLKLVFNHILNLIFFYWLPHSCLNVAEALMQFGDRE	360
A2T092	A2T092	CAPHI	194	MVFAIQSMKPFKMDYSRIVERLLKLVFNHILNLIFFYWLPHSCLNVAEALMQFGDRE	253
Q8MK44	DGAT1_BOVIN		301	MVFAIQSMKPFKMDYSRIVERLLKLVFNHILNLIFFYWLPHSCLNVAEALMQFGDRE	360

Q0ZR39	Q0ZR39	BUBBU	361	FYRDMWNSISITYFWQNNIIVVHKWCIIRHFYKPHLRGSSKWAARTAVFLASAFFHEYL	420
A6YP72	A6YP72	BOSIN	297	FYRDMWNSISITYFWQNNIIVVHKWCIIRHFYKPHLRGSSKWAARTAVFLASAFFHEYL	356
A8VJM4	A8VJM4	SHEEP	361	FYRDMWNSISITYFWQNNIIVVHKWCIIRHFYKPHLRGSSKWAARTAVFLASAFFHEYL	420
A2T092	A2T092	CAPHI	254	FYRDMWNSISITYFWQNNIIVVHKWCIIRHFYKPHLRGSSKWAARTAVFLASAFFHEYL	313
Q8MK44	DGAT1_BOVIN		361	FYRDMWNSISITYFWQNNIIVVHKWCIIRHFYKPHLRGSSKWAARTAVFLASAFFHEYL	420

Q0ZR39	Q0ZR39	BUBBU	421	SIFLRFMLNWAFTGSAQIPLAMIVGRFFRGNHGAAVWLSLIIGQPVAVLMYVHDYVVL	480
A6YP72	A6YP72	BOSIN	357	SIFLRFMLNWAFTGSAQIPLAMIVGRFFRGNHGAAVWLSLIIGQPVAVLMYVHDYVVL	416
A8VJM4	A8VJM4	SHEEP	421	SIFLRFMLNWAFTGSAQIPLAMIVGRFFRGNHGAAVWLSLIIGQPVAVLMYVHDYVVL	480
A2T092	A2T092	CAPHI	314	SIFLRFMLNWAFTGSAQIPLAMIVGRFFRGNHGAAVWLSLIIGQPVAVLMYVHDYVVL	373
Q8MK44	DGAT1_BOVIN		421	SIFLRFMLNWAFTGSAQIPLAMIVGRFFRGNHGAAVWLSLIIGQPVAVLMYVHDYVVL	480

Figure 1. Comparative alignments of multiple sequences of the DGAT1 protein in domestic water buffalo *Bubalus bubalis* (Q0ZR39), Zebu bovine *Bos indicus* (A6YP72), sheep *Ovis aries* (A8VJM4), goat *Capra hircus* (A2T092) and bovine *Bos taurus* (Q8MK44). Different arrows indicate the presence of conservative changes along sequences of the DGAT1 enzyme.

Various physical and chemical parameters were predicted from the primary structure of the non-polymorphic and polymorphic harboring K232A DGAT1 protein sequences of bovine *Bos taurus* by ProtParam analyses (data not shown). The biochemical characterization of the non-polymorphic DGAT1 protein sequence demonstrated a theoretical pI of 9.56 and molecular weight of approximately 56 kDa. The total number of negatively (Asparagine+Glutamic acid) and positively (Arginine+Lysine) charged residues was 37 and 54, respectively. Moreover, this protein was classified as a stable peptide, since its instability index was computed to be 37.94 with grand average of hydropathicity (GRAVY) of 0.163. Regarding to the polymorphic harboring K232A DGAT1 protein of bovine *Bos taurus*, descriptive biochemical parameters demonstrated theoretical pI and molecular weight values of 9.52 and 56 kDa, respectively. The total number of negatively (Asparagine+Glutamic acid) and positively (Arginine+Lysine) charged residues was 37 and 53, respectively. Data of the instability index were computed in 37.94, with grand average of hydropathicity (GRAVY) of 0.175, supporting that the polymorphic harboring K232A DGAT1 protein is also classified as a stable peptide. Interestingly, estimates of half-life inferred from *in vitro* mammalian reticulocytes demonstrated identical values of 30 hours for both the non-polymorphic and polymorphic harboring K232A DGAT1 proteins of bovine *Bos taurus*.

Analyses from secondary structure prediction of amino acid sequences of non-polymorphic and polymorphic harboring K232A DGAT1 proteins of bovine *Bos taurus* were estimated as presented in Figure 3. These results revealed the presence of a coiled coil domain situated between 200 and 250 amino

acid residues in the non-polymorphic DGAT1 protein, whereas this coiled coil domain was not detected at the same position in the polymorphic harboring K232A DGAT1 protein (Figure 3).

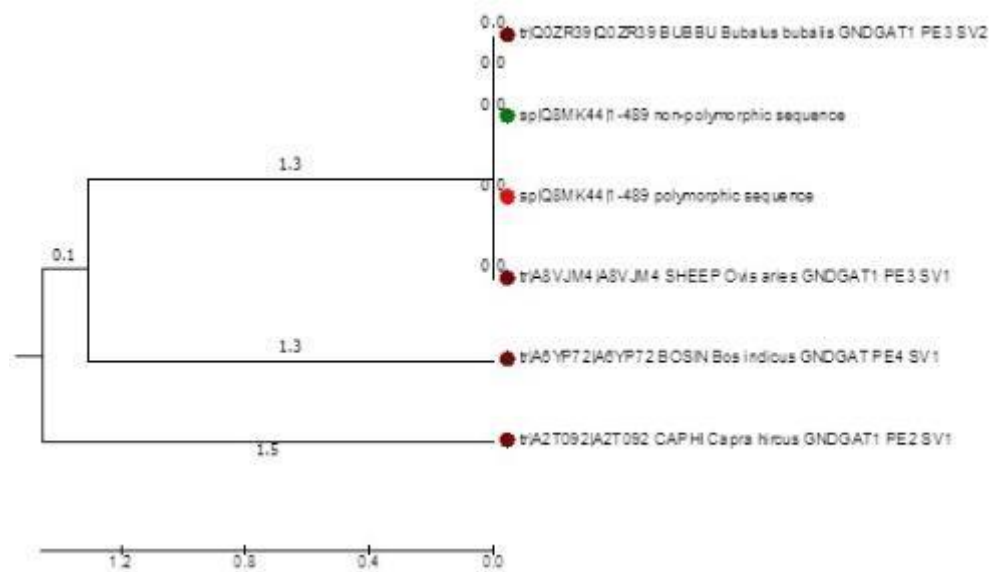


Figure 2. Evolutionary relationships among domestic ruminants inferred from amino acid sequences of the diacylglycerol-acyltransferase-1 (DGAT1) protein. Data of timescale are presented in millions of years.

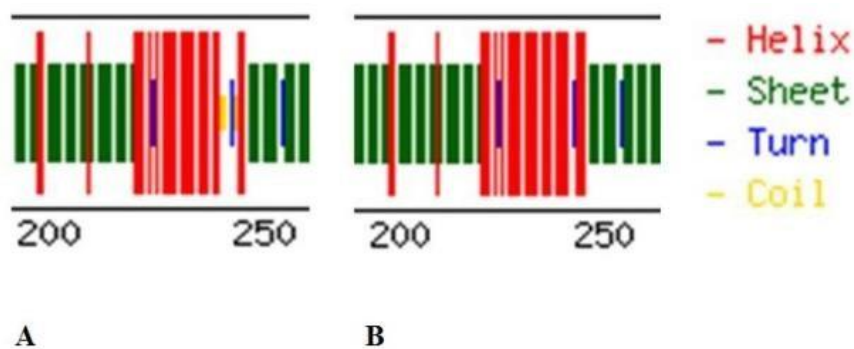


Figure 3. Prediction of the secondary structure of the non-polymorphic (A) and polymorphic harboring K232A (B) DGAT1 proteins of bovine *Bos taurus*.

DISCUSSION

The diacylglycerol-acyltransferase-1 lipogenic enzyme plays a fundamental role in the physiological pathways involved with the metabolism of triacylglycerol in higher eukaryotes, such as intestinal fat absorption, lipoprotein assembly, adipose tissue formation, and lactation (Cases et al., 1998, Yen et al., 2015). In this study, the conservation degree of multiple amino acid sequences of the DGAT1 protein and phylogenetical relationships among distinct species of the suborder Ruminantia were established using computational biology tools. Multiple comparative alignments of the DGAT1 amino acid sequences of domestic ruminants demonstrated an identity value corresponding to 77.505%, identifying 279 and 3 amino acid residues in identical and similar positions, respectively (Figure 1). Although amino acid sequences of the DGAT1 protein of goat and bovine *Bos indicus* available on database are yet considered partial segments, the findings reported herein revealed a phylogenetical clustering of specimens of the suborder Ruminantia intimately grouped in a main unique cluster (Figure 2). Such evidences suggest an accentuated conservation degree of DGAT1 protein sequences in domestic ruminants that seems to reflect possibly the closeness and the convergence of their evolution. Corroborating such findings, some studies have examined

the evolutionary relationships from multiple alignments of amino acid sequences of the DGAT1 protein. Venkatachalapathy et al., (2008) constructed a phylogenetic tree from derived DGAT1 protein sequences of 15 different species and illustrated a unique branch for mammals, fly, nematode and plants. According to Venkatachalapathy et al., (2008), cattle and buffalo were intimately grouped in the same branch, whereas swine was linked to another cluster. Within this perspective, Turchetto-Zolet et al., (2011), inferring the evolutionary phylogenetic analyses of the *dgat1* and *dgat2* genes across eukaryotic organisms, revealed that although these genes are widely divergent in molecular and structural aspects, such genic segments evolved at a functional convergence pattern.

Genic segments encoding the DGAT1 lipogenic enzyme emerge as a potential candidate gene, and the detection of a non-conservative substitution of lysine by alanine (K232A) provides a major effect on milk composition and yield. Notably, Kaupé et al., (2004) speculated that *dgat1^k* appears to be the ancestral allele and the K232A substitution probably occurred after the divergence of the *Bos indicus* and *Bos taurus* lineages. In fact, the evolutionary conservation of the affected lysine residue among mammals indicates the functional importance of a positively charged, hydrophilic residue at that position and that its substitution by a neutral, hydrophobic alanine residue could alter the functionality of this enzyme. In fact, amino acid substitutions with identical chemical properties signal the presence of conservative changes throughout sequences of interest. Moreover, high significance of the matches and conserved domains was identified by constraint-based analyses with e-value of 0.003, indicating low probability of the alignments have occurred by chance (Baxevanis and Ouellette, 2005).

Among all biochemical comparative parameters predicted from non-polymorphic and polymorphic DGAT1 protein, half-life values of 30 hours were estimated for both the DGAT1 proteins of bovine *Bos taurus*. By biological half-time is understood as the time taken by material or substance to lose every one-half of its quantity in the environment, or that of its concentration in a particular body organ (Boyle, 2005). According to Yu et al., (2002), the DGAT1 protein in the presence of a protease inhibitor and do not demonstrated a relatively long half-life of 18 h, supporting that posttranslational protein degradation plays a significant role in controlling steady-state levels of DGAT1 protein. Thus, accentuated half-life values found in the non-polymorphic and polymorphic harboring K232A DGAT1 proteins seem to support the strong ability of this peptide to exert their more extensively biological roles on neutral lipid biosynthesis, and assure consequently higher content of triglycerides on the milk composition and yield in domestic ruminants.

Moreover, the present study predicted biochemical aspects associated with the secondary structure of non-polymorphic and polymorphic harboring K232A DGAT1 sequences of bovine *Bos taurus* (Figure 3). In the biochemistry and structural biology, the protein secondary structure is the general three-dimensional form of local segments of proteins. Secondary structure is defined by the pattern of hydrogen bonds of the protein (such as alpha helices and beta sheets) that are visualized in an atomic-resolution structure. More specifically, the secondary structure is dependent of the patterns of hydrogen bonds formed between amine hydrogen and carbonyl oxygen atoms contained in the backbone peptide bonds of the protein (Boyle, 2005). The results described in this study demonstrated the presence of a coiled coil domain situated between amino acid residues at positions 200 and 250 in the non-polymorphic DGAT1 protein, whereas this coiled coil domain was not revealed in the same position in the polymorphic harboring K232A DGAT1 protein (Figure 3). Theoretical evidences point out that the random coiled coil domain is not a true typical secondary structure, but characterize as a class of conformations that indicate an absence of regular secondary structure (Boyle, 2005). Biophysical aspects associated with basic mechanism by which enzymes catalyze chemical reactions begin with the binding of the substrate (or substrates) to the active site on the enzyme, as postulated largely by Lock and Key model. Based on the prediction of the secondary structure of the polymorphic harboring K232A DGAT1 protein, the absence of coiled coil domain situated between positions 200 and 250 could be responsible for spatial modifications in the unique geometric shape of the active site of this enzyme, affecting the complementary binding of the substrate to the active site on the enzyme. Within this context, differentiated domains of the secondary structure of the DGAT1 protein harboring the polymorphism K232A may be exerting a positive influence on its biological action during neutral lipid biosynthesis.

In conclusion, findings described in this study provide insights into the phylogenetical clustering and evolutionary biology of domestic ruminants and the prediction of primary and secondary structures of the non-polymorphic and polymorphic harboring K232A DGAT1 proteins by comparing multiple alignments of biomolecule sequences using Bioinformatics' tools.

ACKNOWLEDGMENTS

This study was supported by the Bioinformatics and Computational Biology Group, designated "BIO in BYTES."

REFERENCES

- Baxevanis, AD; Ouellette, BFF. Bioinformatics: a practical guide to the analysis of genes and proteins. John Wiley and Sons Inc., New Jersey, 2005, 560p.
- Boutetm, E; Lieberherr, D; Tognolli, M; Schneider, M; et al., UniProtKB/Swiss-Prot. *Methods in Molecular Biology*, 2007, 406, 89-112.
- Boyle, J. Lehninger principles of biochemistry (4th ed.): Nelson D, and Cox M. *Biochemical and Molecular Biology Education*, 2005, 1119 p.
- Cases, S; Smith, SJ; Zheng, YW; Myers, HM; et al., Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proceedings of National Academy of Science USA, 1998, 95, 13018–13023.
- Cases, S; Stone, SJ; Zhou, P; Yen, E; et al., Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members. *Journal of Biological Chemistry*, 2001, 276, 38870–38876.
- Chou, PY; Fasman, GD. Conformational parameters for amino acids in helical, β -sheet, and random coil regions calculated from proteins. *Biochemistry*, 1974a, 13, 211-222.
- Chou, PY; Fasman, GD. Prediction of protein conformation. *Biochemistry*, 1974b, 13, 222-245.
- Gabriel, JE; de Figueiredo, DDLG; de Farias, RP. Revealing highly conserved regions in the E6 protein among distinct human papillomavirus types using comparative analysis of multiple sequence alignments. *Brazilian Journal of Biology*, 2013, 73, 449-450.
- Gabriel, JE; Lidani KC. Molecular conservation of the mammalian leptin protein. *Genetics and Molecular Research*, 2015, 14, 253-258.
- Gasteiger, E; Hoogland, C; Gattiker, A; Duvaud, S; et al., Protein identification and analysis tools on the ExPASy Server. (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press, 2005, 571-607.
- Grisart, B; Farnir, F; Karim, L; Cambisano, N; et al., Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of National Academy of Science USA*, 2004, 101, 2398-2403.
- Kaupe, B; Winter, A; Fries, R; Erhardt, G. DGAT1 polymorphism in *Bos indicus* and *Bos taurus* cattle breeds. *Journal of Dairy Research*, 2004, 71, 182-187.
- Lewin, TM; de Jong, H; Schwerbrock, NJ; Hammond, LE; et al., Mice deficient in mitochondrial glycerol-3-phosphate acyltransferase-1 have diminished myocardial triacylglycerol accumulation during lipogenic diet and altered phospholipid fatty acid composition. *Biochemical and Biophysical Acta*, 2008, 1781, 352-358.
- Liu, Q; Siloto, RM; Lehner, R; Stone, SJ; et al., Acyl-CoA:diacylglycerol acyltransferase: molecular biology, biochemistry and biotechnology. *Progress in Lipid Research*, 2012, 51, 350-377.
- Lu, C; Yang, R; Shen, B; Osman, H; et al., RNA interference-mediated knockdown of DGAT1 decreases triglyceride content of bovine mammary epithelial cell line. *Gene Expression*, 2012, 15, 199-206.
- Mackay, TF. The genetic architecture of quantitative traits. *Annual Reviews in Genetics*, 2001, 35, 303-339.
- Meegalla, RL; Billheimer, JT; Cheng, D. Concerted elevation of acyl-coenzyme A:diacylglycerol acyltransferase (DGAT) activity through independent stimulation of mRNA expression of DGAT1 and DGAT2 by carbohydrate and insulin. *Biochemical and Biophysical Research Communications*, 2002, 298, 317-323.
- Miltiadou, D; Tzamaloukas, O; Genethliou, N; Papachristoforou, C; et al., Do established SNPs affecting bovine milk traits exist in other dairy ruminants? *Small Ruminant Research*, 2010, 88, 72-74.

- Oliani, LC; Lidani, KC; Gabriel, JE. Differentiated evolutionary relationships among chordates from comparative alignments of multiple sequences of MyoD and MyoG myogenic regulatory factors. *Genetics and Molecular Research*, 2015, 14, 12561-12566.
- Papadopoulos, JS; Agarwala, R. COBAL: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, 2007, 23, 1073-1079.
- Schatz, MC; Delcher, AL; Salzberg, SL. Assembly of large genomes using second-generation sequencing. *Genome Research*, 2010, 20, 1165-1173.
- Shi, Y; Cheng, D. Beyond triglyceride synthesis: the dynamic functional roles of MGAT and DGAT enzymes in energy metabolism. *American Journal of Physiology, Endocrinology and Metabolism*, 2009, 297, E10-8.
- Smith, SJ; Cases, S; Jensen, DR; Chen, HC; et al., Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking *dgat*. *Nature Genetics*, 2000, 25, 87-90.
- Spelman, RJ; Ford, CA; Mcelhinney, P; Gregory, GC; et al., Characterization of the DGAT1 gene in the New Zealand dairy population. *Journal of Dairy Science*, 2002, 85, 3514-3517.
- Tăbăran, A; Balteanu, VA; Gal, E; Pusta, D; et al., Influence of DGAT1 K232A polymorphism on milk fat percentage and fatty acid profiles in Romanian Holstein cattle. *Animal Biotechnology*, 2015, 26, 105-111.
- Tamura, K; Battistuzzi, FU; Billington-Ross, P; Murillo, O; et al., Estimating divergence times in large molecular phylogenies. *Proceedings of National Academy of Science USA*, 2012, 109, 19333-19338.
- Tamura, K; Stecher, G; Peterson, D; Filipowski, A; et al., MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular and Biology Evolution*, 2013, 30, 2725-2729.
- Thaller, G; Kuhn, C; Winter, A; Ewald, G; et al., DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle. *Animal Genetics*, 2003, 34, 354-357.
- Turchetto-Zolet, AC; Maraschin, FS; de Moraes, GL; Cagliari, A; et al., Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. *BMC Evolutionary Biology*, 2011, 11, 263.
- Venkatachalapathy, RT; Sharma, A; Sukla, S; Bhattacharya, TK. Cloning and characterization of DGAT1 gene of Riverine buffalo. *DNA Sequence*, 2008, 19, 177-184.
- Winter, A; Kramer, W; Werner, FA; Kollers, S; et al., Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proceedings of National Academy of Science USA*, 2002, 99, 9300-9305.
- Yen, CL; Nelson, DW; Yen, MI. Intestinal triacylglycerol synthesis in fat absorption and systemic energy metabolism. *Journal of Lipid Research*, 2015, 56, 489-501.
- Yen, CL; Stone, SJ; Koliwad, S; Harris, C; et al., Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. *Journal of Lipid Research*, 2008, 49, 2283-2301.
- Yu, Y-H; Zhang, Y; Oelkers, P; Sturley, SL; et al., Posttranscriptional control of the expression and function of diacylglycerol acyltransferase-1 in mouse adipocytes. *Journal of Biological Chemistry*, 2002, 277, 50876-50884.