CARLOS SOUZA DO NASCIMENTO

EXPRESSÃO GÊNICA EM BIBLIOTECAS DE cDNA DE PELE DE BOVINOS F_2 (HOLANDÊS × GIR) INFESTADOS COM O CARRAPATO *RIPHICEPHALUS (BOOPHILUS) MICROPLUS*

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Zootecnia, para obtenção do título de *Doctor Scientiae.*

VIÇOSA
MINAS GERAIS - BRASIL
2009

CARLOS SOUZA DO NASCIMENTO

EXPRESSÃO GÊNICAEM BIBLIOTECAS DE cDNA DE PELE DE BOVINOS F2 (HOLANDÊS x GIR) INFESTEDOS COM O CARRAPATO *RIPHICEPHALUS (BOOPHILUS) MICROPLUS*

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Zootecnia, para obtenção do título de *Doctor Scientiae*.

APROVADA: 19 de fevereiro de 2009.

Dr. Marco Antonio Machado (Co-Orientador)	Dra. Marta Fonseca Martins Guimarães
Prof. Jose Bracini Neto	Prof. Paulo Luiz Souza Carneiro

Profª. Simone E. F. Guimarães
(Orientadora)

Para ser grande, sê inteiro: nada
Teu exagera ou exclui.
Sê todo em cada coisa. Põe quanto és
No mínimo que fazes.
Assim em cada coisa a Lua toda
Brilha, porque alta vive.

Ricardo Reis (heterônimo de Fernando Pessoa)

AGRADECIMENTOS

Agradeço a todos que, de forma direta ou indireta, estiveram presentes durante essa longa caminhada.

À Universidade Federal de Viçosa (UFV) e ao Departamento de Zootecnia (DZO), pela oportunidade de realização do curso.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

À Embrapa Gado de Leite (Juiz de fora, Minas Gerais), pela imprescindível contribuição na realização deste trabalho.

Agradeço à minha orientadora, Profa. Simone E.F. Guimarães, pela oportunidade de aprendizado e pela confiança demonstrada na realização desta pesquisa.

Ao meu co-orientador, Dr. Marco Antonio Machado, pela amizade, pelo incentivo e pelos ensinamentos, que contribuíram de forma significativa para minha formação.

Agradeço ao meu co-orientador, Prof. Paulo Sávio Lopes e a Dra. Marta Fonseca Guimarães, pelo apoio, pelas sugestões e pelos conselhos.

Aos amigos do Laboratório de Biotecnologia: Ana Paula, André, Braccini, Débora, Katiene, Mayara, Kleibe, Lucas, Marcos, Mário, Nicola, Paulinho, Priscila, pela amizade e pela vivência.

Agradeço imensamente aos grandes amigos, Danilo, Rogério, Márcio “Leitão”, Miller “Mister Bean”, pelo companheirismo.

Aos amigos do melhoramento, à Fernanda e Viviane da cantina do DZO.

Aos funcionários do Departamento de Zootecnia, especialmente os da Granja de Melhoramento de Suínos, na pessoa do Zé Geraldo.

Agradeço imensamente à minha família, aos meus pais, Paulo e Cleuza, irmãos e sobrinhos pela paciência com minhas constantes ausências.

Meus sinceros agradecimentos a todos os amigos que fiz durante os anos de Viçosa, Juiz de Fora e Belo Horizonte.

BIOGRAFIA

Carlos Souza do Nascimento, filho de Paulino Couto do Nascimento e Cleuza Teixeira de Souza, nasceu na cidade do Rio de Janeiro, Rio de Janeiro, no dia 7 de dezembro de 1972.

No primeiro semestre de 1998, ingressou no curso de Zootecnia da Universidade Federal de Viçosa (UFV), Viçosa, Minas Gerais, graduando-se em Zootecnia em maio de 2002.

No período de agosto de 2002 a fevereiro de 2005, foi bolsista de Apoio Técnico II (Fapemig), sob a orientação dos Doutores Mario Luiz Martinez e Marco Antonio Machado, ambos da Embrapa Gado de Leite.

No primeiro semestre de 2005, iniciou o curso de mestrado em Zootecnia na Universidade Federal de Minas Gerais (UFMG), onde recebeu o título de Mestre em julho de 2006.

Em agosto de 2006, iniciou o curso de Doutorado em Zootecnia na UFV. Submeteu-se aos exames finais de defesa de tese dia 19 de fevereiro de 2009 para obtenção do título de *Doctor Scientiae* em Zootecnia.

SUMÁRIO

RESUMO.....vi

ABSTRACT... viii

CAPÍTULO 1. 1

REVISÃO DE LITERATURA 2

OBJETIVOS.... 8

CAPÍTULO 2..... 9

Construction and characterization of cDNA libraries generated from skin tissue of F_2 cattle infested with the tick *Riphicephalus (Boophilus) microplus* 10

ANEXO 1.....52

CAPÍTULO 3.... 66

Avalição da cobertura gênica e do número esperado de novos transcritos em bibliotecas de cDNA ..... .67

ANEXO 2.....82

CAPÍTULO 4 87

Differential expression of calcium-binding proteins (TPT1 and S100a7) and calcium channel protein (TRPV6) in F_2 cattle infested with the tick *Riphicephalus (Boophilus) microplus*……...……..88

CAPÍTULO 5 102

Differential expression of the cystatin 6-like cysteine proteinase inhibitor gene (CST6) in bovine's skin tissues infested with the tick *Riphicephalus (Boophilus) microplus* ……103

CAPÍTULO 6…..... 114

DISCUSSÃO GERAL.....115

CONSIDERAÇÕES FINAIS... 120

REFERÊNCIAS…. 121

RESUMO

NASCIMENTO, Carlos Souza do, D.Sc., Universidade Federal de Viçosa, fevereiro de 2009. **Expressão gênica em bibliotecas de cDNA de pele de bovinos F_2 (Holandês × Gir) infestados com o carrapato *Riphicephalus (Boophilus) microplus***. Orientador: Simone Eliza Facioni Guimarães. Co-orientadores: Paulo Sávio Lopes e Marco Antonio Machado.

A resistência bovina ao carrapato *Riphicephalus* (*Boophilus*) *microplus* é herdável e está principalmente associada a animais zebu (*Bos indicus*), embora esteja também presente em menor nível em algumas raças taurinas (*B. taurus*). A elucidação da resistência genética é um dos fatores mais promissores para reduzir as perdas de produção e diminuir o custo de controle desse parasita na pecuária bovina tropical. Com o intuito de caracterizar genes funcionais envolvidos na resistência/susceptibilidade dos bovinos ao carrapato, duas bibliotecas de cDNA foram construídas a partir de pele de animais F_2 (Holandês × Gir) resistentes e susceptíveis infestados com larvas de *R. microplus*. Foram sequenciadas 4.070 etiquetas de sequência expressa (*EST-Expressed Sequence Tag*) geradas a partir de amostras de pele de bovinos F_2 (Holandês × Gir) infestados com o carrapato *R. microplus*. Do total de *EST* geradas para as duas bibliotecas, foram obtidas 2.700 sequências de alta qualidade. Os resultados do agrupamento geraram um conjunto não-redundante de 1.292 sequências únicas. Cerca de 790 destas sequências compartilharam similaridade significativa com sequências de proteínas conhecidas e 502 destas não apresentaram similaridade com sequências de proteínas presentes no banco de proteínas não-redundante (*NCBI - National Center of Biotechnology Information*). A análise do perfil funcional dos transcritos permitiu identificar 54 termos de ontologia gênica (*GO-terms*) significativamente ($P<0,01$) representados nos conjuntos de dados de res e sus quando comparadas ao genoma de *B. Taurus*. Foi estimada a porcentagem de genes presentes nos conjuntos de dados de ESTs. As predições para cobertura gênica foram de 49% (RES) e 40% (SUS). QRT-PCR foi usada para determinar o nível de expressão de quatro genes identificados nas bibliotecas de cDNA. As expressões relativas dos genes S100A7, TPT1, TRV6 e CST6

foram 2,01 (±0.6), 1,32 (±0.9), 1,53 (±1.2), 2,03 (±0.6), respectivamente. Esses resultados indicam que esses transcritos foram diferencialmente expressos (P=0.001) em lesões de pele de animais susceptíveis. No entanto, a expressão aumentada desses genes não parece conferir proteção aos animais suscetíveis à infestação com carrapato. Até o momento, nenhum estudo de genômica funcional com animais cruzados (Holandês × Gir) tem sido relatado. Os transcritos gerados neste estudo podem contribuir de forma substancial para o melhor entendimento da genômica funcional da interação parasita-hospedeiro neste tecido.

ABSTRACT

NASCIMENTO, Carlos Souza do, D.Sc., Universidade Federal de Viçosa, February 2009. **Gene expression in skin cDNA libraries from F_2 cattle (Holstein × Gyr) infested with the tick *Riphicephalus (Boophilus) microplus*.** Adviser: Simone Eliza Facioni Guimarães. Co-Advisers: Paulo Sávio Lopes and Marco Antonio Machado.

It is known that resistance to cattle tick *Riphicephalus (Boophilus) microplus* is inheritable and are primarily associated with zebu animals (*Bos indicus*), although also present at lower levels in some breed taurine (*B. taurus*). The elucidation of genetic resistance is one of the most promising factors to reduce production losses and to reduce the cost of this parasite control in livestock tropical bovine. In order to characterize the functional genes in involved resistance / susceptibility of cattle to ticks, two of cDNA libraries were constructed from skin of F_2 animals (Holstein × Gyr) resistant and susceptible infested with larvae of *R. microplus*. We sequenced 4,070 Expressed sequence tags (ESTs) from skin F_2 cattle (Holstein × Gyr) infested with the tick *R. microplus*. From total of generated ESTs, 2,700 high quality sequences from two cDNA libraries were obtained. Clustering results generated a non-redundant set of 1.292 unique sequences. About 790 of the sequences shared significant similarity with known protein sequences and 502 of these gave no similarity to protein sequences deposited on non-redundant (nr) protein database (NCBI - National Center of Biotechnology Information). The functional profile analysis of transcripts indicated 54 Gene Ontology terms (GO-terms) significant ($P<0.01$) over represented on res and sus datasets in relation to *B. taurus* genome. We also estimated the percentage of genes presents on ESTs dataset. The gene capture predictions were 49% (RES) e 40% (SUS). Real-time polymerase chain reaction was used to determine the gene expression level to four genes identified on cDNA libraries. The relative expression of the S100A7, TPT1, TRV6 and CST6 genes was 2.01 (±0.6), 1.32 (±0.9), 1.53 (±1.2), (2.03±0.6), respectively. Our findings show that these genes were differentially expressed ($P=0.001$) in skin lesions from susceptible animals. However, the increased expression of these genes does not appear to confer protection against to infestation with ticks. To date, no study of functional

genomics to crossbreed animals (Holstein x Gyr) has been reported. The transcripts generated in this study can contribute substantially to a better understanding of the functional genomics of host-parasite interaction in this tissue.

CAPÍTULO 1

1. REVISÃO DE LITERATURA

1.1. O parasita *Riphicephalus (Boophilus) microplus* e sua importância econômica

O *Riphicephalus (Boophilus) microplus* é um ectoparasita de bovinos encontrado em regiões tropicais e subtropicais do planeta. Recentemente sugeriu-se que o gênero *Boophilus* seja parafilético ao gênero *Riphicephalus*, de modo que o *Boophilus* pode ser classificado como subgênero do *Riphicephalus* (Barker e Murrel, 2002). O carrapato é originário da Ásia, com foco de distribuição na Índia e na ilha de Java. No Brasil, segundo Andreoti (2002), sua introdução deu-se com a vinda de animais do Chile no início do século XVIII em regiões onde hoje se localiza o estado do Rio Grande do Sul.

Em termos econômicos, o carrapato representa grande problema na produção de bovinos em diferentes regiões. Atualmente, as perdas econômicas causadas por esse parasita decorrem, principalmente, do parasitismo bovino.

Segundo Gonzales (1995), a ação expoliativa da fêmea do carrapato promove a perda de sangue (2-3 mL) do bovino, promovendo redução na produção de leite e carne, gerando perda média anual de 0,24 kg de peso vivo por carrapato (Sutherst et al., 1982). Além disso, são também transmissores dos protozoários *Babesia bovis* e *B. bigemina* e de riquétsias do gênero Anaplasma, agentes causadores da doença denominada tristeza parasitária bovina (Horn e Arteche, 1985). O *R. microplus* causa ainda perdas associadas a danos ao couro dos bovinos pelas reações inflamatórias provocadas no local de fixação (Seifert et al., 1968; Horn e Arteche, 1985).

Possuidor do maior rebanho comercial de bovinos do mundo com aproximadamente 170 milhões de cabeças (Brasil, 2003), cerca de 18% do rebanho brasileiro é destinado à bovinocultura de leite e o restante à atividade de corte. Estima-se que cerca de 80% do rebanho bovino mundial esteja infestado por carrapatos, sendo este o ectoparasito de maior impacto econômico na pecuária de leite e corte nacional, por ocasionar um prejuízo anual estimado em 2 bilhões de dólares (Grisi et al., 2002). Horn (1988) descreveu gasto anual de 13,8 milhões de dólares somente com acaricidas, o

que representaria 15% do gasto total do País com defensivos na agropecuária. Além disso, a fixação do carrapato em seu hospedeiro provoca lesões, que se tornam portas de entrada para infecções bacterianas e miíases, acarretanto depreciação do couro durante o beneficiamento nos curtumes (Sauer et al., 1995).

Além dos danos diretos, as perdas indiretas causadas pelos carrapatos são representadas pelos gastos com mão-de-obra, medicamentos, construções e medidas preventivas necessárias no combate ao ectoparasito. (Gomes, 1998). O uso de acaricidas, que em sua maioria, são substâncias tóxicas (organoclorados, organofosforados, carbamatos, amitraz, piretróides sintéticos e ivermectina) tem sido a medida de controle mais comum contra esses ectoparasitas. Os principais problemas relacionados a essa prática são o desenvolvimento de linhagens resistentes de carrapatos, o aparecimento de resíduos químicos nos produtos de origem animal e a poluição ambiental causada pelo uso de acaricidas (Bullman et al., 1996).

1.2. Interação hospedeiro-carrapato

Embora algumas espécies hospedeiras desenvolvam mecanismos de resitência contra carrapatos, outras falham em adquiri-los. Essa falha na elaboração de imunidade efetivamente protetora pode estar relacionada ao mecanismo de escape do ectoparasita, que o torna capaz de modular a resposta de algumas espécies de hospedeiros. Além disso, algumas espécies hospedeiras desenvolvem resistência a essa modulação, tornando-se refratárias a carrapatos.

Sabe-se que moléculas liberadas pelo carrapato *R. Microplus* durante a ligação e a alimentação da larva ao hospedeiro estimulam respostas imunológicas inatas e adquiridas. A habilidade do hospedeiro em responder a essas moléculas irá resultar em diferentes níveis de resistência. O carrapato contra-ataca a resposta do hospedeiro com moléculas imunosupressivas presentes na sua saliva. A saliva do *R. microplus* contém substâncias que afetam os linfócitos T, macrófagos, neutrófilos e células naturais de defesa e influencia a classe de imunoglobulina produzida em resposta ao antígeno do

carrapato. Um exemplo é a saliva de carrapatos Ixodes, a qual contém atividade antihemostática, antiinflamatória, inibidora da atividade da bradicinina, de anafilatoxinas e desativadora de neutrófilos (Ribeiro et al., 1985; 1990). Particularmente, as proteinases, peptidases e seus inibidores têm sido foco em pesquisas com ectoparasitas. A importância desse grupo de enzimas e inibidores tem sido evidenciada em diversos trabalhos realizados nos últimos anos, os quais descrevem a existência de diferentes proteinases e seus inibidores (Mulenga et al., 1999; Miyoshi et al., 2004; Seixas et al., 2003; Andreotti et al., 2002).

1.3. Variabilidade genética da resistência ao carrapato em bovinos

Segundo Mattioli et al. (2000), a resistência ao ataque de carrapatos varia entre as raças bovinas. Em geral, animais *B. indicus* são mais resistentes que os *B. taurus*. Ao longo do processo de domesticação, os bovinos foram submetidos à intensa pressão de seleção para vários fenótipos. O gado indiano tem convivido há milhares de anos com o carrapato *R. microplus* e provavelmente houve eliminação natural dos animais mais sensíveis, permitindo maiores oportunidades reprodutivas para os animais geneticamente resistentes (Lemos et al., 1986).

Em bovinos, a resposta da interação carrapato-hospedeiro varia de acordo com a raça bovina estudada e, mesmo depois de repetidas infestações, animais de raças suscetíveis apresentam maior carga de ectoparasitas que os de raças resistentes (Mattioli et al., 2000). A suscetibilidade do bovino é herdável e a resposta imune do hospedeiro exerce importante função, mas os mecanismos envolvidos não são completamente entendidos. Diversos autores (Hewetson, 1968; Wharton et al., 1970; Teodoro et al., 1984; Madalena et al., 1985) obtiveram estimativas de herdabilidade variando de 20 a 49% em zebuínos e mestiços, enquanto Veríssimo et al. (1997) observaram baixas estimativas de herdabilidade (0,89 a 0,91%) em mestiços leiteiros.

A seleção natural de zebus pelo parasitismo do carrapato ocasionou, provavelmente, acúmulo de grande quantidade de genes de pequenos efeitos, o que é característico de herança poligênica (quantitativa). Segundo Martinez et al. (2004), esse tipo de resistência poligênica promove resposta rápida e

efetiva à seleção em raças de moderada a alta resistência, mas não em raças suscetíveis. A seleção artificial produziu tipos distintos, associados principalmente à produção de carne e de leite. Em algumas regiões do mundo, raças se tornaram especializadas, por exemplo, para transporte de cargas em altas altitudes ou se adaptaram a ambientes tropicais, desenvolvendo características como tolerância a temperatura adversa e patógenos específicos, ou mesmo resistência a doenças.

A seleção de animais de raças suscetíveis com base na resistência poligênica não é viável, por isso é necessária a exploração de genes de efeito maior (*Major gene*) que possam estar associados à resistência a carrapatos (Martinez et al., 2004). Vários alelos DRB3 classe II do complexo principal de histocompatibilidade (BoLA) têm sido associados a resistência (Martinez et al., 2006). Em raças suscetíveis, infestações do bovino com *R. microplus* tem comprovado que anticorpos IgG do hospedeiro são suprimidos por moléculas presentes na saliva do parasita (Kashino et al., 2005).

Utech et. al. (1978) observaram em rebanhos zebuínos grande proporção de animais com alta resistência ao carrapato, enquanto Villares (1941) verificou maior resistência da raça Zebu ao comparar o número de carrapatos encontrados em bovinos de diversas raças. Apenas 5% do total de carrapatos eram originários de animais Zebus, 7% de raças nacionais ou criolas e 88% de raças européias.

Na Austrália, os animais Zebus têm sido utilizados intensamente em cruzamentos com raças européias, por sua maior resistência ao carrapato. As diferenças de resistência entre os *B. taurus* e os produtos do cruzamento *B. taurus* × *B. indicus* foram demonstradas por Byford et al. (1976), que observaram, em geral, resistência de moderada a alta nos animais oriundos do cruzamento.

No Brasil, Teodoro et al. (1984) estudaram a resistência de touros mestiços (5/8, 3/4 e 7/8 Taurino × Zebuíno) sob infestação artificial com carrapatos e observaram que os animais 5/8 foram mais resistentes ao carrapato, enquanto os 7/8 apresentaram menor resistência. Lemos et al. (1985), estudando a resistência ao carrapato em novilhas de grupos genéticos com diferentes proporções de gene zebu, também observaram maior carga

parasitária e menor resistência nos animais com maior porcentagem de genes europeus.

Kerr et al. (1994) encontraram evidências para existência de um gene principal para resistência a carrapatos em uma linhagem proveniente do cruzamento das raças Hereford e Shorthorn. Essa linhagem foi parcialmente comercializada com o nome de Belmont Adaptaur ou HS e, em condições genéticas apropriadas, pode conferir até 100% de resistência (Frisch, 1999). Em mais de 30 anos de trabalho, observou-se que cada cópia do gene Adaptaur no DNA dos animais reduziu sequencialmente a contagem dos carrapatos em 75% e que a frequência desse gene na população foi de 25% (Frisch, 1994).

A utilização de cruzamentos visando à combinação de características de resistência ao carrapato e à tolerância ao calor do *B. indicus* com a alta produtividade leiteira e fertilidade do *B. taurus* foi sugerida por Rendel (1971). Observando que o gado zebuíno (*B. indicus*) é menos suscetível ao carrapato, criadores de bovinos têm realizado manejo reprodutivo buscando um rebanho mais resistente a esse ectoparasito. A realização de cruzamentos têm resultado frequentemente em perdas de características desejáveis, como a alta produtividade de leite do gado europeu. Desse modo, um melhor entendimento da ação gênica dos carrapatos sobre os hospedeiros torna-se fundamental na desestabilização dessa relação.

1.4. Genômica funcional de bovinos

Com a conclusão do projeto de sequenciamento do genoma bovino e a disponibilidade de tecnologias de elevado processamento, a genômica funcional pode agora ser usada para investigar simultaneamente alterações na expressão de milhares de genes decorrentes de desafios ambientais e fisiológicos. Um dos principais objetivos nos estudos em genômica realizados em mamíferos é caracterizar o padrão de expressão gênica que corresponde a eventos fisiológicos importantes relacionados à produção e à saúde dos animais. Um recurso necessário para atender esse objetivo é obter sequências dos clones que representam a expressão do mRNA presente em uma célula ou tecido submetido a determinada circunstância de estudo. Cada uma dessas

sequências é chamada Etiqueta de Sequência Expressa (*EST - Expressed Sequence Tag*).

Atualmente, existem inúmeros recursos disponíveis para estudo de genômica em bovinos, entre eles, o rascunho do genoma bovino, baseado na cobertura de 7X do genoma (www.hgsc.bcm.tmc.edu/projects/bovine/). Existem aproximadamente 8.163.902 EST disponíveis de humanos, 4.850.605 de camundongos, 1.532.429 de suínos, 1.517.143 de bovinos taurinos e apenas 19.458 EST de zebuínos (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). Outra importante fonte de informação de EST pode ser encontrada no Bodymap-Xs (http://lifesciencedb.jp/bodymap-xs), um banco de dados que com informações de expressão gênica para inúmeras espécies, permitindo realizar comparações de expressão para genes ortólogos e parálogos (Ogasawara et al., 2006).

Um maior número de informações sobre os transcritos é fundamental para interpretar resultados de experimentos de genômica funcional, especialmente em razão da escassez de informações disponíveis acerca das sequências gênicas de zebuínos. Essas raças dominam o mercado de produção bovina nas regiões tropicais e a geração de EST provenientes desses animais é importante para construir ferramentas de estudos genômicos que possam ser utilizadas na caracterização da diversidade genética entre os animais de origem taurina e zebuína quanto à saúde e produção. Diante da importância dos animais zebus e mestiços na pecuária leiteira nacional, é evidente a necessidade de se aumentar o número de EST. Este trabalho foi realizado com o objetivo de gerar informações de transcritos expressos na pele de bovinos F_2 (Holandês × Gir), avaliados como resistentes e suscetíveis, submetidos a infestação com *R. microplus.*

OBJETIVOS

Este trabalho teve como objetivo geral analisar os perfis de expressão gênica em bovinos infestados com *R. microplus.* Os objetivos específicos constituíram-se em:

1. Comparar a expressão gênômica funcional entre animais resistentes e suscetíveis;
2. Comparar o perfil de cobertura gênica nas *EST* amostradas nas duas bibliotecas de cDNA;
3. Avaliar a expressão gênica diferencial entre grupos de animais resistentes e suscetíveis;

CAPÍTULO 2

Construction and characterization of cDNA libraries generated from skin tissue of F_2 cattle infested with the tick *Riphicephalus (Boophilus) microplus*

C. S. Nascimento[A], M. A. Machado[BC], S. E. F. Guimarães[A], M. F. M. Guimarães[B], J. O. Peixoto[A], J. Furlong[B], M. C. A. Prata[B], R. S. Verneque[B], R. L.Teodoro[B], P. S. Lopes[A]

[A]Department of Animal Science, Federal University of Viçosa - UFV, *Campus Universitário, 36570-000, Viçosa- MG. Brazil*
[B]Embrapa - Brazilian Dairy Cattle Research Center, *Eugênio do Nascimento, 610, Dom Bosco, 36038-330, Juiz de Fora- MG. Brazil*
[C]*Corresponding author. E-mail: machado@cnpgl.embrapa.br*

Abstract

Background: Genetic resistance to bovine tick *Riphicephalus (Boophilus) microplus* is one of the most important factors to reduce costs related to the control of this parasite in tropical cattle systems. Tick resistance in cattle is chiefly associated with zebu (*Bos indicus*) animals, although it is also present in some taurine (*B. Taurus*) breeds.

Results: In order to characterize functional genes involved in tick resistance /susceptibility in cattle, two cDNA libraries was constructed from skin cattle F_2 animals (Holstein × Gyr). The expressed sequence tag (EST) approach was used to compare transcripts distribution from resistant (res) and susceptible (sus) animals infested by tick *R. microplus*. A total of 2,700 high quality sequences (450 contigs and 889 singletons) from two cDNA libraries were sequenced and compared with sequences in the GenBank database. Of the 1,292 EST, 790 (61.1%) sequences shared significant similarity with known sequences in the NCBI non-redundant (nr) protein database, whereas 502 (38.9%) sequences produced no hit and are likely to represent newly described genes. The protein family and domain analysis revealed that the most frequently found gene was that of ribosomal protein with 35 and 55 homologous followed by binding proteins 22 and 14, keratin and collagen 13 and 12, to res

and sus libraries, respectively. Gene Ontology (GO) analysis was performed to indicate the possible functions of these genes. Pearson chi-square tests of gene ontology terms in the 3nd level between res and sus set showed no significance. However, digital gene expression profiling using the *B. taurus* genome showed 54 significant differentialy expressed.

Conclusion: The sequences we have identified in this work are a valuable resource for future studies on functional genomics using cattle and represent an important source for genomic sequencing projects, with partial or complete sequences being made available for over 502 previously unsequenced bovine genes.

Key words: animal breeding, functional genomics, gene expression

Background

Riphicephalus (Boophilus) microplus is a cattle ectoparasite found in tropical and subtropical regions worldwide (Willadsen and Jongejan, 1999). In Brazil, this tick represents a great problem for cattle industry nation-wide and the use of acaricides has been the most common control measure against this ectoparasite. Due to its serious impact on cattle production, there is an urgent need to understand and to unveil the underlying mechanisms involved in tick resistance/susceptibility in cattle.

Genetic resistance to the bovine tick is one of the most important factors to reduce costs related to the control of this parasite in tropical cattle systems. Differences in domestication and selection processes have contributed to considerable phenotypic and genotypic differences between Bos taurus taurus and Bos taurus indicus cattle breeds. In general, *Bos indicus* animals are more resistant to parasitic diseases than *Bos taurus*.Villares (1941) reported high susceptibility of cattle breeds of European origin to *B. microplus* in Brazil, including Holstein and Brown Swiss dairy breeds.

Little is known so far about the genetic mechanisms involved in the genetic resistance of zebuínos animals to ticks. Genetic resistance can contribute to decrease the tick population, costs with medications, mortality and production losses and also to decrease environment pollution and intoxication of the animal. To identify all regions involved in tick resistance, Martinez et al. (2006) developed an F_2 originated from crossing F_1 animals (½ Holstein: ½ Gyr) *B. taurus* x *B. indicus* F_2. An total of 360 F_2 animals were evaluated to parasite load in two seasons (rainy and dry) and genome scan were performed with microsatellite markers a cattered throughout the whole genome (unpublished data).

Expressed sequence tags (*ESTs*) analysis, which survey sequences contained in cDNA libraries, is a powerful approach for identifying new genes and profiling gene expression in tissues or cells. Genomic information is becoming more abundant for many of the species involved, what could be used to develop new control strategies. To understand the function of skin under infestation with ticks parasite, an analysis of abundantly expressed genes in the tissue is needed. To identify these host genes, one strategy is to isolate genes

that are differentially expressed after tick infestation. The identification of new genes and host antigens involved in the mechanism of resistance/susceptibility to the parasite are a promising approach. (Douglas et al., 1999).

In bovine, the epithelium serves as the first line of barrier against parasites and the innate immune response constitutes the first line of defense against pathogen invasion. (Gumbiner, 1993). Disturbances in this barrier can lead to the invasion of microorganisms and pathogenic agents, causing subsequent desease. So far, few genes related to the nonspecific immunity have been isolated and characterized in cattle. The innate immune response plays an important role in protecting cattle against foreign invasion

There is extensive interest in enhancing resistance of dairy cattle to tick and tick-borne diseases. The screening for immune-relevant functional genes in cattle is important to identify the molecular mechanism for tick resistance. Therefore, we conducted this experiment to clone cDNA sequences of skin expressed genes from F_2 cattle to test the hypothesis that in addition to housekeeping genes, the genes involved in resistance/susceptibility mechanism are abundantly or less expressed in skin tissue, and to study the distribution of genes abundantly expressed.

Methods

Animals

The animals used in this study belonged to an F_2 population originating from the crossing of F_1 females (50% Gyr: 50% Holstein) with F_1 sires of the same genetic composition. All F_2 animals were raised together on an Embrapa's experimental farm located on the State of Rio de Janeiro, Brazil. The climate corresponds to Cwa of Koppen's classification (Koppen and Geiger, 1936) mild, dry winter and hot summer, with the dry season extending from April to September (Teodoro and Madalena, 2003). Animals were studied in age contemporary groups ranging from 10 to 14 months old. To determine tick resistance, a total of 360 F_2 animals were artificially infested with approximately 10,000 tick larvae by placing them in the "dorsal-lumbar" region of the animals. The absolute number of ticks determines the level of resistance of each animal. The counts of adult female ticks were done 21st day after infestation. Tick counts ranged from zero to 792 ticks per animal, showing an extreme genetic variability in the F_2 population. Animals with extreme breeding values higtest and lowest for tick resistance/susceptibility in this herd were selected for the gene expression study. Six tick-resistant and six tick-susceptible F_2 animals were selected based on their breeding value for the collection of skin biopsies.

Tissue colleting

Skin samples were collected on day 5 and 12 after artificial infestation. Skin biopsies were obtained from the lesion area provoked by the ticks in susceptible animals and from the healthy area of resistant animals. The animals continued to be monitored until day 21 when semi-engorged females measuring 4.5 to 8.0 mm in diameter were counted. This procedure was used to determine the level of resistance/susceptibility the selected animals.

RNA Preparation

After tissue colleting, 800 mg of biopsies from each animals was pooled in to two pools and the extraction of total RNA was perfomed. Total RNA was

extracted using the RNeasy® Maxi kit (Qiagen, Valencia, CA) according to manufacturer instructions. The quality of the RNA was monitored by examination of the 18S and 28S ribosomal RNA bands after electrophoresis and quantified by spectrophotometry at 260 nm. Poly(A) RNA from skin tissue was isolated using Oligotex® kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

Construction of skin cDNA Libraries

Superscript™ Plasmid System with Gateway TM Technology for cDNA Synthesis and Cloning kit (Invitrogen, Carlsbad, CA, USA) was used for construction of the two cDNA libraries. Poly (A) RNA obtained from a pool of equal amounts of RNA from each of two groups was reverse transcribed and equimolar amounts cDNA of the reverse transcription product was used to construct the libraries. The cDNA products were then size fractioned by column chromatography according following the manufacturer's instructions. Fractions were collected and molecular weight was monitored by agarose gel electrophoresis. Fractions greater than 0.3 kb were combined (fractions 7 to 13) and used for non-normalized skin cDNA libraries construction. The cDNA was then ligated to restriction enzyme digested pSPORT1 vector resulting in directional cloning. After the ligation, plasmids were transformed into Echerichia. coli DH5α Ultracompetent cells (Invitrogen, Carlsbad, CA, USA). Colonies resulting from transformation were plated in Agar medium containing ampicillin. Individual colonies were randomly picked and inoculated into 96-well microtiter plates containing 200 µl of 2xYT medium and incubated for 16 h at 180 rpm. Each microtiter plate was used into 96 deepwell plates for plasmid DNA preparation cultures containing 900 µl of 2xYT medium and incubation took 24 h at 180 rpm (Sambrook and Russel, 2001).

Plasmid DNA Purification

Plasmid DNA purification followed an alkaline lysis procedure developed by Birnboim and Doly (1979), with some modification to fit a high-throughput plasmid DNA preparation using 96-well PVDF filter plates (Corning, NY, USA).

To resulting sobrenadanct, 30 % PEG-8000 in 1.6 M NaCl was added to achieve a final concentration of 7.5% PEG-8000. Plasmid DNA was allowed to precipitate from this solution for 24 hr at 4°C. Plasmid DNA was collected by centrifuging at 10,000 X g for 45 minutes. The pellet containing the plasmid was resuspended in 30 µl of buffer (10mM Tris, 1 mM EDTA , 150 mM NaCl, pH 8.0). Miniprep product quality control was done in 0.8 % agarose gels to assure positive clones and to determine size insert. Positive clones where rearrayed and diluted to appropriate concentrations. Only clones possessing insert size greather than 0.3 kb were selected for DNA sequencing.

Nucleotide single-pass sequencing

Sequencing reactions were performed with 200 ng of plasmid DNA template prepared in a 96-well microplates using DYEnamic ET dye terminator kit (Amersham Pharmacia Biotech, Piscataway, NJ) with M13 Reverse Primer or M13 Forward Primer, followed by sequencing reaction clean up to remove residual dye and enzyme. Unidirectional single-pass sequencing was performed on a MegaBACE 1000 capillary sequencer (GE Healthcare, Germany).

Quality assessment and feature editing of ESTs sequences

To verify the quality of the EST sequences and identify common features before clustering, the raw traces were processed for submission using the trace2dbest (Parkinson et al, 2004) and submitted to dbEST database (NCBI). *ESTs* sequences were extracted from the raw data then rigorously screened with default parameters to remove contaminating vector (minimatch 12, penalty -2, minscore 20), poly-A/T tails (8), primers e adaptor sequences. Contaminating bacterial sequences were masked using software Cross-Match (Green, 1999), and eliminating any sequence of low-quality regions at both ends using the trim option in the Phred with a cutoff of 20 (equivalent to 0.05 of sequencing error) (Ewing et al., 1998). The comparison quality was set greater than 80% identity over at least a 150-bp fragment acceptable after trimming were subject to further analysis.

Clustering and contig assembly

The *ESTs* were clustered and assembled into contigs using CAP3 (Contig Assembling Program 3) (Huang and Madan, 1999) set to default parameters. *ESTs* that did not form contigs designated singlets. Collectively, the resultant contigs and singlets are referred to as unique sequences.

Sequence similarity searches

The trimmed sequences then undergo a standalone Blast against nonredundanct databases (nucleotide and protein databases) from NCBI GenBank using BlastN and BlastX algorithms (Altschul et al., 1990). GenBank databases currently used are available at ftp://ftp.ncbi.nih.gov/blast/db/ using an *E*-value of 1e-05. After analysis, a putative product was assigned to the results from the similarity search to homologous genes and proteins.

Gene ontology analysis and functional profile

Assignments to the Gene Ontology (GO) database were performed using Blast2GO program (Conesa et al., 2005). The GO terms (http://www.geneontology.org) to the resulting contigs and singlets were automatically assigned against curated InterPro database. Frequency of each functional category was then summarized and reported in a pie chart format at an expect value of 10^{-6} and cutoff 80 to similarity. To assess the statistical significance of terms, we used FatiGO (Al-Shahrour *et al.*, 2004) to extract GO categories over-or-under represented.

Results

cDNA libraries and EST sequence quality analysis

Two independent bovine skin cDNA libraries from resistant and susceptible animal's tissues were sequenced to generate EST datasets. Single-pass sequencing was performed from 5'-end of 3,563 cDNA and 3'end of 507 clones randomly picked from the two nonnormalized cDNA libraries. The efficiency and reproducibility of our protocols were corroborated by comparing two sets of ESTs for the *res* and *sus* cDNA libraries performed at same time. Of 4,245 total clones, 175 sequences were excluded due to low sequencing quality and after vector sequence trimming. Thus, finally 4,070 ESTs with high quality sequences were used for computational analysis, 1,885 from the resistant and 2,185 from the susceptible cDNA libraries. A summary with the statistic of ESTs can be visualized in the Table 1.

Table 1: Statistic summary of ESTs from two cDNA libraries

	Libraries		
Sequencing analysis	Resistance (*res*)	Susceptible (*sus*)	Total
Category	number of sequences and percentage (%)		
ESTs sequenced	3'-279 (18.8)	3'-228 (10.4)	507 (12.5)
	5'-1,606 (85.2)	5'-1,957 (89.6)	3,563 (87.5)
subtotal	1,885 (46.3)	2,185 (53.7)	4,070
Clusterization and assembly			
Hight quality ESTs	1,235 (45.7)	1,465 (54.3)	2,700 (66.3)
Unique sequences	738 (55.1)	601 (44.9)	1,339
contigs	226 (30.6)	223 (37.1)	450 (33.6)
singletons	512 (69.4)	366 (62.9)	889 (66.4)
Mean unique sequence lenght (nt)	471,3	424,1	447,7

Clustering and contig assembly

The ESTs that passed through the above quality check procedures were considered high quality ESTs. These sequences from both resistance and susceptible libraries were assembled to form contigs using CAP3 with default

parameters. ESTs that did not form contigs were designated singlets. Collectively, the resultant contigs and singlets were referred to as unique sequences.

After pre-processing, 1,885 clones from the *res* dataset have yielding 1,235 acceptable sequences (65.55 % of the sequenced clones). For the *sus* library, 2,185 clones have been sequenced, producing 1,465 acceptable ESTs (67.00 % of the original total). Higher quality ESTs sequences were deposited at the GenBank dbEST database (Boguski et al., 1993), jointly with their Blast-based preliminary annotation. The accession numbers assigned to them were GO578626 to GO581239 and dbEST-Id 64942695 to 64945308. Low quality sequences, including short sequences (less than 150-bp) were considered uninformative and were eliminated from the analysis. A total of 738 unique sequences were obtained from the *res and* 601 from the *sus* cDNA libraries. The combined set of contigs and singletons from the two libraries, resulted in 1,339 unique sequences 450 (33.6%) contigs and 889 (66.4%) singletons, respectively with an average size of 471,3 and 424,1 nucleotides to *res* and *sus* libraries, respectively (Table 1).

Sequence similarity searches

Each unique sequence was subjected to BlastN and BlastX algoritms searches against the nucleotides and protein non-redundanct (nr) Genbank databases at the National Center for Biotechnology Information (NCBI, Bethesda, MD) for the identification of putative homologies to known genes and proteins.

BlastN homologous searches: The unique sequences from the 1,339 unique sequences were aligned against nr nucleotide database. The sequences alignment and filtered using BlastN searches ($E < 0.00001$ and similarity > 80%) to eliminate putative match to cloning vector and *E.coli* sequences. The alignment resulted in 1,292 unique clean sequences. Thirty six unique sequences produced alignments with vector sequence and two was hits to *E. coli* sequences on library *res*. Nine sequences on library *sus* that showed to be

a contamination vector sequences were removed. Statistics of blast analysis are shown in Table 2. These unique sequences, based on their matches to expression vector sequences and *E. coli* were not further considered (Table 2 S1.xlsx).

Table 2. Statistic summary of BlastN vs Non-redundant nucleotide database (nr) >80%

Number of sequences and percentage (%)			
Libraries	*res*	*sus*	**Total**
No. of unique sequences	738	601	1,339
No. of contaminants	38	9	47
Unique clean sequences	700	592	1,292
No. of no hits	303 (72.8)	113 (27.2)	416 (32.2)
Hits to known gene	397 (45.3)	479 (54.7)	876 (67.8)
Hits to bovine gene	267 (72)	104 (28)	371 (42.3)

Based on the comparison against non-redundant nucleotide database, we were able to assign a putative identity to 876 unique sequences matching the known genes of various organisms and 371 matched sequences bovine gene (Table 2). From the total of the unique clean sequences, 416 showed no match to any nucleotide sequences in dataset.

BlastX homologous searches: Putative functions of the unique sequences were discovered using BlastX to translate each nucleotide query sequence into all reading frames and then searching for matches in the NCBI non-redundant database. BlastX hits that did not meet the following minimum requirements were removed: (*i*) sequence similarity >80% and Evalue <0.00001. Table 3 summarize the statistics of BlastX hits on nr protein database. Based on this comparison we were able to make an estimation of the number of proteins represented by the 1,292 unique sequences (contigs and singletons). Analysis of the positive BlastX on nr protein database revealed 790 hits to known proteins. It indicated that 61.1% of unique sequences contained inserts of known genes. From this, 495 (62.6%) and 295 (37.3%) unique sequences had match to homologous proteins on *res* and *sus* libraries, respectively. Among

them 129 queries exactly matched the sequences of bovine genes on the database (Table 3 S1 and Table 3 S2). From the total of match to known proteins, 129 queries matched known bovine protein and 661 shared high similarities with genes of other species such as human, pig, horse and mouse (data no shown).

Table 3. Summary of BlastX analysis of unique sequences from two skin cDNA libraries on non-redundant protein database >80%.

Libraries	No. of unique sequences	No. of no hits (%)	No. of hits with known protein (%)	No. of hits with protein bovine (%)
res	700	255 (50.8)	495 (62.6)	51 (10.3)
sus	592	247 (49.2)	295 (37.3)	78 (26.5)
Total	1,292	502 (38.9)	790 (61.1%)	129 (16.3)

The hits to known protein were clusterized into more abundant categories classes based on BlastX blasts results with associated predicted or known functions. The partial clusterization results are shown on Table 4 (*res*) and Table 5 (*sus*). Assignment of homologous proteins by BlastX revealed 60 sequences with hits to hypothetical proteins on *res* (n=50) and *sus* (n=10) libraries. The most frequently found genes were ribosomals proteins with 35 and 55 match to *res* and *sus* libraries, respectively, followed by binding proteins 22 and 14; structure cells proteins like keratin and collagen showed 13 and 12 hits. Others sequences showed hits with immune response genes. On the *res* (n=3) library, homologous proteins showed matchs to interferon regulatory factor 3 (Contig 2), cathepsin L2 precursor (Bs_Res_03G10_rc), MHC class antigen I (Contig 200). From the *sus* (n=8) dataset, TIMP metallopeptidase inhibitor 2 (Bs_Sus_26A07), CD44 antigen (Contig 7), CD63 antigen (Bs_Sus_18E06) and neutrophil beta-defensin 12 (Bs_Sus_19A09).

Table 4. Partial list of putative proteins clusterized by more abundant class on res library from BlastX analysis (nr).

Clone ID	Clone Length (pb)	Evalue	Frame	Acession	Protein Match
Hypothetical protein					
Contig39	639	6,00E-57	3	XP_537826.2	PREDICTED: hypothetical protein
Contig95	527	3,00E-26	1	NP_519204.1	hypothetical protein RSc1083
Contig110	663	2,00E-07	2	NP_287959.1	hypothetical protein Z6021
Contig115	475	8,00E-25	-2	YP_257828.1	hypothetical protein PFL_0686
Contig130	576	4,00E-25	1	YP_274711.1	hypothetical protein PSPPH_2515
Contig136	578	5,00E-16	-1	NP_888639.1	hypothetical protein BB2096
Contig154	592	2,00E-89	2	XP_876960.1	PREDICTED: hypothetical protein
Contig156	472	7,00E-55	-2	NP_249572.1	hypothetical protein PA0881
Contig160	638	1,00E-113	-3	YP_258417.1	hypothetical protein PFL_1288
Contig167	602	2,00E-64	-1	YP_234884.1	hypothetical protein Psyr_1799
Contig179	355	2,00E-06	3	NP_973292.1	hypothetical protein TDE2694
Contig182	568	1,00E-17	-2	YP_335576.1	hypothetical protein BURPS1710b_A0417
Contig187	847	5,00E-57	1	YP_262488.1	hypothetical protein PFL_5420
Contig194	582	2,00E-08	-2	YP_276105.1	hypothetical protein PSPPH_3970
Bs_Res_03D03_rc	470	6,00E-27	3	YP_373045.1	hypothetical protein Bcep18194_B2290
Bs_Res_05H06_rc	427	4,00E-57	2	NP_001094725.1	hypothetical protein LOC616332
Bs_Res_06A07_rc	437	6,00E-15	-1	XP_598312.2	PREDICTED: hypothetical protein
Bs_Res_06B09_rc	190	6,00E-14	3	YP_001019219.1	hypothetical protein Mpe_A0022
Bs_Res_00B11	634	2,00E-50	1	NP_780633.1	hypothetical protein LOC109314
Bs_Res_01F08	448	2,00E-15	2	embCAD98105.1	hypothetical protein
Bs_Res_04B11	585	2,00E-80	1	NP_001074179.1	hypothetical protein LOC242286
Bs_Res_04E11	661	2,00E-87	1	XP_544741.2	PREDICTED: hypothetical protein
Bs_Res_87D06	657	1,00E-115	-1	YP_258382.1	hypothetical protein PFL_1251

Table 4. Partial list of putative proteins clusterized by more abundant class on res library from BlastX analysis (nr). Cont.

Clone ID	Clone Length (pb)	Evalue	Frame	Acession	No. Protein Match
Hypothetical protein					
Bs_Res_88B02	609	1,00E-45	1	YP_257297.1	hypothetical protein PFL_0150
Bs_Res_88H05	268	6,00E-23	3	YP_159140.1	hypothetical protein ebA3751
Bs_Res_89G09	492	3,00E-69	3	YP_261732.1	hypothetical protein PFL_4648
Bs_Res_90A12	422	5,00E-66	2	YP_262194.1	hypothetical protein PFL_5115
Bs_Res_91D06	161	8,00E-11	3	NP_886811.1	hypothetical protein BB0262
Bs_Res_92B11	538	2,00E-61	1	YP_261357.1	hypothetical protein PFL_4265
Bs_Res_92F03	538	1,00E-51	-3	YP_261753.1	hypothetical protein PFL_4670
Bs_Res_93A12	296	9,00E-21	-3	YP_772603.1	hypothetical protein Bamb_0710
Bs_Res_93B05	429	3,00E-16	-1	YP_257828.1	hypothetical protein PFL_0686
Bs_Res_93E05	524	1,00E-18	-3	YP_260889.1	hypothetical protein PFL_3788
Bs_Res_93E07	342	3,00E-13	1	pirB34087	hypothetical protein (L1H 3'
Bs_Res_93F05	380	3,00E-12	1	NP_879266.1	hypothetical protein BP0397
Bs_Res_94A03	446	5,00E-70	3	YP_345931.1	hypothetical protein PflO1_0198
Bs_Res_94C09	358	6,00E-19	2	YP_361714.1	hypothetical protein XCVd0155
Bs_Res_94H03	362	7,00E-31	2	YP_350622.1	hypothetical protein PflO1_4894
Bs_Res_95D07	574	1,00E-19	2	NP_644612.1	hypothetical protein XAC4318
Bs_Res_95E04	582	5,00E-44	3	YP_234385.1	hypothetical protein Psyr_1296
Bs_Res_95G09	540	2,00E-06	-3	YP_001269054.1	hypothetical protein Pput_3746
Bs_Res_95H01	433	6,00E-22	1	NP_774381.1	hypothetical protein blr7741
Bs_Res_95H04	635	4,00E-71	-3	YP_262216.1	hypothetical protein PFL_5137
Bs_Res_96B01	321	4,00E-48	1	YP_346328.1	hypothetical protein PflO1_0595
Bs_Res_96E07	365	3,00E-56	-2	YP_258417.1	hypothetical protein PFL_1288
Bs_Res_96G10	519	2,00E-62	-2	gbEAZ57512.1	conserved hypothetical protein

Table 4. Partial list of putative proteins clusterized by more abundant class on res library from BlastX analysis (nr). Cont.

Clone ID	Clone Length (pb)	Evalue	Frame	Acession	Protein Match
Hypothetical protein					
Bs_Res_96H08	176	2,00E-14	-2	NP_881713.1	hypothetical protein BP3148
Bs_Res_96H12	586	5,00E-56	-3	YP_350028.1	hypothetical protein PflO1_4300
Bs_Res_97E06	304	6,00E-27	-2	NP_743702.1	hypothetical protein PP_1545
Bs_Res_97E07	444	9,00E-10	2	YP_350173.1	hypothetical protein PflO1_4445
Ribosomal protein					
Contig4	487	4,00E-73	-2	gbAAX29364.1	ribosomal protein L27
Contig6	543	1,00E-79	-3	NP_035426.1	ribosomal protein S18
Contig8	661	1,00E-110	-3	NP_001070466.1	ribosomal protein L13a
Contig22	432	2,00E-66	-3	NP_000984.1	ribosomal protein L31 isoform
Contig34	937	1,00E-149	-2	NP_000997.1	ribosomal protein S3a
Contig37	342	8,00E-28	-2	gbAAX36170.1	ribosomal protein S29
Contig41	491	9,00E-62	-2	gbAAX29203.1	ribosomal protein S20
Contig57	574	9,00E-59	-3	NP_000994.1	ribosomal protein P1 isoform isoform
Contig80	455	5,00E-62	1	gbAAX43757.1	ribosomal protein S26
Contig96	486	2,00E-78	3	NP_001019712.1	ribosomal protein S15
Contig141	626	1,00E-101	1	gbAAX43400.1	ribosomal protein S5
Contig151	500	7,00E-58	3	NP_001026926.1	ribosomal protein L6
Contig158	248	3,00E-12	-1	XP_534560.2	similar 60S 60S
Contig176	345	1,00E-23	3	gbAAX43773.1	ribosomal protein L39
Contig185	798	1,00E-125	-3	NP_001038008.1	ribosomal protein L10
Contig188	396	2,00E-44	2	XP_590901.1	similar to yeast yeast
Contig196	536	5,00E-81	3	gbAAX43292.1	ribosomal protein S14
Contig212	545	3,00E-39	-1	NP_001076869.1	mitochondrial ribosomal protein

Table 4. Partial list of putative proteins clusterized by more abundant class on res library from BlastX analysis (nr). Cont.

Clone ID	Clone Length (pb)	Evalue	Frame	Acession	Protein Match
Ribosomal protein					
Bs_Res_03C04_rc	450	5,00E-59	-2	NP_001020492.1	ribosomal protein L35a
Bs_Res_00A02	256	7,00E-18	1	XP_537399.1	similar to 40S 40S
Bs_Res_00C06	438	2,00E-51	-1	NP_001005084.1	ribosomal protein S25
Bs_Res_01E11	259	9,00E-26	2	gbAAS55896.1	40S ribosomal protein S28 S28
Bs_Res_01F02	333	2,00E-49	1	gbAAX29338.1	ribosomal protein L8
Bs_Res_04F02	603	2,00E-90	1	XP_537583.2	similar to 40S
Bs_Res_87C06	653	1,00E-118	2	NP_777140.1	ribosomal protein L3
Bs_Res_87E02	300	4,00E-32	1	gbAAX43793.1	ribosomal protein L38
Bs_Res_87F03	291	2,00E-14	2	XP_922073.1	similar to ribosomal protein ribosomal
Bs_Res_88B04	429	1,00E-61	1	NP_001029667.1	ribosomal protein L35
Bs_Res_89C05	527	7,00E-83	2	NP_001003.1	ribosomal protein S8
Bs_Res_89F07	662	1,00E-106	1	gbAAX29348.1	ribosomal protein S9
Bs_Res_92A02	562	4,00E-97	2	XP_231785.2	similar to 60S
Bs_Res_95B12	423	3,00E-60	-2	gbAAX37110.1	ribosomal protein L34
Bs_Res_95D12	401	5,00E-46	-2	dbjBAE75962.1	ribosomal protein L7
Bs_Res_95H03	326	2,00E-32	-1	gbAAH03518.1	Similar to ribosomal protein
Bs_Res_98D12	577	6,00E-98	3	gbAAX29834.1	ribosomal protein L11
Binding protein					
Contig12	482	4,00E-52	-1	NP_777021.1	S100 calcium binding protein
Contig13	561	1,00E-53	-2	NP_777020.1	S100 calcium-binding protein A4
Contig30	461	4,00E-15	3	gbAAV38972.1	insulin-like growth factor binding
Contig71	520	3,00E-45	1	NP_777076.1	S100 calcium binding protein
Contig89	554	4,00E-74	-1	NP_001008670.1	cellular retinoic acid binding

Table 4. Partial list of putative proteins clusterized by more abundant class on res library from BlastX analysis (nr). Cont.

Clone ID	Clone Length (pb)	Evalue	Frame	Acession	No. Protein Match
Binding protein					
Contig119	636	2,00E-70	3	YP_257343.1	cyclic nucleotide-binding protein
Contig131	595	1,00E-08	2	NP_881653.1	putative transport protein ATP-binding
Contig152	347	1,00E-20	-3	YP_348034.1	periplasmic binding protein/LacI
Contig166	503	1,00E-43	1	XP_581277.2	similar to odorant
Contig172	391	2,00E-46	2	NP_001107197.1	S100 calcium binding protein
Contig190	441	6,00E-06	-1	gbAAP94989.1	salivary androgen-binding protein beta
Bs_Res_03F04_rc	439	3,00E-09	-3	gbAAT85299.1	glycine-rich RNA-binding protein,
Bs_Res_01A09	356	3,00E-40	1	NP_071528.1	histidine triad nucleotide binding
Bs_Res_01F06	545	7,00E-73	2	NP_776740.1	fatty acid binding protein
Bs_Res_87C09	657	1,00E-114	1	NP_886542.1	putative extracellular solute-binding
Bs_Res_89D02	607	1,00E-106	3	NP_001032906.1	RAN binding protein 6
Bs_Res_90D12	311	4,00E-10	1	YP_523095.1	extracellular ligand-binding receptor
Bs_Res_92H06	217	1,00E-26	-3	YP_258203.1	GTP-binding protein Era
Bs_Res_93A01	590	1,00E-43	2	YP_236842.1	GAF:ATP-binding region,
Bs_Res_93B12	490	2,00E-53	-1	NP_776739.1	fatty acid binding protein
Bs_Res_95B06	247	1,00E-11	-1	YP_261563.1	DNA-binding response regulator
Bs_Res_96B02	467	5,00E-38	2	NP_252362.1	ABC transporter ATP-binding protein
Keratin/collagen					
Contig18	977	1,00E-93	-1	NP_001008663.1	keratin 5 (epidermolysis bullosa
Contig47	1065	1,00E-108	2	NP_776802.1	keratin 10
Contig59	613	4,00E-65	2	NP_001003392.1	keratin 1
Contig65	882	1,00E-104	2	gbAAP36646.1	Homo sapiens keratin 14
Contig124	628	1,00E-102	3	XP_543647.2	similar to keratin

Table 4. Partial list of putative proteins clusterized by more abundant class on res library from BlastX analysis (nr). Cont.

Clone ID	Clone Length (pb)	Evalue	Frame	Acession	Protein Match
Keratin/collagen					
Contig216	615	5,00E-66	1	spP04264K2C1	Keratin, type II cytoskeletal
Contig225	498	1,00E-16	2	spO77727K1C15	Keratin, type I cytoskeletal
Bs_Res_02B12	646	8,00E-91	3	XP_523648.1	keratin 35
Bs_Res_89G01	581	2,00E-44	2	NP_000414.2	keratin 2
Bs_Res_94A01	315	1,00E-34	3	XP_548101.2	similar to Keratin,
Contig11	1113	1,00E-160	1	NP_001029211.1	collagen, type I, alpha
Contig67	1376	0.0	3	NP_000081.1	collagen type III alpha
Bs_Res_05G03_rc	346	2,00E-23	-3	NP_001029211.1	collagen, type I, alpha
Unmamed protein/unknown					
Contig168	479	5,00E-66	1	dbjBAC25377.1	unnamed protein product
Contig217	317	4,00E-07	-1	dbjBAB29079.1	unnamed protein product
Bs_Res_01H02	209	4,00E-19	1	dbjBAE38285.1	unnamed protein product
Bs_Res_89H06	627	5,00E-15	-2	dbjBAC30170.1	unnamed protein product
Contig28	461	7,00E-37	2	gbAAC25388.1	unknown
Contig140	644	4,00E-37	2	gbAAC09350.1	unknown
Bs_Res_88A01	331	1,00E-06	3	gbAAG22475.1	unknown
Bs_Res_89G07	433	5,00E-13	-2	gbAAK16226.1	unknown
Immune response					
Contig2	327	6,00E-20	-3	gbAAZ38325.1	interferon regulatory factor 3
Bs_Res_03G10_rc	595	1,00E-100	-2	spQ5E998CATL2	Cathepsin L2 precursor
Contig200	1136	1,00E-144	-2	gbAAO91983.1	MHC class I antigen

Table 5. Partial list of putative proteins clusterized by more abundant class on sus library from BlastX analysis (nr).

Clone ID	Clone Length	Evalue	Frame	Acession No.	Protein Match
ribosomal protein n=55					
Contig1	554	3,00E-78	-2	NP_001019712.1	ribosomal protein S15
Contig5	585	9,00E-59	-1	NP_000994.1	ribosomal protein P1 isoform 1
Contig12	607	3,00E-48	3	NP_001020492.1	ribosomal protein L35a
Contig35	581	3,00E-84	-3	NP_001019642.1	ribosomal protein L27a
Contig37	209	2,00E-19	3	NP_001026926.1	ribosomal protein L6
Contig43	554	5,00E-87	3	gbAAX29372.1	ribosomal protein S11
Contig55	488	3,00E-36	2	gbAAH17386.1AAH17386	ribosomal protein S19
Contig58	339	1,00E-07	2	NP_001070466.1	ribosomal protein L13a
Contig62	316	1,00E-30	-2	gbAAX43326.1	ribosomal protein S13
Contig65	687	5,00E-81	2	NP_001014862.1	ribosomal protein L29
Contig79	463	8,00E-25	-1	gbAAS55896.1	40S ribosomal protein S28
Contig82	752	1,00E-104	1	gbAAI02075.1	Ribosomal protein, large, P0
Contig86	677	2,00E-99	-2	NP_001030383.1	ribosomal protein L5
Contig89	324	2,00E-36	-3	gbAAX29203.1	ribosomal protein S20
Contig90	259	5,00E-17	-1	spP51417RL15_CHICK	60S ribosomal protein L15 (L10)
Contig96	713	1,00E-106	1	NP_001038008.1	ribosomal protein L10
Contig99	583	5,00E-71	2	gbAAX36164.1	ribosomal protein L32
Contig101	523	1,00E-59	-2	gbAAX43301.1	ribosomal protein L30
Contig102	322	3,00E-32	2	XP_542250.2	PREDICTED: similar to ribosomal
Contig105	662	2,00 E-100	3	NP_000998.1	ribosomal protein S4, X-linked X
Contig107	562	2,00E-75	-3	NP_001070466.1	ribosomal protein L13a
Contig115	606	2,00E-82	1	XP_868296.1	PREDICTED: similar to ribosomal
Contig117	324	1,00E-27	1	gbAAH59443.1	Ribosomal protein S5
Contig127	467	4,00E-31	1	gbAAH03518.1AAH03518	Similar to ribosomal protein L23

Table 5. Partial list of putative proteins clusterized by more abundant class on sus library from BlastX analysis (nr) (Cont.)

Clone ID	Clone Length	Evalue	Frame	Acession No.	Protein Match
binding protein n=55					
Contig134	529	8,00E-58	1	XP_231785.2	PREDICTED: similar to 60S ribosomal
Contig140	355	1,00E-28	3	gbAAX36170.1	ribosomal protein S29
Contig145	453	8,00E-64	2	NP_000984.1	ribosomal protein L31 isoform 1
Contig147	212	2,00E-09	1	XP_510169.1	PREDICTED: similar to ribosomal
Contig148	549	4,00E-82	-1	XP_533657.2	PREDICTED: similar to ribosomal
Contig152	528	2,00E-67	3	spP26452.4RSSA_BOVIN	40S ribosomal protein SA
Contig153	263	6,00E-22	1	gbAAW82138.1	ubiquitin-like/S30 ribosomal fusion
Contig156	595	5,00E-82	-1	NP_035426.1	ribosomal protein S18
Contig157	704	1,00E-112	3	NP_001070466.1	ribosomal protein L13a
Contig161	513	1,00E-62	-3	XP_520251.1	PREDICTED: similar to ribosomal
Contig164	742	1,00E-122	3	NP_001035610.1	ribosomal protein L7a
Contig173	434	4,00E-32	3	gbAAX43793.1	ribosomal protein L38
Contig179	510	3,00E-57	3	NP_001026926.1	ribosomal protein L6
Contig181	734	1,00E-103	-2	gbAAX29111.1	ribosomal protein S7
Contig186	737	1,00E-98	-2	NP_001019640.2	ribosomal protein L9
Contig188	660	1,00E-77	2	NP_001016.1	ribosomal protein S23
Contig196	516	1,00E-70	-2	NP_001007.2	ribosomal protein S12
Contig207	463	6,00E-59	1	gbAAK95209.1AF402835_1	40S ribosomal protein S26-2
Contig214	247	2,00E-14	-3	XP_208281.1	PREDICTED: similar to ribosomal
Contig218	523	3,00E-57	3	gbAAK31162.1	ubiquitin A-52 residue ribosomal protein
Bs_Sus_10H01	288	2,00E-36	-3	NP_777213.1	ribosomal protein P2
Bs_Sus_13E04	525	1,00E-66	3	gbAAX29348.1	ribosomal protein S9
Bs_Sus_16F12	456	3,00E-73	3	NP_073163.1	ribosomal protein S14
Bs_Sus_18F09	291	6,00E-27	2	NP_001030580.1	ribosomal protein L28

Table 5. Partial list of putative proteins clusterized by more abundant class on sus library from BlastX analysis (nr) (Cont.)

Clone ID	Clone Length	Evalue	Frame	Acession No.	Protein Match
binding protein n=55					
Bs_Sus_21E04	274	8,00E-22	2	XP_863240.1	PREDICTED: similar to ribosomal
Bs_Sus_23B09	441	2,00E-51	1	NP_062621.2	ribosomal protein L21
Bs_Sus_23H03	458	3,00E-60	1	gbAAI02075.1	Ribosomal protein, large, P0
Bs_Sus_26F05	547	5,00E-68	2	NP_075238.1	ribosomal protein L14
Bs_Sus_29H12	469	2,00E-52	3	gbAAX29866.1	ribosomal protein S12
Bs_Sus_30B03	426	2,00E-54	2	XP_851596.1	PREDICTED: similar to ribosomal
Bs_Sus_30D07	435	3,00E-60	3	NP_001029846.1	ribosomal protein L14
binding protein n=14					
Contig24	443	2,00E-46	-3	NP_001107197.1	S100 calcium binding protein A8
Contig67	340	3,00E-20	-2	NP_777076.1	S100 calcium binding protein A12
Contig69	821	3,00E-97	3	XP_581277.2	PREDICTED: similar to odorant binding
Contig113	860	6,00E-38	-3	XP_581277.2	PREDICTED: similar to odorant binding
Contig171	513	6,00E-49	3	NP_001029539.1	S100 calcium binding protein A2
Contig198	432	9,00E-53	1	NP_777020.1	S100 calcium-binding protein A4
Bs_Sus_09A01	694	1,00E-115	2	NP_776993.1	poly(A) binding protein, cytoplasmic 1
Bs_Sus_13C04	628	4,00E-62	1	NP_001095828.1	pirin (iron-binding nuclear
Bs_Sus_13G05	508	2,00E-09	3	NP_001093934.1	androgen binding protein beta
Bs_Sus_16A05	424	2,00E-54	2	XP_857871.1	PREDICTED: similar to Poly(rC)-binding
Bs_Sus_22E02	491	1,00E-14	3	gbAAP44465.1	salivary androgen-binding protein beta
Bs_Sus_24D03	234	5,00E-09	3	gbAAB28336.1	retinol-binding protein; RBP
Bs_Sus_26A11	500	2,00E-77	2	NP_001106714.1	coronin, actin binding protein, 1B
Bs_Sus_26E09	290	2,00E-21	1	gbAAP97261.1AF136171_1	heparin-binding protein HBp15
keratin/collagen n=12					
Contig44	441	2,00E-50	2	XP_580946.1	PREDICTED: similar to keratinocyte

Table 5. Partial list of putative proteins clusterized by more abundant class on sus library from BlastX analysis (nr) (Cont.)

Clone ID	Clone Length	Evalue	Frame	Acession No.	Protein Match
keratin/collagen n=12					
Contig103	384	1,00E-14	1	XP_580946.1	PREDICTED: similar to keratinocyte
Contig138	784	4,00E-23	-3	gbAAD05191.1	type II keratin subunit protein
Contig162	345	6,00E-19	-1	spA1L595.1K1C17_BOVIN	Keratin, type I cytoskeletal
Contig197	409	7,00E-34	3	gbAAR89460.1	keratin-associated protein 16.3
Contig201	652	1,00E-38	-3	NP_776802.1	keratin 10
Contig203	589	1,00E-08	2	spP04261K2C3_BOVIN	Keratin, type II cytoskeletal 60
Contig222	572	1,00E-77	2	NP_776802.1	keratin 10
Bs_Sus_11H05	502	1,00E-66	1	XP_548101.2	PREDICTED: similar to Keratin, type I
Bs_Sus_19A03	661	4,00E-35	2	NP_001032911.1	keratin associated protein 5-5
Bs_Sus_27B12	419	4,00E-50	1	NP_001003392.1	keratin 1
Bs_Sus_29B10	423	7,00E-08	1	XP_543641.2	PREDICTED: similar to keratin 6L
hyphotetical protein n=10					
Contig68	661	2,00E-41	3	XP_875921.1	PREDICTED: hypothetical protein
Contig74	591	5,00E-12	3	XP_601599.2	PREDICTED: hypothetical protein
Contig136	205	3,00E-08	2	NP_001016640.1	hypothetical protein LOC549394
Contig190	338	3,00E-07	3	gbAAO52805.1	hypothetical protein
Contig195	675	2,00E-12	2	XP_660558.1	hypothetical protein AN2954.2
Bs_Sus_14B12	608	2,00E-41	3	XP_414732.1	PREDICTED: hypothetical protein
Bs_Sus_23B02	240	2,00E-32	1	XP_609413.1	PREDICTED: hypothetical protein
Bs_Sus_23F10	510	5,00E-61	1	embCAD97677.1	hypothetical protein
Bs_Sus_24E07	546	6,00E-90	3	XP_876960.1	PREDICTED: hypothetical protein
Bs_Sus_31B01	446	3,00E-09	2	dbjBAB64457.1	hypothetical protein
Immune reponse n=8					
Bs_Sus_25D08	688	1,00E-88	2	NP_777056.1	ADAM metallopeptidase

Table 5. Partial list of putative proteins clusterized by more abundant class on sus library from BlastX analysis (nr) Cont.

Clone ID	Clone Length	Evalue	Frame	Acession No.	Protein Match
Immune reponse n=8					
Bs_Sus_26A07	489	1,00E-61	3	NP_776897.2	TIMP metallopeptidase inhibitor 2
Contig7	917	1,00E-131	-1	NP_776438.1	CD44 antigen
Bs_Sus_09H02	507	5,00E-43	2	NP_777073.1	male-enhanced antigen
Bs_Sus_18E06	651	1,00E-98	3	spQ9XSK2CD63_BOVIN	CD63 antigen
Contig128	433	2,00E-33	3	XP_873295.1	PREDICTED: similar to beta-defensin 2
Bs_Sus_19A09	482	5,00E-30	1	gbAAD43032.1	neutrophil beta-defensin 12
Bs_Sus_28B03	515	5,00E-51	2	embCAA70406.1	MHC class I molecule

Significant hits on the nr database were followed up with protein function searches in the Swissprot database, which provides value-added information reports for protein functions. Roughly, we found 300 (23.2%) hits to homologous proteins with similarity > 80%, whereas 992 (76.9%) have no putative identification or showed similarity < 80% and 121 (40.3%) of the hits represent bovine proteins. The Table 6 summarizes BlastX analysis to putative proteins against Swissprot database.

Table 6: Statistic summary of BlastX vs Uniprot-Swissprot protein database >80% (%)

Libraries	No. of unique sequences	No. of no hits	No. of hits with known protein	No. of hits with known bovine	No. of hits bovine full proteins
res	700	561 (56.6)	139 (46.3)	52 (43.7)	46 (45.1)
sus	592	431 (43.4)	161(53.7)	69 (56.3)	56 (54.9)
Total	1,292	992 (76.9)	300(23.2)	121 (40.3)	102 (84.3)

One particularly important criterion to determine the quality of a cDNA library is the percentage of the clones with full length sequences. So, we analyzed the similarity of these full ESTs with match to bovine homologous proteins. The range of insert sizes (150 to 1,600 bp) suggested that the libraries contained some fraction of full-length cDNAs. The Figure 1 shows the distribution of the number of sequences by length to the BlastX hits against Swissprot protein database.

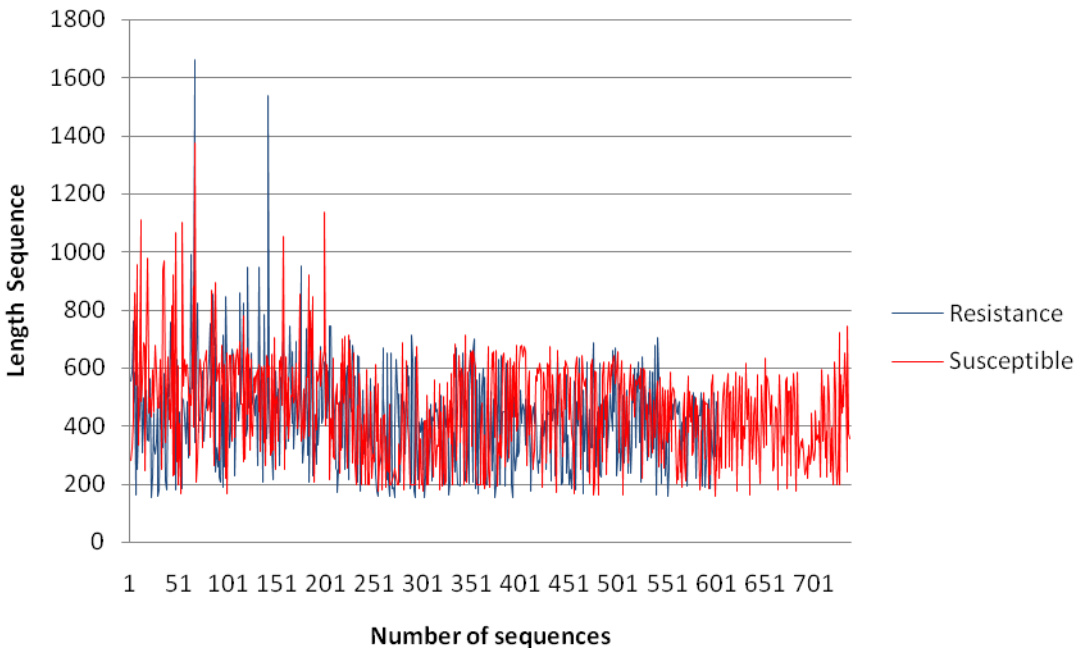


Figure 1. Distribution of the number of sequences by length to the BlastX hits against Swissprot protein database to resistance and susceptible datasets.

From the total of 300 hits to homologous proteins 102 was annotated on Swissprot database as full sequences proteins. The *res* dataset had contributed with 45.1 % total of number of hits and sus with 54.9%.

Gene Ontology analysis and functional profile

The automatic procedure by Blast2GO assigned GO terms and the sequences were categorized in terms of their gene ontology terms at the level 3-rd annotation Al-Shahrour et al. (2004). Subsets of the unique sequences were annotated with the GO terms. The GO database was used to identify the biological process (P), cellular process (C) and molecular function (F) of the putative homologous proteins found on the cDNA libraries. Figure 2 shows the GO terms distribution. The two most representative ontology's terms were biological process and molecular function to both libraries. We annotated 1,175 successfully GO terms at a mean level 621 from to *res* datasets (Figure 2.A) and 554 *sus* datasets (Figure 2.B).

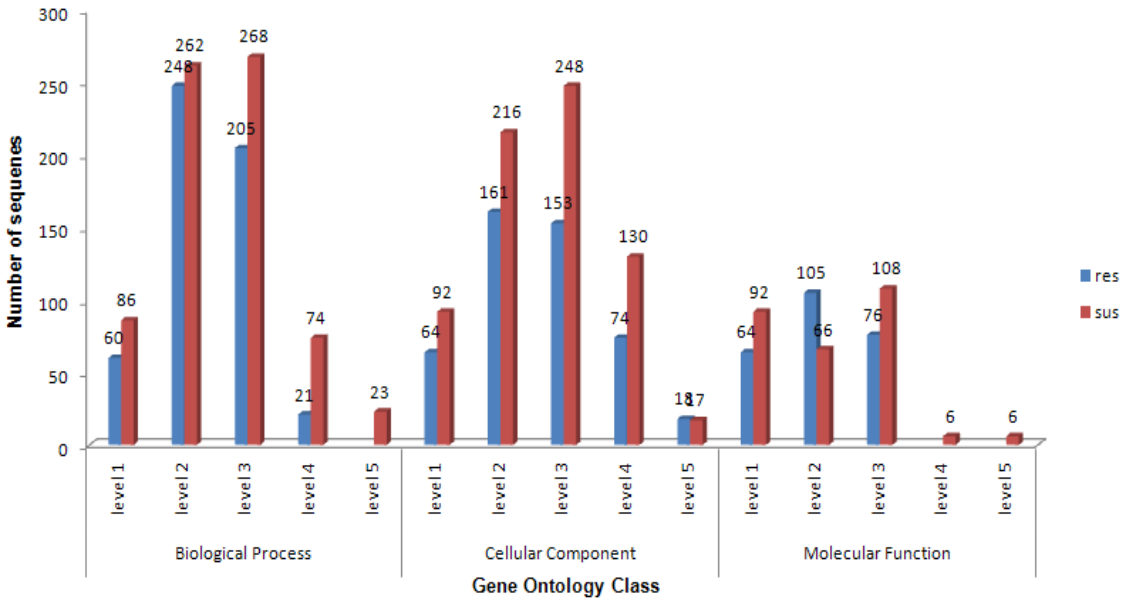
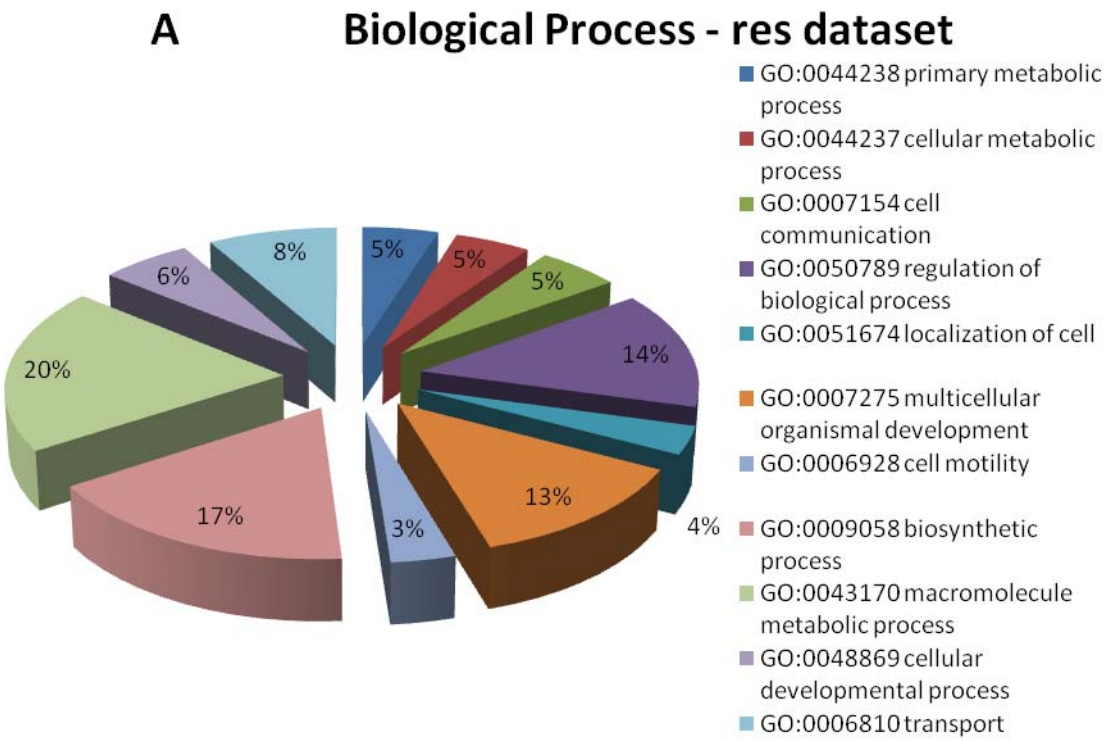


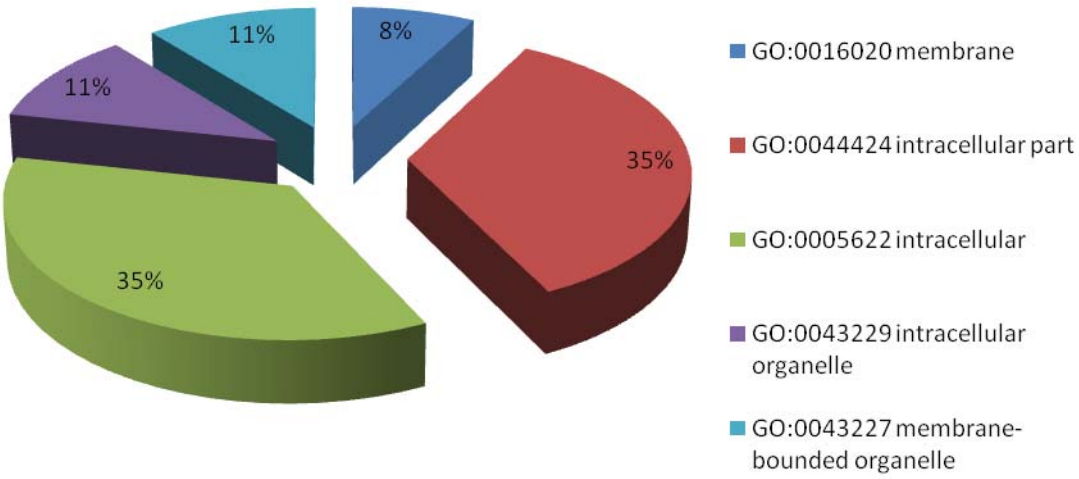
Figure 2. GO level distribution bar chart for skin cDNA libraries.
GO terms allowed to assign 621 annotations to *res* (A) library, mean level = 4.47, standard deviation = 1.372 and 554 annotations to *sus* (B) library, mean level = 4.13, standard deviation = 1.302.

We examined the assigned biological process to determine differences in the distribution of these processes among the libraries. From the total of 60 sequences mapped on biological process, two major of annotations on *res* library were involved in macromolecule metabolic process (n=41), followed by biosynthetic process (n=35) (Figure 3.A). In the *sus* library, others unique sequences were mapped, respectively, to macromolecule metabolic process (n=64) and biosynthesis process (n=56) from total of 86 unique sequences on this dataset (Figure 3.D).

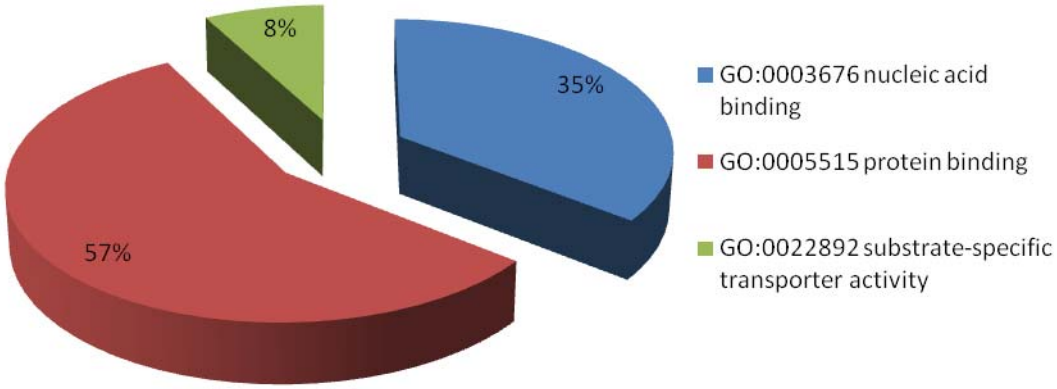
A


Biological Process - res dataset
5%
5%
5%
14%
4%
13%
3%
17%
20%
6%
8%
GO:0044238 primary metabolic process
GO:0044237 cellular metabolic process
GO:0007154 cell communication
GO:0050789 regulation of biological process
GO:0051674 localization of cell
GO:0007275 multicellular organismal development
GO:0006928 cell motility
GO:0009058 biosynthetic process
GO:0043170 macromolecule metabolic process
GO:0048869 cellular developmental process
GO:0006810 transport

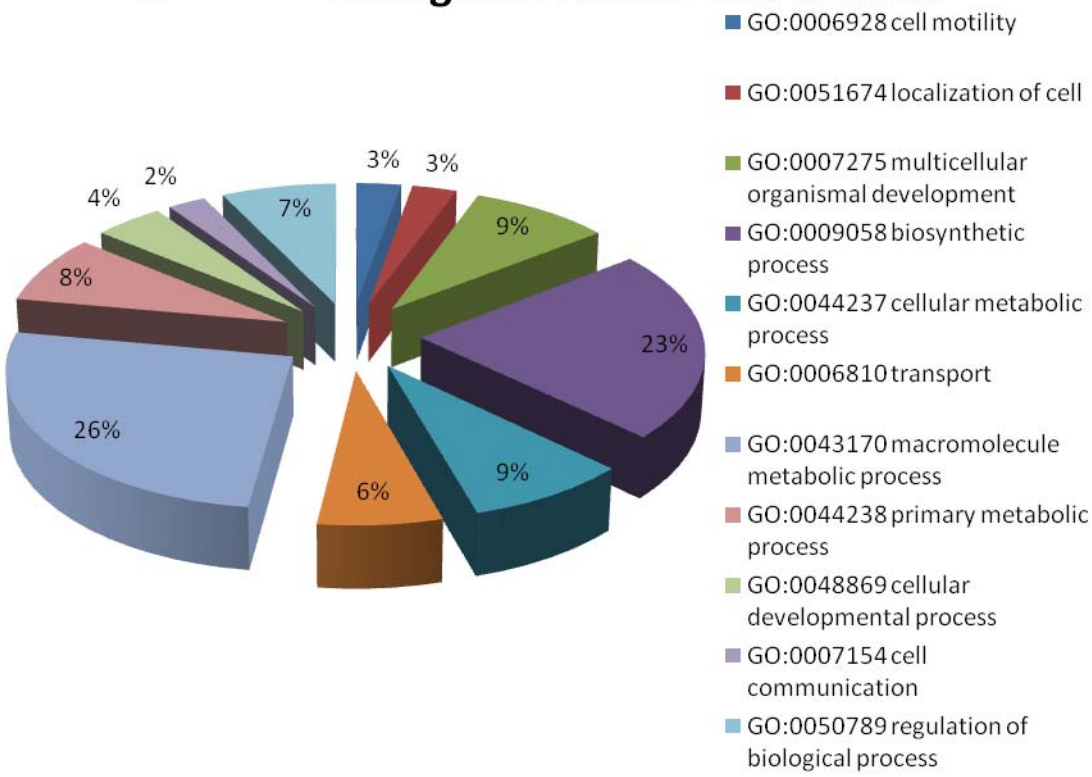
B


Cellular Component - res dataset
8%
35%
35%
11%
11%
GO:0016020 membrane
GO:0044424 intracellular part
GO:0005622 intracellular
GO:0043229 intracellular organelle
GO:0043227 membrane-bounded organelle

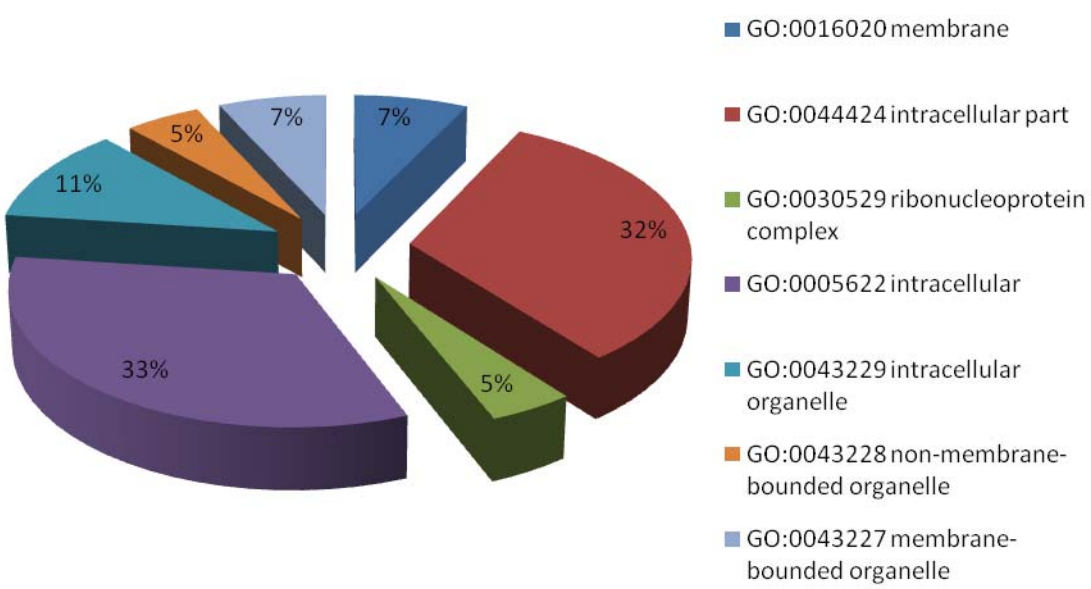
C Molecular Function - res dataset



D Biological Process - sus dataset



E Cellular Component - sus dataset



F Molecular Function - sus dataset

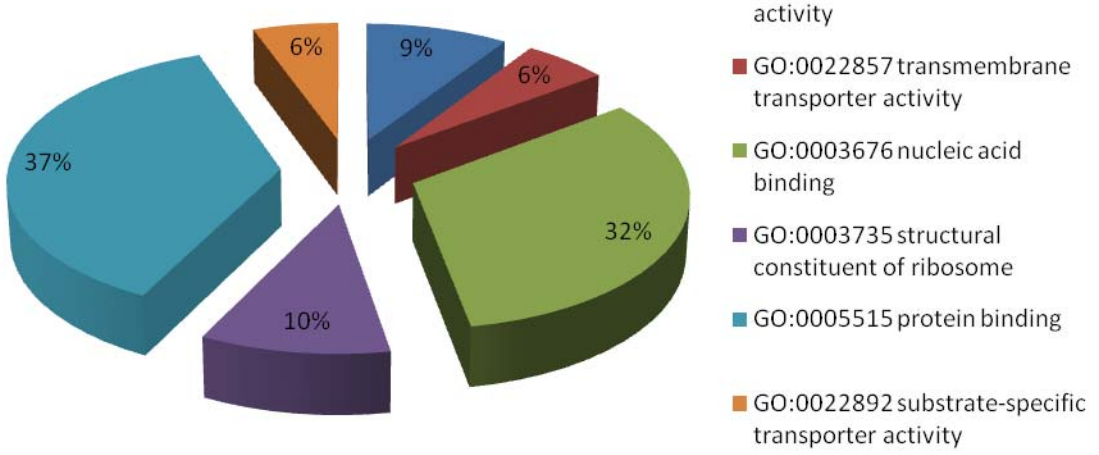


Figure 3. Functional annotation of sequence consensus based on Gene Ontology (GO) categorization.
Sequences were evaluated for their predicted involvement in Biological Processes (A and D), Cellular Components (B and E) and Molecular Functions (C and F), res and sus datasets, respectively. Data are presented at level three GO categorization. Piers chart indicate the relative percentage of sequence consensus in each group.

We also to attribute according to cellular components, distributed GO terms into many: cell, cell part, organelle, extracellular region, extracellular region part and extracellular matrix to library (Figure 3B and 3E). In total, 64 and 92 of unique sequences that have match to GO terms to both libraries. The major categories were intracellular (n=59 and n=82) and intracellular part (n=58 and n=80), respectively, to *res* and *sus* cDNA libraries.

According to the molecular function classification, genes were distributed into the following categories: binding, structural molecule activity, catalytic activity, transport activity, enzyme regulatory activity and transcription regulatory activity, motor activity and molecular transducer activity (Figure 3C and 3F). The functional class protein binding (n=43 and n=40) and nucleic acid binding (n=27 and n=35) were the main categories found on both cDNA libraries.

In this work, the FatiGO tool was used to find protein terms that are over-represented on skin cDNA dataset with respect to a reference *Bos Taurus* genome. To reflect our prior expectation that some functional term related to immune response was enriched, we used the adjusted p-value to control the FDR (*False Discovery Rate*) with a threshold of 5% as the criterion for significance. No significant GO terms was found using the corrected p-values when both libraries were compared (data not were shown). However, 54 protein terms were found to be significant over represented on both libraries in relation to *Bos taurus* genome. These genes belong to processes such as biosynthetic process or oxidative phosphorylation which are not necessarily immune response specific and fall into a wide range of GO categories. The most significant protein terms to *res* and *sus* datasets within the 3 main categories from GO, Swissprot, Interpro and KEGG databases on bovine genome are reported in Table 7.

Table 7. Significant proteins terms over-represented relation to functional profile on *sus* and res datasets compared to genome of Bos taurus p>0.05

Terms of DB	Term Description	Term ID	Adjusted p value	
GO-biological process (n=8)			**SUS library**	**RES library**
level 3	biosynthetic process	GO:0009058	3,172	_
level 3	cellular metabolic process	GO:0044237	0,037	_
level 4	cellular biosynthetic process	GO:0044249	1,562	_
level 5	macromolecule biosynthetic process	GO:0009059	0,087	_
level 5	electron transport	GO:0006118	0,007	_
level 6	translation	GO:0006412	0,000	_
level 6	biopolymer biosynthetic process	GO:0043284	0,029	_
level 7	translational elongation	GO:0006414	0,002	_
GO-cellular component (n=19)				
level 3	non-membrane-bound organelle	GO:0043228	_	0,029
level 6	proton-transporting ATP synthase complex	GO:0045259	_	0,008
level 7	proton-transporting ATP synthase complex, coupling factor F	GO:0045263	_	0,013
level 8	proton-transporting two-sector ATPase complex	GO:0016469	_	0,010
level 3	non-membrane-bound organelle	GO:0043228	0,094	0,029
level 3	organelle part	GO:0044422	0,001	_
level 4	intracellular	GO:0005622	0,004	_
level 5	intracellular part	GO:0044424	0,001	
level 6	cytoplasm	GO:0005737	0,000	0,008
level 6	ribonucleoprotein complex	GO:0030529	0,006	0,010
level 6	intracellular organelle	GO:0043229	0,006	_
level 7	cytoplasmic part	GO:0044444	0,000	0,008
level 7	intracellular non-membrane-bound organelle	GO:0043232	0,300	_
level 7	intracellular organelle part	GO:0044446	0,005	_
level 8	ribosome	GO:0005840	0,000	0,010
level 8	organelle envelope	GO:0031967	0,010	_

Table 7. Significant proteins terms over-represented relation to functional profile on *sus* and res datasets compared to genome of Bos taurus p>0.05. cont.

Terms of DB	Term Description	Term ID	Adjusted p value	
GO-cellular component (n=19)			**SUS librariy**	**RES library**
level 9	large ribosomal subunit	GO:0015934	0,001	_
level 9	organelle inner membrane	GO:0019866	0,001	_
level 9	mitochondrial part	GO:0044429	0,014	_
GO-Molecular function (n=13)				
level 5	hydrolase activity, catalyzing transmembrane movement of substances	GO:0016820	_	0,046
level 3	structural constituent of ribosome	GO:0003735	0,000	0,004
level 3	oxidoreductase activity	GO:0016491	0,002	_
level 4	electron carrier activity	GO:0009055	0,002	_
level 4	heme-copper terminal oxidase activity	GO:0015002	0,002	0,009
level 4	oxidoreductase activity, acting on heme group of donors	GO:0016675	0,002	0,009
level 4	cation transporter activity	GO:0008324	0,015	0,002
level 5	monovalent inorganic cation transporter activity	GO:0015077	0,003	6,199
level 5	oxidoreductase activity, oxygen as acceptor	GO:0016676	0,004	0,031
level 5	NADH dehydrogenase activity	GO:0003954	0,042	_
level 5	oxidoreductase activity, acting on NADH or NADPH, quinone	GO:0016655	0,042	_
level 6	hydrogen ion transporter activity	GO:0015078	0,003	2,400
level 7	cytochrome-c oxidase activity	GO:0004129	0,045	_
KEGG (n=1)	Oxidative phosphorylation	bta00190	0,000*	0,028*
Interpro (n=15)				
Interpro	IPR001751	_	0,000	_
Interpro	IPR013787	_	0,001	_
Interpro	IPR005755	_	0,001	_
Interpro	IPR005822	_	0,001	_
Interpro	IPR001813	__	0,002	___

*ANEXO

Table 7. Significant proteins terms over-represented relation to functional profile on *sus* and res datasets compared to genome of Bos taurus p>0.05. cont.

Terms of DB	Term Description	Term ID	Adjusted p value	
Interpro (n=15)			**SUS librariy**	**RES library**
Interpro	IPR000883	_	0,015	_
Interpro	IPR000915	_	0,028	_
Interpro	IPR001380	_	0,028	_
Interpro	IPR001505	_	0,028	_
Interpro	IPR002429	_	0,028	_
Interpro	IPR005568	_	0,028	_
Interpro	IPR005798	_	0,028	_
Interpro	IPR011759	_	0,028	_
Interpro	IPR000298	_	0,048	_
Interpro	IPR005797	_	0,048	_
Swissprot (n=2)				
Swissprot	COX1_BOVIN	_	0,049	_
Swissprot	RL13A_BOVIN	_	0,049	_

Within the biological process, 8 GO terms were identified for the up-regulated genes in the *sus* group. The most significant GO terms were translational (GO: 0006412, p=0.00013) and translational elongation (GO:0006414, p=0.00162) on level 6 and level 7, respectively. For the cellular component category the number of up-regulated genes were 9 *res* and 19 in *sus* datasets. The best cluster for this category was found for the ribosome sub-category (GO: 0005840, p=0.00026) on level 8 and cytoplasmic part (GO: 0044444, p=0.00002) with p value on level 7. The third GO category, molecular function, included 13 GO terms, which were up-regulated in *res* and *sus* groups. Sub-category oxidoreductase activity (GO: 0016675, p=0.00169) on level 4 was present on both libraries and hydrolase activity (GO: 0016820, p =0.046376) was only represented on *res* dataset (level 5).

Discussion

The identification and sequencing of EST have been useful to functional genomics and contributed to research in microarray construction and interspecies comparison. Previous studies on bovine had placed most attention on its quantitative trait loci. *EST* analysis is an efficient and fast method for gene discovery. In the present study, two cDNA libraries of F_2 cattle skin were constructed and the 2,700 *EST*s were analyzed to identify immune-related genes. These *EST*s were grouped in contigs and singletons. However, because the sequences were 5'-reads the possibility exists that clones from the same gene might be more or less complete at the 5'-end and fail to assemble into a contig. On the other hand, this number can be underestimated do to the fact that genes with high sequence similarities, such as ribosomal, might be grouped together into one contig.

Of the analyzed *EST*s, a high percentage showed no significant similarity to entries in the public protein databases. This was expected in view of the small number of genetic studies in Zebu and/or Holstein x Gyr crossbred animals Thus a significant achievement from our studies is to increase the existing sequence information for this group with 661 new sequences. In this study, 52 (*res*) and 69 (*sus*) transcripts matched with bovine proteins in the Swissprot database. Twenty six transcripts matched no bovine genes in the database but those with a high similarity (> 90%) to proteins of other mammals, like human, are considered newly described bovine genes. The finding that most of these proteins showed high similarities with human is consistent with many previous studies on similarity between human and bovine (Wind et al , 2005).

An reason may explain the fact that only 300 proteins were identified from the 1,292 unique clean sequences: many genes could be novel and therefore there are no orthologs existing in the GenBank and Swissprot. These unidentified clones remain to be characterized in the future. These proteins should be useful for analyzing gene function during tick infestation and for developing molecular markers related to disease-resistance.

The most abundant group of unique sequences in this study belonged to the ribosomals proteins. The high number of unique sequences associated with

housekeeping processes such as metabolism and ribosomal proteins was to be expected, due to the energy production and protein synthesis requirements in both normal and regenerating tissues. This result was expected because ribosomal protein genes are expressed ubiquitously at all development stages. Moreover, the ribosomal protein family is generally well conserved and contains 55 proteins in prokaryotes and 88 in eukaryotes (Doudna and Rath, 2002). An increasing number of studies have reported that numerous ribosomal proteins show extra-ribosomal functions, such as involvement with several human genetic disorders (Wool, 1996). Wang et al. (2008) studied the gene expression profiling on cattle skin and found a huge number of ribosomals proteins up regulated in response to a challenge with ticks in both High Resistance and Low Resistance define animals. It is known that global changes in the expression of ribosomal proteins can occur as a response to stress (Causton et al., 2001). However, changes in the expression of individual ribosomal proteins could be explained if the protein has an additional extra ribosomal function (Wool, 1996).

Binding proteins followed ribosomal proteins in abundance. These included calcium binding protein, lipid binding protein and ATP/GTP binding protein. For example, S100 proteins act as mediators of calcium-dependent signal transduction. The S100A7 gene encodes a low molecular weight calcium-binding protein responsible for the chemotactic migration of CD4+ lymphocytes to the affected region (Jinquan et al., 1996). The extracellular matrix (ECM), or connective tissue, provides the scaffolding in which all tissues are associated. This tissue is organized in a complex array of molecules that consists of collagens, proteoglycans, and glycosaminoglycans. The gene TIMP metallopeptidase inhibitor 2 (Bs_Sus_26A07) encoding extracellular proteinase inhibitor. Matrix metalloproteinases and their inhibitors are responsible for the control of extracellular matrix degradation. Alterations in this enzyme system might be involved in the extracellular matrix alterations associated with glaucoma in humans (Rosa et al., 2000). Hence, the skin ephitelium can control the status of the extracellular matrix through the secretion of metalloproteinases inhibitors. A balance between proteinases and their inhibitors in the aqueous humor can play a role in the processing, targeting and turnover of protein messengers and/or membrane receptors involved in the modulation of the aqueous humor out flow or in other biological functions. During wound healing,

active remodeling of the ECM components is initiated. New matrix molecules are produced by activated fibroblasts, whereas damaged tissue is systematically degraded by native enzymes within the tissue as well as those elaborated by the same activated fibroblasts, macrophages, and other leukocytes.

Other gene, like cysteine proteinase inhibitor (Contig 26), also present on sus library, are natural reversible inhibitors of papain-like cysteine proteinases found in many plants and animals, such as cathepsins B, L, H and S. Cystatins exert numerous specific functions in vertebrates, including the presentation of antigens (Honey and Rudensky, 2003), development of the immune system (Lombardi *et al*., 2005), epidermal homeostasis (Reinheckel *et al*., 2005), extracellular matrix degradation, and neutrophil chemotaxis during inflammation (Serveau-Avesque *et al*., 2006).

Some unique sequences are related to molecules of the immune system. Some of them, such as Contig200 (res) and Bs_Sus_28B03 (sus), are similar to MHC class I, that are not restricted to immune cells. Others, as such the transcription factor interferon regulatory factor 3 (IRF-3) regulates genes in the innate immune response.

CD44 (Contig7) present on library *res*, is a cell surface glycoprotein involved in cell/cell and cell/matrix interactions. The CD44 antigen is a highly glycosylated cell-surface polypeptide involved in diverse cellular functions, including cell adhesion and lymphocyte-homing receptor activity. Expression of cell adhesion molecules regulates epithelial cell differentiation and organization of complex tissues such as skin. CD44 is also expressed in vivo by several tumors, including astrocytomas, meningiomas, and colonic adenocarcinomas. In addition, it has been shown that expression of CD44 appears to confer metastatic potential to cell lines derived from certain adenocarcinomas. In the skin, CD44 is normally expressed in epidermal keratinocytes and hair follicular, sebaceous, and ecrine epithelial cells (Hale et al., 1995).

Contig128 and Bs_Sus_19A09, both from *sus* dataset, were homologous to defensins. . The percentage of sequences associated with inflammatory response in the sus dataset was more to double that of the res library. At this tissue, wound healing is occurring and the immune response must be activated against tick infestation. Defensins are members of an

evolutionarily old family of related peptides (Crovella et al, 2005). Zhu (2008) has studied the evolutionary relationships of defensins as a class of effectors of innate immunity in three eukaryotic kingdoms. In the cow, β-defensins are encoded by a large gene family expressed in a wide variety of tissues. Thirteen β-defensins have been isolated from bovine neutrophils, although gene expression is restricted to mature myelopoietic cells. To date, β-defensin gene expression has not been reported in macrophages, although defensins have been shown to exhibit antimicrobial activity against intracellular pathogens of macrophages (Selsted et al.1993).

The sequences from each library were processed using Blast2GO, but showed very similar compositions when defined in terms of their biological process, cellular components and molecular function (GO annotation level 3). An example of this output is shown for both libraries in Figure 3 with the majority of clones having either catalytic or binding activities. However, FatiGO was used to perform statistical analyses using pair-wise comparisons between libraries to identify any potential functional enrichment. Explicar melhor No GO terms were found statistical significant between the libraries. However, GO terms were found significant to level of genome bovine. For example, translational elongation (GO: 0006414) was elevated only on *sus* library compared to the control genome with single test p-values of 0.01. This was not significant using the False Discovery Rate on *res* library. The problem with such pair wise comparisons is that even though the comparative library was always the control animals, the GO categories listed for each pair wise comparison varied considerably and it was not possible using this technique to make global statements of certain molecular functions being statistically enhanced in *sus* library compared to *res* library.

Conclusion

We successfully constructed two cDNA libraries from skin tissue of F_2 animals. On hundred twenty nine (129) unique sequences could be assigned a protein homologous in bovine. Six hundred sixthy one (661) sequences in the non-redundant clustered dataset showed be new expression tags or do not have proteins homologous in the bovine.

Annotation searches on biologically important putative functions showed that the immune response GO term is sub-represented in the libraries datasets, suggesting the presence of a relatively lower number of immune-related genes.

References

Alowami S, Qing G, Emberley E, Snell L, Watson PR (2003) Psoriasin (S100A7) expression is altered during skin tumorigenesis. *BMC Dermatology*, v.3, p.1-7.

Al-Shahrour F, Diaz-Uriarte R, Dopazo J (2004) FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, v.20, p.578–580.

Birnboim HC, Doly J (1979) A Rapid Alkaline Extraction Procedure for Screening Recombinant Plasmid DNA. *Nucleic Acids Res,* v.7(6)p.1513-1523.

Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander ES, Young RA. (2001) Remodeling of Yeast Genome Expression in *Response to Environmental Changes*. Mol Biol Cell, v.12, p.323–337.

Conesa A, Götz S, García-Gómez J M, Terol J, Talón M, Robles M (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics,* v.21, p.3674-3676.

Crovella S, Antcheva N, Zelezetsky I, Boniotto M, Pacor S, Verga Falzacappa MV, Tossi A (2005). Primate beta-defensins - structure, function and evolution. *Current Protein and Peptide Science*, v.6(1), p.7-21.

Douglas SE, Gallant JW, Bullerwell CE, Wolff C, Munholland J, Reith ME (1999): Winter flounder expressed sequence tags: establishment of an *EST* database and identification of novel fish genes. *Mar Biotechnol*, v.1, p.458-464.

Doudna JA, Rath VL (2002) Structure and function of the eukaryotic ribosome: The next frontier. *Cell,* v.109, p.153-156.

Ewing B (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research,* v.8, p.175-185.

Green P (1999) Documentation for PHRAP and Cross_Match (version 0.990319). Available at http://www.phrap.org/phrap.docs/phrap.html. (verified 5 October 2006)

Gumbiner BM (1993) Breaking through the tight junction barrier. *J. Cell Biol*., v.123, p.1631-1633.

Hale LP, Patel DD, Clark RE, Haynes BF (1995) Distribution of CD44 variant isoforms in human skin: differential expression in components of benign and malignant epithelia. *J Cutan Pathol,* v.22(6), p.536-45.

Honey K, Rudensky AY (2003) Lysosomal cysteine proteases regulate antigen presentation. *Nature Reviews Immunology,* v.3, p.472-482.

Jinquan T, Vorum H, Larsen CG, Madsen P, Rasmussen H, Gesser B, Etzerodt M, Honore B, Celis JE, Thestrup-Pedersen K (1996) Psoriasin: a novel chemotactic protein. *Journal of Investigative Dermatology,* v.107, p.5-10.

Koppen W, Geiger R (1936). *Handbuch der Klimatologie*. v.1. Part C. Verlag, Börntraeger, Berlin, Germany.

Lombardi G, Burzyn D, Mundinano J, Berguer P, Bekinschtein P, et al. (2005) Cathepsin-L influences the expression of extracellular matrix in lymphoid organs and plays a role in the regulation of thymic output and of peripheral T cell number. *Journal of Immunology,* v.174, p.7022-7032.

Martinez ML, Machado MA, Nascimento CS, Silva MVGB, Teodoro RL, Furlong J, Prata MCA, Campos AL, Guimarães MFM, Azevedo ALS, Pires MFA, Verneque RS (2006) Association of BOLA-DRB3.2 alleles with tick (*Boophilus* microplus) resistance in cattle. *Genetics and Molecular Research*, v.5, p.513-524.

Montell C, Birnbaumer L, Flockerzi V The (2002) TRP channels, a remarkably functional family. *Cell,* v.108, p.595–598.

Parkinson J, Anthony A, Wasmuth J, Schmid R, Hedley A, Blaxter M (2004) PartiGene—constructing partial genomes. *Bioinformatics* v.20, p.1398–1404.

Reinheckel T, Hagemann S, Dollwet-Mack S, Martinez E, Lohmuller T, Zlatkovic G, Tobin DJ, Maas-Szabowski N, Peters C. (2005) The lysosomal cysteine protease cathepsin L regulates keratinocyte proliferation by control of growth factor recycling. *Journal of Cell Science,* v.118, p.3387-3395.

Rosa, FA, Lee D A (2000) Collagen degradation in glaucoma: will it gain a therapeutic value? *Current Opinion in Ophthalmology*, v. 11 (2), p.90-93.

Sambrook, J; Russel, D.W. *Molecular cloning. A laboratory manual*. New York : Cold Spring Harbor, 2001 (volumes 1, 2 e 3).

Selsted M E, Tang Y-Q, Morris W L, McGuire P A, Novotny M J, Smith W, Henschen A H, Cullor J S (1993) Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils. *J Biol Chem*., v.268, p.6641–6648.

Serveau-Avesque C, Martino, MF, Herve-Grepinet V, Hazouard E, Gauthier F, Diot E and Lalmanach G (2006) Active cathepsins B, H, K, L and S in human inflammatory bronchoalveolar lavage fluids. *Biology of the Cell,* v.98, p.15-22.

Teodoro R L, Pires MFA, Furlong J, Prata MCA, Machado MA, Silva MVBG, Freitas C, Peixoto MGCD, Martinez ML (2006) Estimativas de parâmetros genéticos e fenotípicos para a resistência a carrapatos e bernes em uma população experimental F_2 Holandês x Gir. In 'Anais da 43th Reunião Anual da Sociedade Brasileira de Zootecnia'. (Ed.SBZ) (CD ROM). (43th Reunião Anual da Sociedade Brasileira de Zootecnia: João Pessoa - PB: Brazil).

Teodoro RL, Madalena FE (2003) Dairy production and reproduction by crosses of Holstein, Jersey or Brown Swiss sires with Holstein-Friesian/Gir dams. Trop. *Anim. Health Prod*., v.35, p.105-115.

Villares JB (1941) Climatologia zootecnica III Contribuição ao estudo de resistência e susceptibilidade genética dos bovinos ao Boophilus microplus. In.: Boletim da Indústria Animal, Nova Serie, São Paulo v.4, p.70-86.

Wang YH, Reverter A, Kemp D, McWilliam SM, Ingham A, Davis C, Moore RJ, Lehnert, SA (2007) Gene expression profiling of Hereford Shorthorn cattle following challenge with Boophilus microplus tick larvae. *Aust. J. Exp. Agric*., v.47, p.1397-1407.

Webb M, Ethan Emberley D, Lizardo M, Alowami S, Qing G, Alfia'ar A, Snell-Curtis LJ, Niu Y, Civetta A, Myal Y, Shiu R, Murphy LC, Watson PH (2005) Expression analysis of the mouse S100A7/psoriasin gene in skin inflammation and mammary tumorigenesis. *BMC Cancer,* v.5, p.17.

Willadsen P, Jongejan F (1999) Immunology of the Tick-Host Interaction and the Control of Ticks and Tick-borne Diseases *Parasitology Today*, v.15(7), p.258-62.

Wind AE, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA (2005) A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution *Proc. Natl. Acad. Sci*. USA December 20, v.102, p.18526-18531

Wool IG (1996) Extraribosomal functions of ribosomal proteins. *Trends Biochem Sci.,* v.21, p.164–165.

Zhu S (2008) Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CSalphabeta defensins. *Molecular Immunology,* v.45(3), p.828-838.

ANEXO 1

Table 2 S1. List of contaminants sequences from cloning vector and Echerichia coli identified to both libraries

sequence name	sequence description	length	min. eValue[1]	sim. mean[2]
res (n=38)				
Contig40	cloning vector pexpress-complete sequence	422	1.4E-25	95,20%
Bs_Res_03D02_rc	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_03E04_rc	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_03F11_rc	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_03G07_rc	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_03G11_rc	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_03H02_rc	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_00A06	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_00E06	cloning vector pexpress-complete sequence	348	4.8E-24	95,80%
Bs_Res_00E09	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_00F03	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_00F10	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_00G05	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_00G08	cloning vector pexpress-complete sequence	196	5.9E-26	99,40%
Bs_Res_01A01	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_01A04	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_01A07	cloning vector pexpress-complete sequence	310	1.0E-25	98,44%
Bs_Res_01A10	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_01A12	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_01B07	cloning vector pexpress-complete sequence	416	1.3E-25	93,20%
Bs_Res_01B08	cloning vector pexpress-complete sequence	419	1.3E-25	94,40%
Bs_Res_01B11	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_01C05	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_01C07	cloning vector pexpress-complete sequence	459	1.5E-25	94,68%
Bs_Res_01G09	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_02B08	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_02H04	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_04G02	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_04G03	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_04G09	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_04G10	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_04G11	cloning vector pexpress-complete sequence	391	5.5E-24	92,90%
Bs_Res_04H03	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_98H03	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_99A03	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_99B06	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Res_88F01	escherichia colicomplete genome	672	1.0E-0.0	74,90%
Bs_Res_88H05	escherichia colicomplete genome	268	6.1E-34	72,40%
sus (n=9)				
Contig59	expression vectorcomplete sequence	481	3.1E-123	100,00%
Contig72	expression vectorcomplete sequence	415	2.3E-35	93,50%
Bs_Sus_10E12	expression vectorcomplete sequence	390	4.8E-50	100,00%
Bs_Sus_18C05	cloning vector pexpress-complete sequence	197	5.9E-26	99,40%
Bs_Sus_20A01	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Sus_20A02	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Sus_20H03	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Sus_21F04	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%
Bs_Sus_24E01	cloning vector pexpress-complete sequence	195	5.8E-26	99,40%

Table 3 S1. Putative protein with significatives hits from bovine homologous protein on nr protein database (BlastX, cutoff 80%, E value < 0.00001) to *res* library, n=51

QueryID	QueryLen	Accession No.	Score	bit	Evalue	Frame	Organism	Protein ID	Description
Contig2	327	71564269	98.6	244	6,00E-20	-3	Bos	AAZ38325.1	interferon regulatory factor 3
Contig5	859	19919708	464	1193	1,00E-129	1	Bos	AAM08332.1	cytochrome oxidase subunit III
Contig7	955	77736201	587	1513	1,00E-166	-1	Bos taurus	NP_001029799.1	gelsolin isoform b
Contig8	661	116000000	399	1026	1,00E-110	-3	Bos	NP_001070466.1	ribosomal protein L13a
Contig11	1113	77404252	567	1462	1,00E-160	1	Bos	NP_001029211.1	collagen, type I, alpha 1
Contig12	482	27807077	205	522	4,00E-52	-1	Bos	NP_777021.1	S100 calcium binding protein A7
Contig13	561	27807075	211	537	1,00E-53	-2	Bos	NP_777020.1	S100 calcium-binding protein A4
Contig14	685	4105170	414	1063	1,00E-114	-1	Bos taurus	AAD02283.1	transcription factor AEBP1
Contig26	679	27806675	282	721	7,00E-75	-3	Bos taurus	NP_776454.1	cystatin C
Contig29	468	23305876	213	541	2,00E-54	-2	Bos taurus	AAN17325.1	ferritin heavy chain
Contig45	923	60101830	403	1035	1,00E-111	3	Bos taurus	YP_209210.1	ATP synthase F0 subunit 6
Contig47	1065	27805977	393	1009	1,00E-108	2	Bos taurus	NP_776802.1	keratin 10
Contig48	276	60101827	96.7	239	2,00E-19	3	Bos	YP_209207.1	cytochrome c oxidase subunit I
Contig63	469	61097917	200	509	1,00E-50	3	Bos taurus	NP_001012782.1	cystatin E/M
Contig96	486	66792868	293	749	2,00E-78	3	Bos taurus	NP_001019712.1	ribosomal protein S15
Contig151	500	72534798	210	534	7,00E-58	3	Bos taurus	NP_001026926.1	ribosomal protein L6
Contig157	1052	4239700	192	487	2,00E-47	-1	Bos taurus	CAA10770.1	reverse transcriptase-like
Contig180	401	28461285	130	328	1,00E-29	-3	Bos	NP_787025.1	cytochrome c oxidase subunit VIIc
Contig199	660	77736201	270	689	4,00E-71	-2	Bos taurus	NP_001029799.1	gelsolin isoform b
Contig200	1136	30313725	514	1323	1,00E-144	-2	Bos indicus	AAO91983.1	MHC class I antigen
Contig201	401	114000000	162	410	2,00E-57	-3	Bos	NP_001040046.1	churchill domain containing 1
Contig204	464	50234067	228	582	2,00E-69	-1	Bos	YP_052697.1	NADH dehydrogenase subunit 1
Contig221	539	78042556	246	627	3,00E-64	3	Bos taurus	NP_001030276.1	ligatin
Bs_Res_03C04_rc	450	70778730	228	581	5,00E-59	-2	Bos taurus	NP_001020492.1	ribosomal protein L35a
Bs_Res_05C02_rc	303	115000000	170	431	2,00E-43	-2	Bos taurus	NP_001069271.1	fumarate hydratase

Table 3 S1. Putative protein with significatives hits from bovine homologous protein on nr protein database (BlastX, cutoff 80%, E value < 0.00001) to *res* library, n=51

QueryID	QueryLen	Accession No.	Score	bit	Evalue	Frame	Organism	Protein ID	Description
Bs_Res_05G03_rc	346	77404252	110	274	2,00E-23	-3	Bos	NP_001029211.1	collagen, type I, alpha 1
Bs_Res_06A07_rc	437	76637791	82.0	201	6,00E-15	-1	Bos	XP_598312.2	PREDICTED: hypothetical protein
Bs_Res_01F06	545	27805805	275	702	7,00E-73	2	Bos	NP_776740.1	fatty acid binding protein 5
Bs_Res_01F10	192	114000000	100	249	2,00E-20	3	Bos taurus	NP_001039446.1	TRM112-like
Bs_Res_02H03	566	71037405	339	869	3,00E-92	1	Bos	NP_001020740.1	heat shock 27kDa protein 1
Bs_Res_87C06	653	27807287	426	1094	1,00E-118	2	Bos taurus	NP_777140.1	ribosomal protein L3
Bs_Res_88B04	429	77735941	236	603	1,00E-61	1	Bos taurus	NP_001029667.1	ribosomal protein L35
Bs_Res_88E03	678	27807407	350	899	2,00E-95	1	Bos	NP_777217.1	ferritin, light polypeptide
Bs_Res_89D02	607	83715972	387	994	1,00E-106	3	Bos taurus	NP_001032906.1	RAN binding protein 6
Bs_Res_89D10	576	41386683	232	592	5,00E-60	2	Bos taurus	NP_776318.1	beta-2-microglobulin
Bs_Res_89F09	403	114000000	169	428	3,00E-41	-1	Bos taurus	NP_001039738.1	cyclin D1
Bs_Res_91B03	202	114000000	63.5	153	2,00E-09	1	Bos	NP_001039976.1	N-acetyltransferase ARD1
Bs_Res_91C04	592	77735771	363	931	2,00E-99	3	Bos taurus	NP_001029584.1	F-box protein 9
Bs_Res_91G11	621	27806853	368	944	1,00E-101	1	Bos taurus	NP_776359.1	lumican
Bs_Res_92E03	404	84000291	268	685	4,00E-71	-3	Bos	NP_001033247.1	myeloid leukemia factor 1
Bs_Res_93G09	384	37960066	226	577	1,00E-58	-3	Bos	AAP47876.1	NADH dehydrogenase subunit 4
Bs_Res_94D06	252	50872151	149	375	4,00E-35	3	Bos taurus	NP_001002891.1	COX5A protein
Bs_Res_94D07	384	77735849	163	412	2,00E-39	2	Bos	NP_001029619.1	fructose-1,6-bisphosphatase 1
Bs_Res_94G02	575	59857657	344	882	1,00E-93	3	Bos taurus	AAX08663.1	RAD9 homolog
Bs_Res_95C10	584	77735877	121	303	2,00E-26	-1	Bos taurus	NP_001029635.1	es1 protein
Bs_Res_96C01	440	4239700	103	256	1,00E-27	-3	Bos taurus	CAA10770.1	reverse transcriptase-like
Bs_Res_97F11	353	41386780	186	473	2,00E-46	2	Bos	NP_776637.1	UDP-glucose pyrophosphorylase 2
Bs_Res_97H08	418	28195402	172	435	4,00E-42	3	Bos taurus	NP_777186.1	major allergen BDA20
Bs_Res_98B12	289	86438072	117	292	2,00E-25	3	Bos taurus	AAI12658.1	Vl1a protein
Bs_Res_98D07	223	47824862	87.4	215	1,00E-16	3	Bos	NP_001001439.1	cytochrome c oxydase subunit 4

Table 3 S2. Putative protein with significatives hits from bovine homologous protein on nr protein database (BlastX, cutoff 80%, E value < 0.00001) to res library (n=78)

QueryID	QueryLen	Accession No.	Score	bit	Evalue	Organism	Protein ID	Description
Contig1	554	66792868	293	749	3,00E-78	Bos taurus	NP_001019712.1	ribosomal protein S15
Contig7	917	27806703	469	1207	1,00E-131	Bos taurus	NP_776438.1	CD44 antigen
Contig12	607	70778730	193	491	3,00E-48	Bos taurus	NP_001020492.1	ribosomal protein L35a
Contig19	347	81294367	143	361	2,00E-33	Bos taurus	AAI08137.1	RNF25 protein
Contig35	581	66792780	313	801	3,00E-84	Bos taurus	NP_001019642.1	ribosomal protein L27a
Contig37	209	72534798	96.7	239	2,00E-19	Bos taurus	NP_001026926.1	ribosomal protein L6
Contig39	546	37960066	311	796	9,00E-84	Bos	AAP47876.1	NADH dehydrogenase subunit 4
Contig46	557	78042530	235	600	5,00E-61	Bos taurus	NP_001030210.1	MYG1 protein
Contig58	339	116000000	57.8	138	1,00E-07	Bos	NP_001070466.1	ribosomal protein L13a
Contig61	533	60101827	150	378	2,00E-35	Bos	YP_209207.1	cytochrome c oxidase subunit I
Contig64	527	27806923	313	801	2,00E-84	Bos taurus	NP_776316.1	brain ribonuclease
Contig65	687	62460424	302	774	5,00E-81	Bos taurus	NP_001014862.1	ribosomal protein L29
Contig68	661	76644727	171	432	2,00E-41	Bos	XP_875921.1	PREDICTED: hypothetical protein
Contig74	591	76635551	73.2	178	5,00E-12	Bos	XP_601599.2	PREDICTED: hypothetical protein
Contig77	445	61097917	153	386	2,00E-36	Bos taurus	NP_001012782.1	cystatin E/M
Contig81	629	81673144	164	416	5,00E-40	Bos taurus	AAI09753.1	LOC504773 protein
Contig82	752	74353873	381	979	1,00E-104	Bos	AAI02075.1	Ribosomal protein, large, P0
Contig83	669	59042836	329	844	4,00E-89	Bos taurus	AAW83829.1	cytochrome b
Contig86	677	78369655	362	930	2,00E-99	Bos taurus	NP_001030383.1	ribosomal protein L5
Contig92	223	74267768	113	282	1,00E-26	Bos taurus	AAI02866.1	RPL13 protein
Contig98	844	76612440	207	526	4,00E-52	Bos	XP_875693.1	PREDICTED: similar to BDA11
Contig107	562	116000000	283	725	2,00E-75	Bos	NP_001070466.1	ribosomal protein L13a
Contig108	229	76612440	73.2	178	3,00E-12	Bos	XP_875693.1	PREDICTED: similar to BDA11
Contig116	825	432627	236	601	4,00E-91	Bos taurus	CAA44700.1	anti-testosterone antibody
Contig118	519	4379166	189	479	4,00E-47	Bos taurus	CAA24008.1	unnamed protein product

Table 3 S2. Putative protein with significatives hits from bovine homologous protein on nr protein database (BlastX, cutoff 80%, E value < 0.00001) to res library (n=78)

QueryID	QueryLen	Accession No.	Score	bit	Evalue	Organism	Protein ID	Description
Contig130	505	60101827	196	497	1,00E-53	Bos	YP_209207.1	cytochrome c oxidase subunit I
Contig132	945	19919708	462	1188	1,00E-129	Bos	AAM08332.1	cytochrome oxidase subunit III
Contig143	477	619	140	352	2,00E-32	Bos taurus	CAA68702.1	unnamed protein product
Contig149	449	41386679	162	410	3,00E-39	Bos taurus	NP_776322.1	carbonic anhydrase IV
Contig157	704	116000000	406	1043	1,00E-112	Bos	NP_001070466.1	ribosomal protein L13a
Contig164	742	94966839	441	1133	1,00E-122	Bos taurus	NP_001035610.1	ribosomal protein L7a
Contig165	526	31982947	271	692	9,00E-72	Bos taurus	NP_776555.1	ornithine decarboxylase 1
Contig166	410	77736047	110	275	7,00E-32	Bos	NP_001029722.1	chromatin modifying protein 2A
Contig170	690	28195402	353	905	3,00E-96	Bos taurus	NP_777186.1	major allergen BDA20
Contig175	950	50234072	403	1035	1,00E-111	Bos	YP_052702.1	ATP synthase F0 subunit 6
Contig176	521	37545835	234	598	7,00E-61	Bos	AAM95739.1	NADH dehydrogenase subunit 4
Contig179	510	72534798	223	567	3,00E-57	Bos taurus	NP_001026926.1	ribosomal protein L6
Contig186	737	164000000	277	708	1,00E-98	Bos taurus	NP_001019640.2	ribosomal protein L9
Contig187	229	37545832	84.7	208	9,00E-16	Bos	AAM95736.1	cytochrome c oxidase subunit 3
Contig198	432	27807075	207	527	9,00E-53	Bos	NP_777020.1	S100 calcium-binding protein A4
Contig201	652	27805977	162	409	1,00E-38	Bos taurus	NP_776802.1	keratin 10
Contig205	742	27543932	413	1062	1,00E-114	Bos taurus	BAC54760.1	Cytochrome b
Contig212	246	7547266	64.3	155	1,00E-09	Bos	AAB37381.2	IgG1 heavy chain constant region
Contig222	572	27805977	291	744	1,00E-77	Bos taurus	NP_776802.1	keratin 10
Bs_Sus_09H02	507	27807169	175	444	5,00E-43	Bos taurus	NP_777073.1	male-enhanced antigen
Bs_Sus_10H01	288	27807523	153	387	2,00E-36	Bos taurus	NP_777213.1	ribosomal protein P2
Bs_Sus_10H06	537	77735813	194	494	9,00E-49	Bos taurus	NP_001029601.1	RNase K
Bs_Sus_13A09	413	87196501	59.7	143	3,00E-08	Bos taurus	NP_776474.2	enolase 1
Bs_Sus_13G11	489	156000000	170	430	2,00E-41	Bos taurus	NP_001095385.1	aldolase A
Bs_Sus_14C10	571	86821596	369	946	1,00E-101	Bos	AAI05480.1	Phosphoglycerate dehydrogenase
Bs_Sus_14G04	413	76253709	116	290	3,00E-25	Bos	NP_776770.2	heat shock 70kDa protein 8

Table 6 S1. Lista of homologous proteins with match to significatives hits from bovine homologous protein on Swissprot protein database (BlastX, cutoff 80%, E value < 0.00001) to res library (n=52)

Query ID	Protein ID	Protein Description	Length	Min. eValue	Sim mean
Contig2	irf3	interferon regulatory factor 3 short	327	7,61E-16	88,00%
Contig4	rl27	60s ribosomal protein	487	1,15E-61	80,11%
Contig5	cox3	cytochrome c oxidase subunit 3	859	3,37E-117	93,00%
Contig8	rl13a	60s ribosomal proteinTable	661	1,32E-98	87,10%
Contig11	co1a1	collagen alpha-1chain	1113	2,68E-147	84,00%
Contig18	k2c5	type ii cytoskeletal 5	977	9,08E-24	95,75%
Contig29	frih	ferritin heavy chain short	468	4,14E-50	94,35%
Contig45	atp6	atp synthase subunit a	923	1,03E-74	92,75%
Contig47	k1c10	type i cytoskeletal 10	1065	6,42E-42	82,65%
Contig48	cox1	cytochrome c oxidase subunit 1	276	2,95E-15	90,60%
Contig83	zc3he	zinc finger ccch domain-containing protein 14	359	3,52E-29	92,00%
Contig90	mfap2	microfibrillar-associated protein 2 short	592	1,52E-52	85,50%
Contig96	rs15	40s ribosomal protein	486	3,71E-52	87,85%
Contig124	k2c74	type ii cytoskeletal 74ame	628	5,76E-101	90,20%
Contig135	orc3	origin recognition complex subunit 3	604	3,46E-68	87,50%
Contig143	nmrl1	-like family domain-containing protein 1	606	1,78E-96	87,80%
Contig158	rl32	60s ribosomal protein	248	1,78E-04	86,50%
Contig180	cox7c	cytochrome c oxidase subunitmitochondrial	401	1,41E-25	93,55%
Contig185	rl10	60s ribosomal protein	798	1,31E-120	89,60%
Contig201	chur	protein churchill	401	1,19E-53	91,57%
Contig221	liga	ligatin	539	6,08E-60	93,20%
Contig226	at5f1	atp synthase subunitmitochondrial flags	664	4,30E-81	88,33%
Bs_Res_03D11_rc	capg	macrophage-capping proteiname	267	1,93E-14	82,80%
Bs_Res_03H07_rc	rla2	60s acidic ribosomal protein p2	519	1,01E-24	80,85%
Bs_Res_00B11	prr9	proline-rich protein 9	634	4,43E-40	88,30%

Table 6 S1. Lista of homologous proteins with match to significatives hits from bovine homologous protein on Swissprot protein database (BlastX, cutoff 80%, E value < 0.00001) to res library (n=52)

Query ID	Protein ID	Protein Description	Length	Min. eValue	Sim mean
Bs_Res_01A09	hint1	histidine triad nucleotide-binding protein 1	356	3,61E-36	80,15%
Bs_Res_01E05	vatf	v-type proton atpase subunit f short	417	1,17E-14	84,30%
Bs_Res_01F10	tr112	trm112-like protein	192	2,06E-16	84,80%
Bs_Res_01H10	armet	protein armetame	469	7,55E-52	86,30%
Bs_Res_02C09	cot2	coup transcription factor 2	681	6,12E-116	80,45%
Bs_Res_04F01	gatl3	gats-like protein 3	626	7,19E-104	85,30%
Bs_Res_04H04	rs27a	40s ribosomal protein s27a	370	3,05E-20	85,50%
Bs_Res_87C06	rl3	60s ribosomal protein l3	653	6,33E-115	88,65%
Bs_Res_88B04	rl35	60s ribosomal protein l35	429	3,64E-46	89,20%
Bs_Res_88E03	fril	ferritin light chain short	678	2,81E-91	85,90%
Bs_Res_88F04	myl9	myosin regulatory light polypeptide 9	654	9,31E-82	89,15%
Bs_Res_89A12	amd	peptidyl-glycine alpha-amidating monooxygenase short	581	7,27E-12	91,30%
Bs_Res_89D10	b2mg	beta-2-microglobuliname	576	8,10E-56	86,50%
Bs_Res_89E01	lerl1	leptin receptor overlapping transcript-like 1	494	6,56E-44	86,90%
Bs_Res_91C04	fbx9	f-box only protein 9	592	3,02E-93	97,00%
Bs_Res_91G07	tm111	transmembrane protein 111	550	1,31E-12	96,25%
Bs_Res_92E03	mlf1	myeloid leukemia factor 1	404	5,64E-67	85,00%
Bs_Res_93B12	fabp4	fatty acid-bindingadipocyteame	490	3,90E-49	81,45%
Bs_Res_94C11	vkor1	vitamin k epoxide reductase complex subunit 1	471	3,61E-54	87,00%
Bs_Res_94D04	ssrd	translocon-associated protein subunit delta precursor (trap-delta)	177	9,06E-09	99,25%
Bs_Res_94D06	cox5a	cytochrome c oxidase subunitmitochondrial	252	1,76E-20	89,85%
Bs_Res_94D07	f16p1	fructose--bisphosphatase 1 short	384	2,53E-35	86,15%
Bs_Res_94D09	echd1	enoyl-hydratase domain-containing protein 1	391	3,14E-62	84,70%
Bs_Res_97F11	ugpa	utp--glucose-1-phosphate uridylyltransferaseame	353	2,14E-42	83,30%
Bs_Res_98B07	rsmn	small nuclear ribonucleoprotein-associated protein n short	347	1,51E-11	95,50%
Bs_Res_98D07	cox41	cytochrome c oxidase subunit 4 isoformmitochondrial	223	1,76E-12	85,20%
Bs_Res_99D04	at5g3	atp synthase lipid-bindingmitochondrial	543	4,96E-49	86,90%

Table 6 S2. Lista of homologous proteins with match to significatives hits from bovine homologous protein on Swissprot protein database (BlastX, cutoff 80%, E value < 0.00001) to *sus* library (n=69)

Query ID	Protein ID	Protein description	Length	Min. eValue	Sim. mean
Contig1	rs15	40s ribosomal protein s15	554	4,09E-55	87,55%
Contig7	cd44	cd44 antigen	917	3,27E-89	81,50%
Contig16	lonp2	peroxisomal lon protease homolog 2	343	2,59E-03	84,40%
Contig19	rnf25	e3 ubiquitin-protein ligase rnf25	347	9,44E-14	89,70%
Contig21	glrx1	glutaredoxin-1	564	7,05E-17	87,20%
Contig35	rl27a	60s ribosomal protein l27a	581	7,90E-67	85,15%
Contig37	rl6	60s ribosomal protein l6	209	2,94E-15	88,30%
Contig61	cox1	cytochrome c oxidase subunit 1	533	4,41E-31	91,40%
Contig63	nu1m	nadh-ubiquinone oxidoreductase chain 1	990	4,53E-120	92,75%
Contig64	rnbr	brain ribonuclease short	527	1,94E-63	88,95%
Contig65	rl29	60s ribosomal protein l29	687	4,96E-59	84,41%
Contig67	s10ac	protein s100-a12	340	2,05E-08	98,00%
Contig80	in80e	ino80 complex subunit e	558	1,15E-27	100,00%
Contig82	rla0	60s acidic ribosomal protein p0	752	9,81E-75	85,65%
Contig86	rl5	60s ribosomal protein l5	677	2,28E-85	82,50%
Contig91	cox2	cytochrome c oxidase subunit 2	349	1,82E-41	96,75%
Contig92	rl13	60s ribosomal protein l13	223	2,99E-20	91,50%
Contig96	rl10	60s ribosomal protein l10	713	8,52E-102	87,40%
Contig99	rl32	60s ribosomal protein l32	583	2,76E-60	84,95%
Contig104	rs17	40s ribosomal protein s17	564	3,98E-37	81,75%
Contig107	rl13a	60s ribosomal protein l13a	562	7,72E-64	85,00%
Contig108	s10a7	protein s100-a7	229	5,09E-04	83,00%
Contig115	rl3	60s ribosomal protein l3	606	3,73E-78	81,25%
Contig118	nu3m	nadh-ubiquinone oxidoreductase chain 3	519	4,25E-23	87,90%
Contig132	cox3	cytochrome c oxidase subunit 3	945	8,70E-117	92,30%

Table 6 S2. Lista of homologous proteins with match to significatives hits from bovine homologous protein on Swissprot protein database (BlastX, cutoff 80%, E value < 0.00001) to *sus* library (n=69). Cont.

Query ID	Protein ID	Protein description	Length	Min. eValue	Sim. mean
Contig137	rs2	40s ribosomal protein s2	726	9,80E-101	84,20%
Contig142	cox1	cytochrome c oxidase subunit 1	1540	0.0	95,40%
Contig143	pgs2	decorin	477	3,53E-28	89,45%
Contig149	cah4	carbonic anhydrase 4	449	5,95E-21	85,20%
Contig157	rl13a	60s ribosomal protein l13a	704	7,79E-100	86,50%
Contig161	rl35	60s ribosomal protein l35	513	1,55E-46	89,25%
Contig162	k1c17	type i cytoskeletal 17	345	8,54E-15	81,50%
Contig164	rl7a	60s ribosomal protein l7a	742	2,42E-102	80,70%
Contig168	med25	mediator of rna polymerase ii transcription subunit 25	318	6,32E-02	100,00%
Contig183	co1a1	collagen alpha-1chain	500	9,65E-28	86,20%
Contig186	rl9	60s ribosomal protein l9	737	6,93E-87	81,05%
Contig196	rs12	40s ribosomal protein s12	516	1,79E-66	81,84%
Contig205	cyb	cytochrome b	742	2,37E-66	93,80%
Contig206	vime	vimentin	745	1,72E-47	93,85%
Contig222	k1c10	type i cytoskeletal 10	572	1,41E-12	83,65%
Bs_Sus_09H02	mea1	male-enhanced antigen 1 short	507	2,04E-27	99,30%
Bs_Sus_10F01	hb21	bola class ii histocompatibilitydqb*0101 beta chain short	532	5,42E-14	94,70%
Bs_Sus_10F11	ef1d	elongation factor 1-delta short	561	2,65E-64	84,40%
Bs_Sus_10H06	rnk	ribonuclease kappa short	537	1,61E-44	81,25%
Bs_Sus_11F06	ndub9	nadh dehydrogenase1 beta subcomplex subunit 9	652	2,26E-103	90,30%
Bs_Sus_14F02	at5g2	atp synthase lipid-bindingmitochondrial	636	4,48E-32	90,00%
Bs_Sus_16F11	tyb10	thymosin beta-10	450	5,16E-12	90,70%
Bs_Sus_17F08	ef1g	elongation factor 1-gamma short	459	1,60E-51	81,10%
Bs_Sus_17G10	cox7c	cytochrome c oxidase subunitmitochondrial	375	1,01E-23	90,10%

Table 6 S2. Lista of homologous proteins with match to significatives hits from bovine homologous protein on Swissprot protein database (BlastX, cutoff 80%, E value < 0.00001) to *sus* library (n=69). Cont.

Query ID	Protein ID	Protein description	Length	Min. eValue	Sim. mean
Bs_Sus_18D06	co1a2	collagen alpha-2chain	587	7,50E-97	81,35%
Bs_Sus_18F06	wdr70	wd repeat-containing protein 70	210	2,74E-13	97,20%
Bs_Sus_18F09	rl28	60s ribosomal protein l28	291	1,59E-13	98,50%
Bs_Sus_20G02	pihd1	pih1 domain-containing protein 1	627	1,80E-78	82,30%
Bs_Sus_21C06	ndus5	nadh dehydrogenaseiron-sulfur protein 5	445	1,08E-54	90,20%
Bs_Sus_21D10	apod	apolipoprotein d short	615	7,78E-64	90,40%
Bs_Sus_22E07	psme1	proteasome activator complex subunit 1	378	1,04E-28	81,90%
Bs_Sus_22F11	qcr8	cytochrome b-c1 complex subunit 8	465	1,23E-38	87,10%
Bs_Sus_22G04	eif3m	eukaryotic translation initiation factor 3 subunit m short	407	6,48E-39	85,80%
Bs_Sus_23C07	tf2h3	general transcription factor iih subunit 3	415	2,06E-32	85,30%
Bs_Sus_23F02	rl13	60s ribosomal protein l13	444	3,60E-34	84,65%
Bs_Sus_24E04	dd19a	atp-dependent rna helicase ddx19a	553	2,22E-92	80,25%
Bs_Sus_24F04	ndus6	nadh dehydrogenaseiron-sulfur proteinmitochondrial	485	1,87E-40	88,90%
Bs_Sus_24F05	qcr10	cytochrome b-c1 complex subunit 10	398	7,77E-24	96,70%
Bs_Sus_26A07	timp2	metalloproteinase inhibitor 2	489	2,28E-57	81,85%
Bs_Sus_28C10	nucb1	nucleobindin-1 flags: precursor	583	6,53E-77	89,10%
Bs_Sus_29C05	nduad	nadh dehydrogenase1 alpha subcomplex subunit 13	481	2,73E-60	85,10%
Bs_Sus_29C12	nu6m	nadh-ubiquinone oxidoreductase chain 6	448	2,65E-45	87,20%
Bs_Sus_31B08	apoe	apolipoprotein e short	359	2,52E-27	83,40%
Bs_Sus_31C04	ars2	arsenite-resistance protein 2	384	9,00E-25	98,80%

Table 6 S3. Full length sequences bovine proteins matched with Swissprot protein database to sus dataset (n=56)

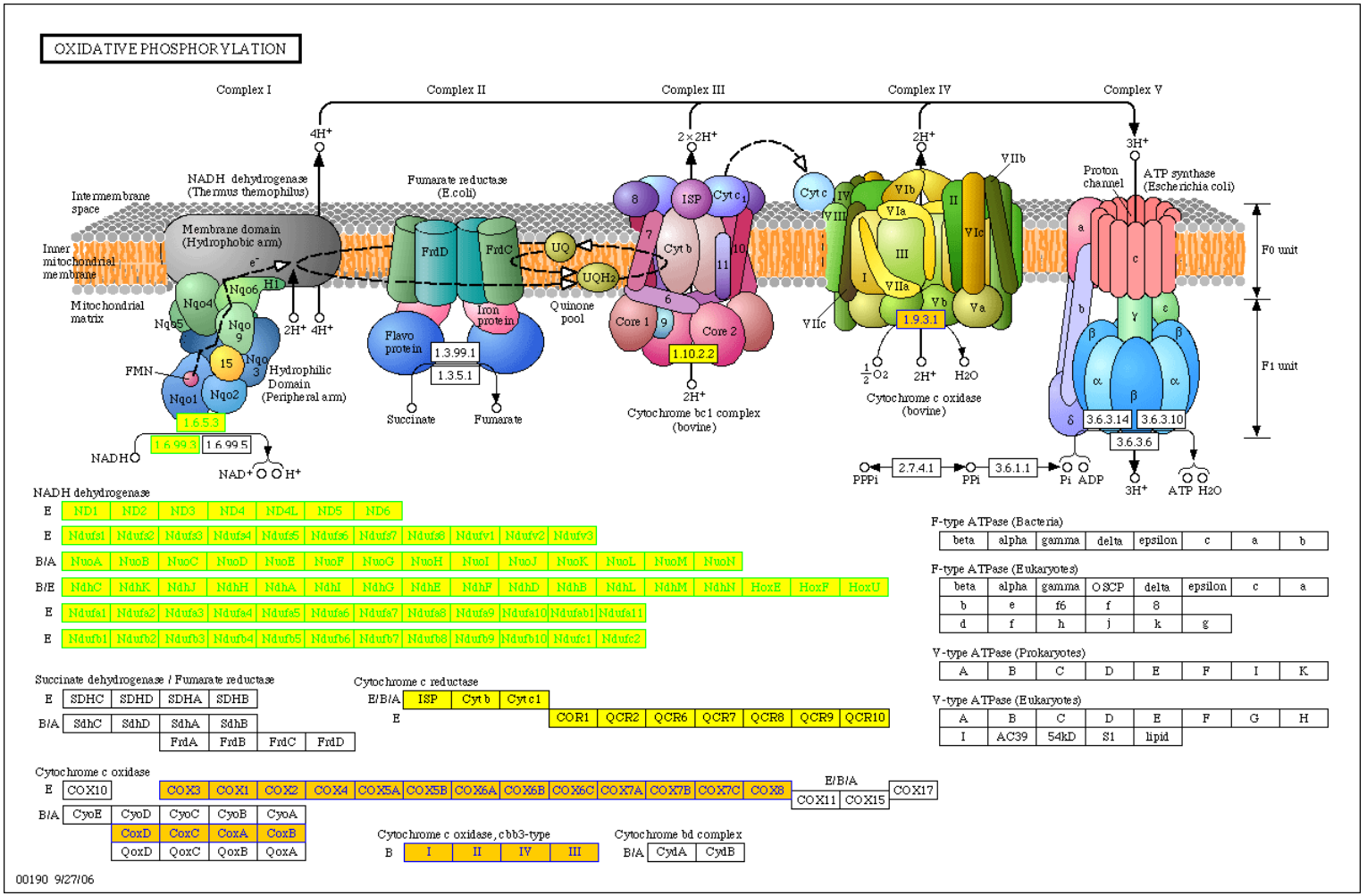
Query ID	Protein ID	Seq description	Length	Min. eValue	Sim mean
Contig1	rs15	40s ribosomal protein s15	554	4,09E-55	87,55%
Contig7	cd44	cd44 antigen	917	3,27E-89	81,50%
Contig16	lonp2	peroxisomal lon protease homolog 2	343	2,59E-03	84,40%
Contig19	rnf25	e3 ubiquitin-protein ligase rnf25	347	9,44E-14	89,70%
Contig21	glrx1	glutaredoxin-1	564	7,05E-17	87,20%
Contig35	rl27a	60s ribosomal protein l27a	581	7,90E-67	85,15%
Contig61	cox1	cytochrome c oxidase subunit 1	533	4,41E-31	91,40%
Contig63	nu1m	nadh-ubiquinone oxidoreductase chain 1	990	4,53E-120	92,75%
Contig64	rnbr	brain ribonuclease short	527	1,94E-63	88,95%
Contig67	s10ac	protein s100-a12	340	2,05E-08	98,00%
Contig80	in80e	ino80 complex subunit e	558	1,15E-27	100,00%
Contig82	rla0	60s acidic ribosomal protein p0	752	9,81E-75	85,65%
Contig86	rl5	60s ribosomal protein l5	677	2,28E-85	82,50%
Contig91	cox2	cytochrome c oxidase subunit 2	349	1,82E-41	96,75%
Contig92	rl13	60s ribosomal protein l13	223	2,99E-20	91,50%
Contig99	rl32	60s ribosomal protein l32	583	2,76E-60	84,95%
Contig104	rs17	40s ribosomal protein s17	564	3,98E-37	81,75%
Contig108	s10a7	protein s100-a7	229	5,09E-04	83,00%
Contig115	rl3	60s ribosomal protein l3	606	3,73E-78	81,25%
Contig118	nu3m	nadh-ubiquinone oxidoreductase chain 3	519	4,25E-23	87,90%
Contig132	cox3	cytochrome c oxidase subunit 3	945	8,70E-117	92,30%
Contig142	cox1	cytochrome c oxidase subunit 1	1540	0.0	95,40%
Contig143	pgs2	decorin	477	3,53E-28	89,45%
Contig149	cah4	carbonic anhydrase 4	449	5,95E-21	85,20%
Contig162	k1c17	type i cytoskeletal 17	345	8,54E-15	81,50%
Contig168	med25	mediator of rna polymerase ii transcription subunit 25	318	6,32E-02	100,00%
Contig183	co1a1	collagen alpha-1chain	500	9,65E-28	86,20%
Contig205	cyb	cytochrome b	742	2,37E-66	93,80%
Contig206	vime	vimentin	745	1,72E-47	93,85%

Table 6 S3. Full length sequences bovine proteins matched with Swissprot protein database to sus dataset (n=56)

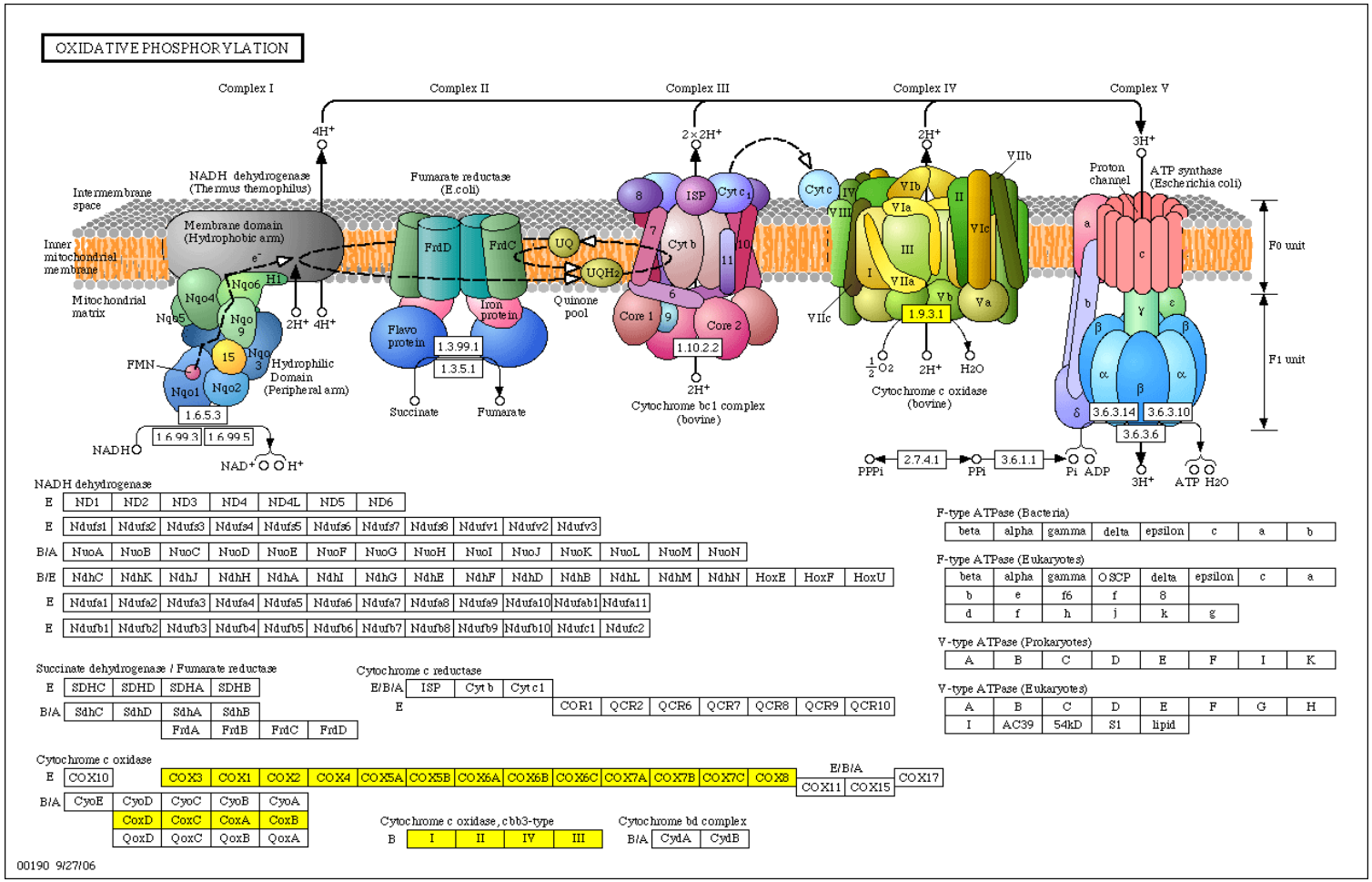
Query ID	Protein ID	Seq description	Length	Min. eValue	Sim mean
Contig222	k1c10	type i cytoskeletal 10	572	1,41E-12	83,65%
Bs_Sus_09H02	mea1	male-enhanced antigen 1 short	507	2,04E-27	99,30%
Bs_Sus_10F01	hb21	bola class ii histocompatibilitydqb*0101 beta chain short	532	5,42E-14	94,70%
Bs_Sus_10F11	ef1d	elongation factor 1-delta short	561	2,65E-64	84,40%
Bs_Sus_10H06	rnk	ribonuclease kappa short	537	1,61E-44	81,25%
Bs_Sus_11F06	ndub9	nadh dehydrogenase1 beta subcomplex subunit 9	652	2,26E-103	90,30%
Bs_Sus_14F02	at5g2	atp synthase lipid-bindingmitochondrial	636	4,48E-32	90,00%
Bs_Sus_16F11	tyb10	thymosin beta-10	450	5,16E-12	90,70%
Bs_Sus_17F08	ef1g	elongation factor 1-gamma short	459	1,60E-51	81,10%
Bs_Sus_17G10	cox7c	cytochrome c oxidase subunitmitochondrial	375	1,01E-23	90,10%
Bs_Sus_18D06	co1a2	collagen alpha-2chain	587	7,50E-97	81,35%
Bs_Sus_21C06	ndus5	nadh dehydrogenaseiron-sulfur protein 5	445	1,08E-54	90,20%
Bs_Sus_21D10	apod	apolipoprotein d short	615	7,78E-64	90,40%
Bs_Sus_22E07	psme1	proteasome activator complex subunit 1	378	1,04E-28	81,90%
Bs_Sus_22F11	qcr8	cytochrome b-c1 complex subunit 8	465	1,23E-38	87,10%
Bs_Sus_22G04	eif3m	eukaryotic translation initiation factor 3 subunit m short	407	6,48E-39	85,80%
Bs_Sus_23C07	tf2h3	general transcription factor iih subunit 3	415	2,06E-32	85,30%
Bs_Sus_23F02	rl13	60s ribosomal protein l13	444	3,60E-34	84,65%
Bs_Sus_24E04	dd19a	atp-dependent rna helicase ddx19a	553	2,22E-92	80,25%
Bs_Sus_24F04	ndus6	nadh dehydrogenaseiron-sulfur proteinmitochondrial	485	1,87E-40	88,90%
Bs_Sus_24F05	qcr10	cytochrome b-c1 complex subunit 10	398	7,77E-24	96,70%
Bs_Sus_26A07	timp2	metalloproteinase inhibitor 2	489	2,28E-57	81,85%
Bs_Sus_28C10	nucb1	nucleobindin-1 flags: precursor	583	6,53E-77	89,10%
Bs_Sus_29C05	nduad	nadh dehydrogenase1 alpha subcomplex subunit 13	481	2,73E-60	85,10%
Bs_Sus_29C12	nu6m	nadh-ubiquinone oxidoreductase chain 6	448	2,65E-45	87,20%
Bs_Sus_31B08	apoe	apolipoprotein e short	359	2,52E-27	83,40%
Bs_Sus_31C04	ars2	arsenite-resistance protein 2	384	9,00E-25	98,80%

- Via metabólica:

Fosforilação Oxidativa sus



Esquema representando os transcritos identificados no *sus dataset* como pertencentes a via metabólica Fosforilação Oxidativa.



Esquema representando os transcritos identificados no grupo *sus* como pertencentes a via metabólica Fosforilação Oxidativa.

Capítulo 3

Avalição da cobertura gênica e do número esperado de novos transcritos em bibliotecas de cDNA

Resumo

As Etiquetas de Sequências Expressas (*Expressed Sequence Tags – EST*) são geradas pelo sequenciamento parcial de transcritos isolados de mRNA que foram convertidos em cDNA. Em análises de *EST* , dois tipos de informações podem ser obtidos: sequência do transcrito e abundância do transcrito. O número de *EST* dentro dos grupos representa a abundância desse transcrito ou dessa espécie de mRNA em cada biblioteca. A redundância dos transcritos tem efeito direto na taxa de descoberta gênica, além de fornecer informações sobre a eficência do sequenciamento e a diversidade dos genes expressos no tecido. Nesse trabalho, procurou-se avaliar a redundância nas *EST* amostradas e o número esperado de novos genes em uma amostragem futura. As *EST* imativas de cobertura gênica indicam a presença de 401,8 (49%) genes únicos no grupo resistente (RES). Para o grupo de animais suscetível (SUS), essa *EST* imativa foi de 392,4 (40%) genes únicos. O número esperado de *EST* para descobrir um novo gene foi maior na RES (1,94%) que no SUS (1,66%). *EST* es resultados indicam que, para a descoberta de um novo gene, são necessárias 1,94 e 1,66 *EST* na RES e SUS, respectivamente. Isto indica maior número de genes redundantes na biblioteca RES e, provavelmente, uma nova amostragem de transcritos conduzirá a menor descoberta de novos genes. Além disto, os resultados sugerem que RES pode ter maior número de erros associados ao agrupamento feito pelo programa de clusterização podem estar inflacionando essas estimativas.

Palavras-chave: cobertura gênica, *EST*, sequenciamento

Introdução

As Etiquetas de Sequências Expressas (*Expressed Sequence Tags – EST*) são geradas pelo sequenciamento parcial de transcritos isolados de mRNA que foram convertidos em cDNA. Em análises de *ESTs* , dois tipos de informações podem ser obtidos: sequência e abundância do transcrito, que podem ser obtidas pelo agrupamento que identifica e monta *EST* que compartilham determinada similaridade entre elas. A montagem de *EST* em cada grupo de sequência pode ser representada por transcritos parciais ou completamente restaurados, desde que não haja erros no processo de agrupamento. O número de *EST* dentro de cada grupo representa a abundância desse transcrito ou dessa espécie de mRNA em cada biblioteca. Informação sobre a abundância de transcrito em dados de *EST* pode ser usada para estimar a cobertura da predição gênica. Por outro lado, informação de sequência facilita a construção de sequências consensos (Okubo et al., 1992).

As *EST* desempenham papel importante na identificação, detecção e caracterização do transcriptoma de organismos e de partes deles, tornando-se uma alternativa atraente e eficaz ao sequenciamento completo do genoma. As sequências de transcritos resultantes e suas correspondentes abundâncias têm sido usadas na identificação de genes e no nível de expressão desses genes (Okubo et al., 1992).

O sequenciamento de *EST* ainda é visto como um procedimento de custo elevado. Portanto, adequar a relação custo-eficácia ao experimento é fundamental (Emrich et al., 2007), o que sugere a necessidade de avaliar a redundância relativa de várias bibliotecas preparadas a partir do mesmo organismo ou dos mesmos tecidos, a fim de detectar quais dessas bibliotecas produzem novos genes a uma taxa mais elevada. Além disso, existem protocolos de “normalização” que visam tornar as frequências de genes na biblioteca mais uniformes, melhorando a taxa de descoberta gênica. No entanto, a realização desses protocolos ainda tem custo elevado. Portanto, a decisão de se avançar com o sequenciamento de uma biblioteca não-normalizada ou recorrer a um procedimento de normalização depende do

equilíbrio cuidadoso dos custos envolvidos. Essa decisão é baseada em estimativas estatísticas da cobertura dos transcritos e do número esperado de novos genes em uma amostragem futura (Adams et al., 1991).

Material e métodos

Os conjuntos de dados analisados consistem de amostras de *EST* obtidas de bibliotecas de cDNA não-normalizadas geradas a partir de dois *pools* de tecidos de pele proveniente de animais F_2 (1/2Holandês:1/2 Gir) avaliados como resistentes (RES) e suscetíveis (SUS) e infestados com carrapato *Riphicephalus (Boophilus) microplus*. As *EST* foram agrupadas e montadas em *contigs* pelo programa CAP3 (Contig Assembling Program 3) (Huang & Madan, 1999) e configuradas para parâmetros padrão (definir), de modo que *EST* que não formaram *contigs* foram nomeadas *singlets*. Esses conjuntos de dados foram descritos previamente no capítulo 2.

Estimativas para predição gênica e número esperado de novos genes foram obtidas como descrito por Susko & Roger (2004): os dados foram ajustados para um modelo binomial negativo derivado da distribuição misturada Gamma-Poisson e o número de genes preditos foi obtido por:

$$\eta_x = \eta_1 \frac{\Gamma(x+\alpha)}{x!\Gamma(1+\alpha)} \gamma^{x-1} \qquad (1)$$

em que α, β e γ são parâmetros da distribuição gama. O parâmetro α pode assumir valores entre -1 e 1 e o parâmetro β, entre 0 e 1. O parâmetro γ é determinado pela expressão:

$$\gamma = \frac{\beta}{(1+\beta)} \qquad (2)$$

Esse modelo tem a vantagem de desconsiderar o problema com os genes não-observados $(x_0 = 0)$ e isso permite que o parâmetro α tenha valores menores que zero e qualquer valores maiores que -1 quando valores finitos para $\eta_1, \eta_2, \ldots$ são considerados.

2) [0, *t*].

A imativa do número de *EST* para descobrir um novo gene pode ser obtida pela seguinte equação:

$$\Delta_{\alpha\gamma}(t) = -\eta_1 \frac{\{(1+\gamma t)^{-\alpha} - 1\}}{(\gamma\alpha)} \quad (3)$$

Os parâmetros foram obtidos por métodos iterativos (algoritmo EM - *Expectation and maximization*) por máxima verossimilhança (Efron & Thisted, 1976).

Resultados

Agrupamento das EST em genes únicos

Duas bibliotecas de cDNA não-normalizadas foram geradas a partir de biópsias de pele obtidas de animais F_2 infestados com *Riphicephalus B. microplus*. As *EST* foram agrupadas usando CAP3 com os parâmetros *default*. Para os dados de *EST* , o perfil agrupamento de n genes foi diretamente sumarizado dos resultados do CAP3. A Figura 1 sumariza os dois conjuntos de *EST* quanto à distribuição em possíveis grupos de genes. Para a biblioteca do grupo dos aniamais resistentes, foram gerados 1.207 transcritos com 820 genes únicos, os quais foram distribuídos em dez grupos com níveis de expressão 619, 135, 29, 17, 5, 6, 5, 1, 2, 1. No primeiro nível, 619 genes apareceram uma única vez e, no nível mais baixo, 1 gene foi representado 31 vezes no total de 820 genes.

Para a biblioteca SUS, foram gerados 1.350 transcritos com 981 genes únicos, agrupados em 12 clusters com níveis de expressão 806, 108, 21, 19, 8, 7, 3, 3, 2, 2, 1, 1. Nesse conjunto de dados, 819 genes apareceram apenas uma vez e 1 gene com 13 possíveis transcritos representando o mesmo gene.

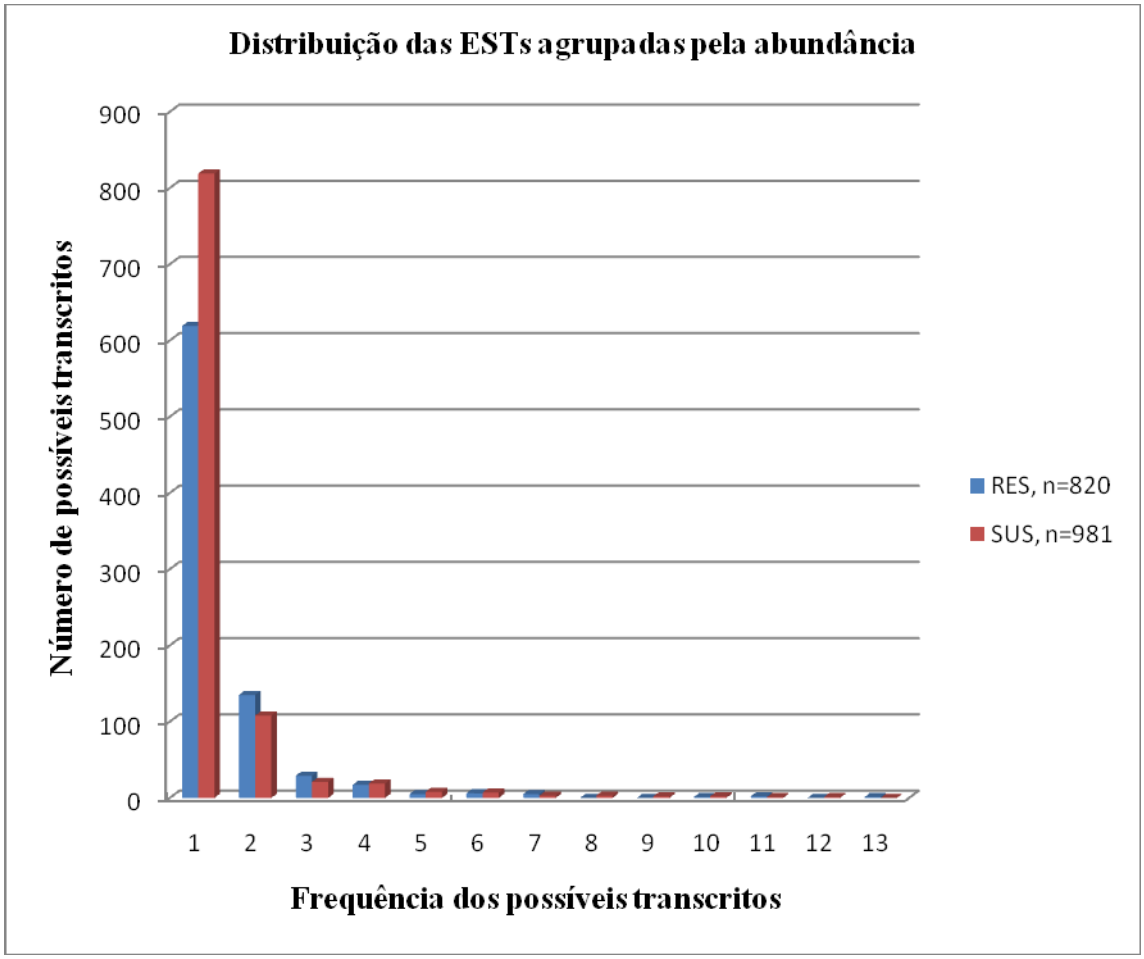


Figura 1. Distribuição das *ESTs* em relação ao número possível de transcritos para as bibliotecas RES e SUS feitos pelo programa CAP3.

Estimativas para cobertura gênica e número de novos genes

Os parâmetros α e γ da equação (1) convergiram após 16 (RES) e 15 (SUS) iterações. Os valores estimados de α e γ para os conjuntos de dados foram: $\alpha_{estimado}$ = -0,4944 e $\gamma_{estimado}$=0,8549 para RES e $\alpha_{estimado}$ =-0,7115 e $\gamma_{estimado}$=1,0586 para SUS.

As estimativas para cobertura foram 0,49 (±0,02) e 0,40 (±0,02) para as bibliotecas RES (n=820) e SUS (n=981) e indicam , respectivamente, que as *ESTs* amostradas representam redundância de 49 e 40%. Estimativas para o número de *EST* requeridas para descobrir um novo gene, considerando amostragem adicional de tamanho *t,* também foram obtidas. As estimativas para RES foram 1,94 (RES) e 1,66 (SUS) e os intervalos de confiança preditos

(α=5%), foram 1,87(±2,02) e 1,61(±1,72), para RES e SUS, respectivamente (Tabela 1). Esses valores indicam que, em média, serão necessários aproximadamente duas *ESTs* para descobrir um novo gene.

Tabela 1. Estimativas paramétricas para predições gênicas e para o número esperado de novos genes

Conjuntos de *EST*	Predição gênica (Erro-padrão)	Número esperado de novos genes	
		EST (Erro-padrão)	Intervalo de confiança 95%
RES	0,49 (±0,02)	1,94 (±0,04)	LI=1,87; LS=2,02
SUS	0,40 (±0,02)	1,66 (±0,03)	LI=1,61; LS=1,72

As estimativas para o intervalo de confiança foram obtidas para um valor de α=0,05. LI e LS representam os limites inferiores e superiores do intervalo de confiança, respectivamente.

Nas Figuras 2 e 3 são apresentadas as estimativas paramétricas para o número esperado de novos genes em uma segunda amostragem de *EST* com relação aos dois conjuntos de dados RES e SUS, respectivamente.

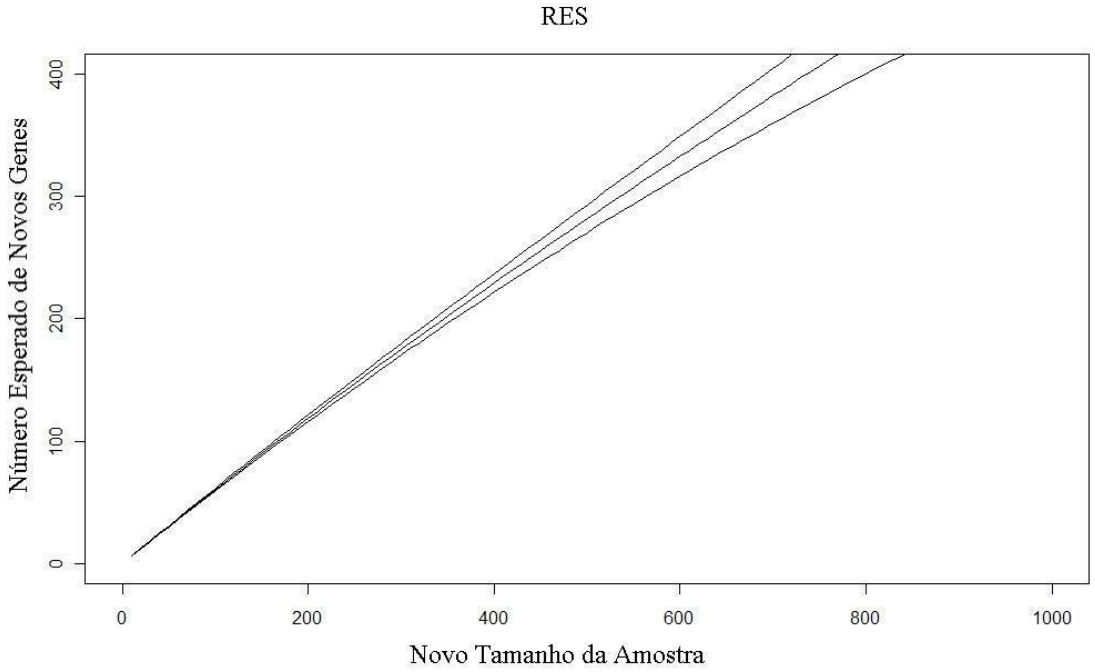


Figura 2. Estimativas para número esperado de novos genes em função de uma nova amostragem de transcritos para o grupo resistente (RES). A linha central fornece a estimativa e as linhas em torno o intervalo de confiaança a 95%.

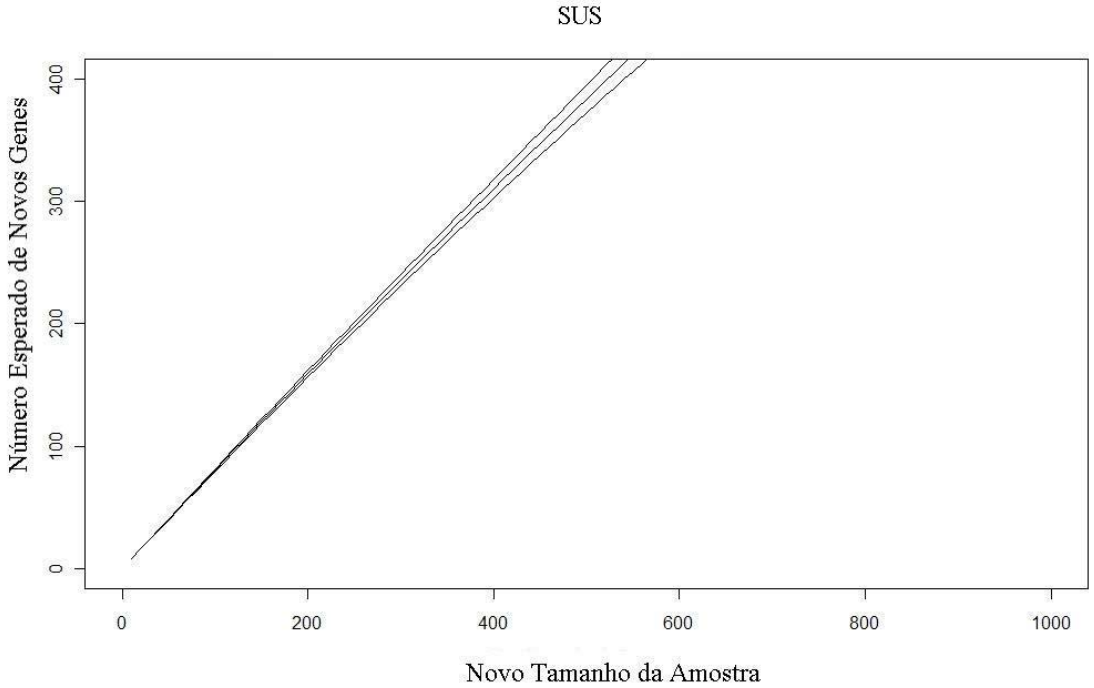


Figura 3. Estimativas para número esperado de novos genes em função de uma nova amostragem de transcritos para o grupo suscetível (SUS). A linha central fornece a estimativa e as linhas em torno do intervalo de confiança a 95%.

Os agrupamentos das *EST* feitos pelo programa CAP3 foram avaliados com estimativas paramétricas, como descritos por Susko e Roger (2004) e comparadas a estimadores não-paramétricos. As Figuras 4 e 5 mostram o padrão côncavo no tamanho da *EST* amostrada, o qual indica tendência de declínio da eficiência com sequenciamento adicional quando se consideram estimativas não-paramétricas.

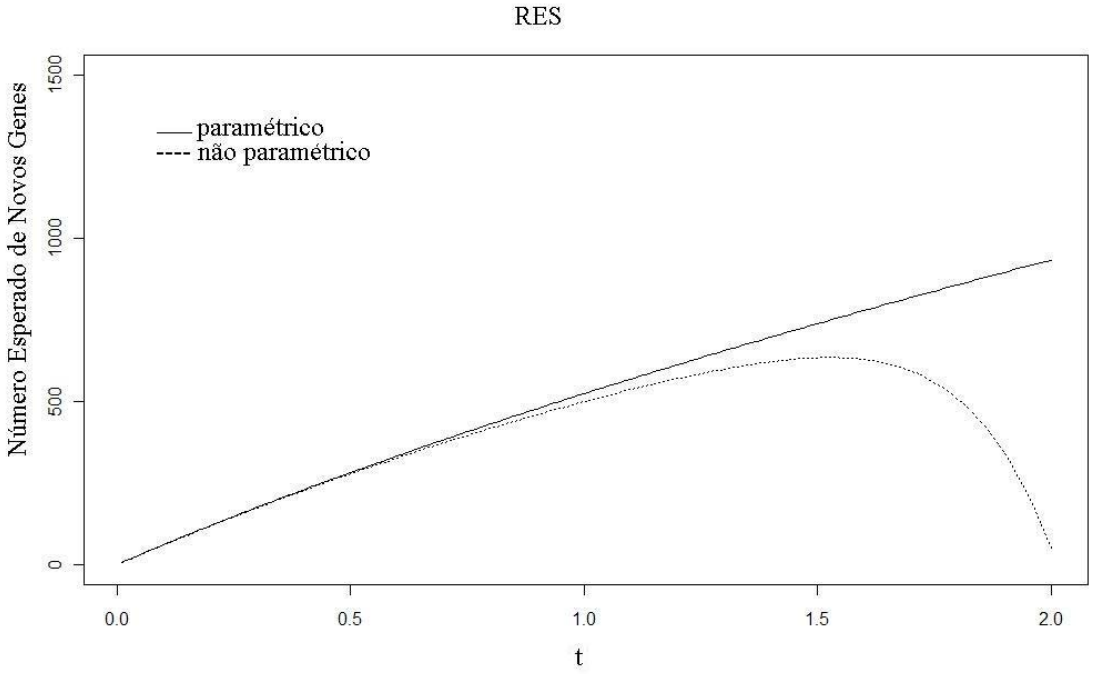


Figura 4. Estimativa para o número esperado de novos genes em uma nova amostra de tamanho *tn* como uma função de múltiplos *t* do tamanho inicial *n*.
Os números estimados são dados para estimativas paramétricas e não paramétricas paras a *EST* das bibliotecas RES. Deve ser observada a diferença no comportamento para *t*>1.

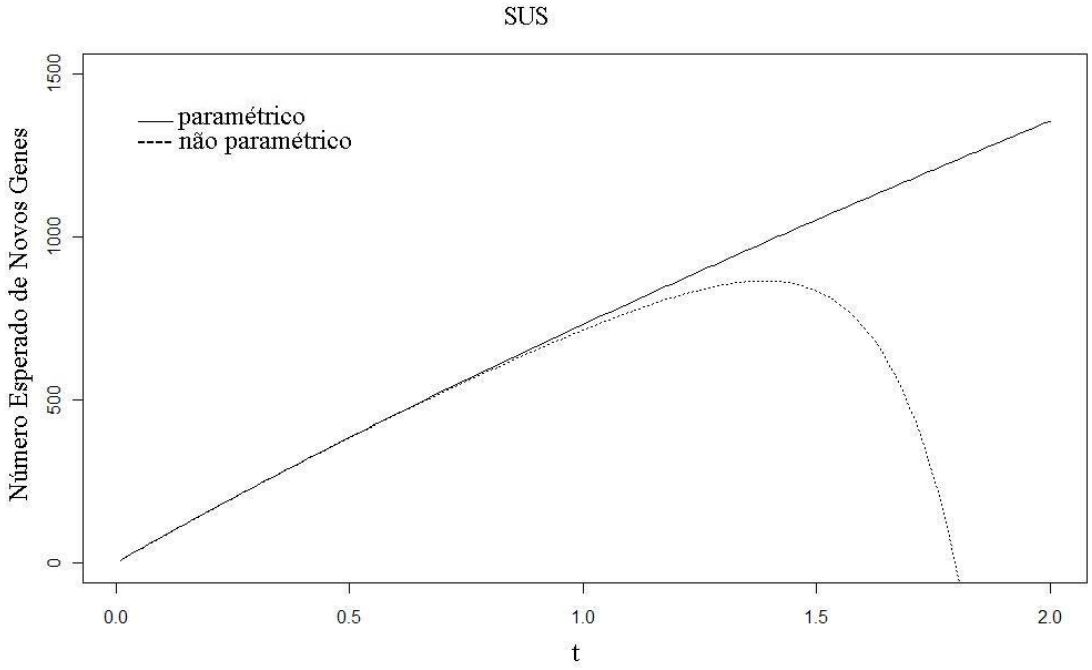


Figura 5. Estimativa para o número esperado de novos genes em uma nova amostra de tamanho *tn* como uma função de múltiplos *t* do tamanho inicial *n*.
Os números estimados são dados para estimativas paramétricas e não-paramétricas para as *ESTs* das bibliotecas SUS.

Discussão

Dados de *EST* são gerados pelo sequenciamento de DNA de muitos clones com cada clone representando uma cópia do cDNA para um gene esperado no tecido-alvo. Desse modo, *n* representa o número desconhecido de genes expressos no conjunto de dados de *EST* . Com base nas informações de sequências, as *EST* são agrupadas e, para cada grupo de transcritos, é assumido representar um único gene expresso. Um agrupamento de etiquetas de sequências expressas deve representar a verdadeira distribuição das *EST* amostradas na biblioteca de cDNA (Wang et al., 2004)

Os resultados do agrupamento de *EST* (Tabela 2) sugerem que aproximadamente 820 e 981 genes estavam presentes na biblioteca RES e SUS, respectivamente. Por outro lado, a cobertura imada indica 401,8 (49%) genes presentes na RES e 392,4 (40%) genes presentes na biblioteca SUS. As estimativas para cobertura apresentada por ambas as bibliotecas indicam que RES possui maior redundância que a apresentada pelo agrupamento SUS para um número menor de genes. Não foram observadas sobreposições nas regiões do intervalo de confiança entre os dois grupos de *EST* e o número esperado de *EST* para descobrir um novo gene é maior na RES que para a SUS, logo é menos provável que uma nova amostragem para RES leve a novos genes. Esses resultados sugerem que RES pode ter maior número de erros associados ao agrupamento feito pelo CAP3.

Erros podem ser atribuídos a diferentes fontes, por exemplo, qualidade e direção do sequenciamento e contaminação com organismos simbiontes presentes em amostras tomadas do ambiente. Esses erros podem viesar o número de genes observados entre 35-40% (Wang et al., 2004). Em análise dos valores de qualidade das sequências atribuídas pelo programa Phred, os menores valores médios de qualidade foram observados para RES. Portanto, os erros de agrupamento podem estar associados a menor qualidade média das sequências de *EST* obtidas pelo sequenciamento para RES.

Outros fatores também foram considerados como possíveis causas desses erros. Etiquetas de sequências expressas geradas pelo sequenciamento da extremidade 5’ podem conter erros associados à falsa

separação dos genes em grupos distintos. Esses erros podem ser problemáticos, em razão da insuficiente sobreposição entre sequências para o mesmo gene, o que pode explicar até 80% desses erros de agrupamento (Wang et al., 2004). Para os dados de *EST* analisados, 82,6% e 89,6% dos transcritos são representados por sequências obtidas pelo sequenciamento da extremidade 5', respectivamente, RES e SUS. No entanto, nenhuma inferência pôde ser feita quanto aos dados analisados em decorrência da representação similar de transcritos 5' em ambas as bibliotecas.

Para o modelo paramétrico, estimativas razoáveis de parâmetros foram obtidas para $0 \leq t \leq 2$ (Figuras 4 e 5). Em razão dos pequenos valores do parâmetro α, o número esperado de novos genes foi praticamente linear como uma função de *t* com o sequenciamento de novas *EST* e decaimento acentuado para estimativas não-paramétricas para t>1. Essa questão foi discutida recentemente por Susko & Roger (2004) ao considerarem a predição gênica em uma amostra adicional de *EST* maior que a amostra inicial requer ajustes paramétricos para a distribuição da abundância dos transcritos para evitar variabilidade do estimador.

Em RES, a distribuição dos transcritos em grupos indica que 619 genes aparecem uma única vez e um gene foi representado 31 vezes no total de 820 genes. Como $\Delta(t)$ representa a soma dos índices de *t*, para *t*>1 as contribuições dominantes para a soma virão de n_x para um grande valor de *x*, por exemplo, se *t* =2 e a maioria dos genes aparecer com menos de 10 transcritos; mas, se um gene aparecer 30 vezes, a contribuição para a soma desse gene será -2^{30} e $\Delta(t)$ estimado será, possivelmente, negativo. Se outro *read* for obtido e corresponder ao gene representado 30 vezes, a contribuição seria 2^{31} resultando em uma enorme mudança ($2^{30} + 2^{31}$) no $\Delta(t)$ estimado. Para bibliotecas não-uniformes, em que alguns grandes grupos de transcritos podem ser esperados, $\Delta(t)$ estimado é altamente variável e instável para *t*>1.

Conclusões

Em comparação à biblioteca SUS gênica, a RES apresentou maior redundância para menor número de *EST* o que indica, possivelmente, erros de agrupamento neste conjunto de dados.

O número esperado de novos genes é praticamente linear e com inclinação menos acentuada para RES, o que indica menor probabilidade de descoberta gênica em amostragem futura.

Referências

ADAMS, M.D.; KELLEY J.M.; GOCAYNE, J.D.; DUBNICK, M.; POLYMEROPOULOS, M.H.; XIAO, H.; MERRIL, C.R.; WU, A.; OLDE, B.; MORENO, R.F.; KERLAVAGE, A.R.; MCCOMBIE, W.R.; VENTER, J.C. Complementary DNA sequencing: expressed sequence tags and human genome project. **Science**, v.252, p.1651-1656, 1991.

EFRON, B.; THISTED, R. Estimating the number of unseen species: How many words did Shakespeare know? **Biometrika**, v.63, p.435-447, 1976.

EMRICH, S.; BARBAZUK W.; LI, L.; SCHNABLE, P. Gene discovery and annotation using LCM-454 transcriptome sequencing. **Genome Research**, v.17, p.69–73, 2007

FISHER, R.A.; CORBET, A.S.; WILLIANS, C.B. The relation between the number of species and the number of individuals in a random sample of an animal population. **Journal Animal Ecology**, v.12, p.42–58, 1943.

HUANG, X.; MADAN, A. Cap 3: a DNA sequence assembly program. **Genome Research**, v.9, p.868-877, 1999.

OKUBO, K., HORI, N., MATOBA, R., NIIYAMA, T., FUKUSHIMA, A. et al. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. **Nature. Genetics**. v.2, p**.**173–179, 1992.

O' SULLIVAN, D.B.; O'GARA, F. Traits of fluorescent Pseudomonas spp. involved in supression of plant root pathogens. **Microbiology Review***,* v.56, p.662-676, 1992.

SUSKO, E.; ROGER, A.J. Estimating and comparing the rates of gene discovery and expressed sequence tag (*EST*) frequencies in *EST* surveys. **Bioinformatics**. v.20, p.2279–2287, 2004.

WANG, J.P.Z.; LINDSAY, B.G.; LEEBENSMACK, J.; CUI, L.; WALL, P.K.; WEBB, C.M.; DE PAMPHILIS C.W. *EST* clustering error evaluation and correction. **Bioinformatics***,* v.20, p.2973-2984, 2004.

ANEXO 2. Código-fonte do programa escrito em linguagem R para estimativas paramétrica e não-paramétrica referentes à cobertura e ao número esperado de novos genes, segundo a metodologia descrita por Susko & Roger (2004).

```
a=read.table("c:/carlosreal2.txt")
##ÍNDICE 1 É SUS###
##ÍNDICE 2 É RES###

###Coverage - biblioteca individual###

L1=a$V1*a$V2
L2=a$V1*a$V3
C1=(sum(L1)-a$V2[1])/sum(L1)
C2=(sum(L2)-a$V3[1])/sum(L2)

C1
C2

###Erro-padrão coverage - biblioteca individual###

se_C1=(sqrt(sum(L1))^-1)*sqrt((a$V2[1]/sum(L1))+(2*a$V2[2]/sum(L1))-((a$V2[1]/sum(L1))^2))
se_C2=(sqrt(sum(L2))^-1)*sqrt((a$V3[1]/sum(L2))+(2*a$V3[2]/sum(L2))-((a$V3[1]/sum(L2))^2))

se_C1
se_C2

alfa=0.05

Li_ic_C1= C1 - qnorm((1-alfa/2))*se_C1
Ls_ic_C1= C1 + qnorm((1-alfa/2))*se_C1

Li_ic_C1
Ls_ic_C1

Li_ic_C2= C2 - qnorm((1-alfa/2))*se_C2
Ls_ic_C2= C2 + qnorm((1-alfa/2))*se_C2

Li_ic_C2
Ls_ic_C2

###Número esperado de genes - biblioteca individual###

eta1=1/(1-C1)
eta2=1/(1-C2)

eta1
eta2

###Erro-padrão para o número esperado de genes - biblioteca individual###

se_eta1=se_C1/(1-C1)
se_eta2=se_C2/(1-C2)

se_eta1
se_eta2
```

```
Li_ic_eta1= eta1 - qnorm((1-alfa/2))*se_eta1
Ls_ic_eta1= eta1 + qnorm((1-alfa/2))*se_eta1

Li_ic_eta1
Ls_ic_eta1

Li_ic_eta2= eta2 - qnorm((1-alfa/2))*se_eta2
Ls_ic_eta2= eta2 + qnorm((1-alfa/2))*se_eta2

Li_ic_eta2
Ls_ic_eta2

a=read.table("c:/carlosreal2.txt")

######método não paramétrico - biblioteca SUS#########

t=seq(0.01,2,0.01)
del0=matrix(0,12,length(t))
eta=a$V2
eta1=819
for(x in 1:12)
{
  for(i in 1:length(t))
  {
   del0[x,i]=((-1)^(x+1))*(t[i]^x)*eta[x]
  }
}
s=matrix(0,length(t),1)
for(i in 1:length(t))
  {
   s[i]=sum(del0[,i])
  }
plot(t,s,ylim=c(0,1500),type="l",lty=3)#gráfico do não paramétrico#
alfa1=-0.7115
gama1= 1.0586
delta_lib1=eta1*(alfa1^-1)*(gama1^-1)*(1-((1+gama1*t)^-alfa1))#gráfico do paramétrico - idem
            susko Fig2. pag 2283#
lines(t,delta_lib1,type="l")

######método não paramétrico - biblioteca RES#########

tr=seq(0.01,2,0.01)
del0r=matrix(0,10,length(t))
etar=a$V3
eta1r=619
for(x in 1:10)
{
  for(i in 1:length(tr))
  {
   del0r[x,i]=((-1)^(x+1))*(tr[i]^x)*etar[x]
  }
}
sr=matrix(0,length(tr),1)
for(i in 1:length(tr))
  {
   sr[i]=sum(del0r[,i])
  }
plot(tr,sr,ylim=c(0,1500),type="l",lty=3)#gráfico do não paramétrico#
alfa2=-0.4944
gama2= 0.8549
```

```
delta_lib2=eta1r*(alfa2^-1)*(gama2^-1)*(1-((1+gama2*t)^-alfa2))#gráfico do paramétrico - idem
            susko Fig2. pag 2283#
lines(t,delta_lib2,type="l")

############plotando as duas bibliotecas - não-paramétricas############

plot(t,s,ylim=c(0,1000),type="l",lty=3,col="red")#gráfico do não paramétrico sus#
lines(tr,sr,ylim=c(0,1000),type="l",lty=3)#gráfico do não paramétrico res#

t=seq(0.01,1,0.01)

#####método paramétrico - - biblioteca RES##############

k <- c(1,2,3,4,5,6,7,8,9,10,11,13,31)
Nk <- c(619,135,29,17,5,6,5,0,0,1,2,0,1)
sum(Nk)

p1=(k*Nk)/sum(k*Nk)
d1=data.frame(cbind(p1,k))

fit1=nls(p1 ~ ((gamma(k + b1)/(factorial(k)*gamma(1 + b1))))*(b2^(k-1)), start = list(b1=-
            0.87,b2=1.5),data=d1)
par1=coef(fit1)
alfa1=-0.4944
gama1= 0.8549
eta1=619

deltat1=eta1*(1-((1+gama1*t)^-alfa1))/(alfa1*gama1)
plot(1000*t,deltat1,xlim=c(0,1000),ylim=c(0,400),type="l")

 ######intervalo confiança paramétrico - biblioteca RES########
n=1000
t=seq(0.01,1,0.01)
part1=matrix(0,10,length(t))
part2=matrix(0,10,length(t))
part3=matrix(0,10,length(t))
for(x in 1:7)
{
  for(i in 1:length(t))
  {
   part1[x,i]=(t[i]^(2*x))*Nk[x]
   part2[x,i]=((-1)^(x+1))*(t[i]^(x))*((x*Nk[x])-(x+1)*Nk[x+1])
   part3[x,i]=(Nk[x]*(-1)^(x))*(1-2*((1+t[i])^x)+((1+2*t[i])^x))
  }
}
part2=part2[1:9,]

es1=matrix(0,length(t),1)
es2=matrix(0,length(t),1)
es3=matrix(0,length(t),1)
for(i in 1:length(t))
  {
   es1[i]=sum(part1[,i])
   es2[i]=sum(part2[,i])
   es3[i]=sum(part3[,i])
  }
sig11=(es1-es3)/10
sig12=sqrt(1/n)*sqrt(1/7)*es2
var1=sig11 - sig12^2
linf1=deltat1-1.96*sqrt(var1)
```

```
lsup1=deltat1+1.96*sqrt(var1)

plot(t*n,deltat1,xlim=c(0,1000),ylim=c(0,400),type="l")
lines(t*n,linf1,xlim=c(0,1000),ylim=c(0,400),type="l")
lines(t*n,lsup1,xlim=c(0,1000),ylim=c(0,400),type="l")

#####método paramétrico - biblioteca SUS##############

k1=c(1,2,3,4,5,6,7,8,9,10,11,13,31)
Nk1=c(819,108,21,19,8,7,3,3,2,2,1,1,0)
sum(Nk1)

p2=(k1*Nk1)/sum(k1*Nk1)
d2=data.frame(cbind(p2,k1))

fit2=nls(p2 ~ (gamma(k1 + b1)/(factorial(k1)*gamma(1 + b1)))*(b2^(k1-1)), start = list(b1=-
          0.87,b2=1.92),data=d2)

par2=coef(fit2)
alfa2=-0.7115
gama2= 1.0586
eta2=819

deltat2=eta2*(1-((1+gama2*t)^-alfa2))/(alfa2*gama2)

plot(1000*t,deltat2,xlim=c(0,1000),ylim=c(0,400),type="l",lty=3) ###pontilhado é library 2####
lines(1000*t,deltat1,xlim=c(0,1000),ylim=c(0,400),type="l")  ######idem ao gráfico fig.1  pag.
          2282################

  ######intervalo confiança paramétrico - biblioteca SUS########
n=1000
t=seq(0.01,1,0.01)
part1=matrix(0,12,length(t))
part2=matrix(0,12,length(t))
part3=matrix(0,12,length(t))
for(x in 1:12)
{
  for(i in 1:length(t))
  {
   part1[x,i]=(t[i]^(2*x))*Nk1[x]
   part2[x,i]=((-1)^(x+1))*(t[i]^(x))*((x*Nk1[x])-(x+1)*Nk1[x+1])
   part3[x,i]=(Nk1[x]*(-1)^(x))*(1-2*((1+t[i])^x)+((1+2*t[i])^x))
  }
}
part2=part2[1:11,]

es1=matrix(0,length(t),1)
es2=matrix(0,length(t),1)
es3=matrix(0,length(t),1)
for(i in 1:length(t))
  {
   es1[i]=sum(part1[,i])
   es2[i]=sum(part2[,i])
   es3[i]=sum(part3[,i])
  }
sig11=(es1-es3)/12
sig12=sqrt(1/n)*sqrt(1/12)*es2
var2=sig11 - sig12^2
linf2=deltat2-1.96*sqrt(var1)
lsup2=deltat2+1.96*sqrt(var1)
```

```
plot(t*n,deltat2,xlim=c(0,1000),ylim=c(0,400),type="l")
lines(t*n,linf2,xlim=c(0,1000),ylim=c(0,400),type="l")
lines(t*n,lsup2,xlim=c(0,1000),ylim=c(0,400),type="l")
```

Capítulo 4

Differential expression of calcium-binding proteins (TPT1 and S100A7) and calcium channel protein (TRPV6) in F_2 cattle infested with the tick *Riphicephalus (Boophilus) microplus*

Carlos Souza do Nascimento[1], Marco Antonio Machado[2,3], Simone Eliza Facioni Guimarães[1] Marta Fonseca Martins Guimarães[2], Jane de Oliveira Peixoto[1], John Furlong[2], Marcia CA Prata[2], Rui da Silva Verneque[2], Roberto Luis Teodoro[2] , Paulo Sávio Lopes[1]

[1]Department of Animal Science, Federal University of Viçosa - UFV, *Campus Universitário, 36570-000, Viçosa- MG. Brazil*
[2]Embrapa - Brazilian Dairy Cattle Research Center, *Eugênio do Nascimento, 610, Dom Bosco, 36038-330, Juiz de Fora- MG. Brazil*
[3]*Corresponding author. E-mail: machado@cnpgl.embrapa.br*

Abstract

The co-existence of Zebu animals with the tick *Riphicephalus (Boophilus) microplus* over the centuries seems to have conferred greater resistance to these animals compared to Taurine animals. Little is known so far about the genetic mechanisms involved in the genetic resistance of Zebu animals to ticks. The identification of new genes and host antigens involved in the mechanism of resistance/susceptibility to the parasite are a promising approach. In this work, we compared the relative expression of susceptible and resistant animals groups using real-time polymerase chain reaction to determine the expression level of the calcium-binding proteins translationally-controlled tumor protein 1 (TPT1) and allergen Bos d3 (S100A7), and of the calcium channel protein transient receptor potential vanilloid 6 (TRPV6). The three genes were identified in cDNA libraries prepared from skin lesions of susceptible animals and from healthy skin of resistant animals. Skin biopsies were obtained from F_2 cattle previously phenotyped to the tick *R. microplus* based on the number of engorged female ticks and classified as resistant or susceptible. The relative expression of the S100A7, TPT1 and TRV6 genes was 2.01 (±0.6), 1.32 (±0.9) and 1.53 (±1.2) fold higher in the susceptible group, respectively. Our findings show that the S100A7, TPT1 and TRPV6 genes were differentially expressed ($p=0.001$) in skin lesions from susceptible animals. In the present study, some highly expression genes related to skin hypersensitivity could be identified on

susceptible group. However, this hypersensitivity does not seem to protect susceptible animals against tick infestation

Keywords: molecular genetics, qRT-PCR, tick resistance

Introduction

The co-existence of Zebu animals with the tick *Riphicephalus (Boophilus) microplus* over the centuries seems to have conferred greater resistance to these animals compared to Taurine animals (Villares, 1941; Utech et al., 1978). Genetic resistance to the bovine tick *R. microplus* is one of the most important factors to reduce costs related to the control of this parasite in tropical cattle systems. Little is known so far about the genetic mechanisms involved in the genetic resistance of Zebu animals to ticks. However, studies regarding this genetic resistance may contribute to the biological control of ticks, since the use of resistant animals is one of the most effective solutions to control this parasite. Genetic resistance can contribute to decrease the tick population, costs with medications, personnel, mortality, and production losses and also to decrease environment pollution and intoxication of the animal. Development of new control methods is of great importance for cattle breeding programs. The identification of new genes and host antigens involved in the mechanism of resistance/susceptibility to the parasite are a promising approach.

In cattle, one of the main defense mechanisms of resistant animals is self-grooming. Rocha (1976) observed that the host directs its defenses to the region affected by the parasite, trying to eliminate the larvae by licking and rubbing the area against rough surfaces. Translationally controlled tumor protein (TPT1), also known as IgE-dependent histamine-releasing factor, is a growth-related tumor protein. TPT1 is involved in both cell growth and human late allergy reaction, as well as having a calcium binding property; however, its primary biological functions remain to be clearly elucidated (MacDonald et al. 2001).

The gene S100A7 encodes a low molecular weight calcium-binding protein responsible for the chemotactic migration of CD4+ lymphocytes to the affected region The S100A7 or allergen Bos d3 gene is involved in calcium-dependent signal transduction and is related to cellular changes in response to extracellular stimuli. Works suggest that the S100A7 gene promotes an increase in leukocyte recruitment to the affected region and thus contributes to proinflammatory activity (Jinquan et al., 1996).

The TRPV6 gene encodes a transmembrane protein responsible for the absorption of calcium. Neutrophils are the first defense cells arriving at the affected site, followed by eosinophils and basophils. In granulocytic neutrophils, important cellular responses are mediated or essentially regulated by the concentration of free Ca^{2+}, including the production and release of arachidonic acid, neutrophil degranulation and respiratory stress (release of superperoxide anions and other free radicals). In addition, Ca^{2+} is important for chemotaxis, particularly for the activation of integrins permitting strong adhesion to the blood vessel endothelium, and for integrin recycling during neutrophil migration (Heiner et al., 2003).

Based on the above described functions of these genes associated with immune response and other important functions, it is necessary to determine gene expression on skin tissue. However, until today the functions of these genes has not been reported in cattle yet. This work aimed to validate the differential expression of three genes identified in cDNA skin libraries from resistant and susceptible animals to *R. microplus,* using real-time polymerase chain reaction (qRT-PCR).

MATERIAL AND METHODS

Animals and tissue collecting

The animals used in this study belonged to an F_2 population originated from crossing F_1 animals (½ Holstein: ½ Gir) to determine QTLs (Martinez et al., 2006). On the basis of the results of previously performed genetic evaluations (Teodoro et al., 2006), animals with extreme breeding values for tick resistance/susceptibility were selected for the experimental groups. Six tick-resistant and six tick-susceptible F_2 animals were selected based on their breeding value for the sample collection.

To guarantee that they were free of natural infestation before being submitted to experimental infestation, the animals were bathed in acaricides. After, the animals were kept in a picket free of ticks until full clearance of the chemical acaricides and submitted to artificial infestation with 10,000 tick larvae. Skin biopsies were obtained from the lesion area provoked by the ticks in

susceptible animals and from the healthy region of resistant animals after artificial infestation.

RNA extraction

Two pools were prepared for the extraction of total RNA, one consisting of skin biopsies from resistant animals and the other of skin biopsies from susceptible animals. Skin biopsies were removed from RNAlater (Ambion) solution, sliced into several pieces and excess RNAlater removed. Total RNA was extracted using the RNeasy® Midi kit (Qiagen, Valencia, CA) according to manufacturer instructions. A DNase step (RNase-free DNase I, Invitrogen Life Technologies, Carlsbad, CA) was performed to the total RNA to remove genomic DNA. The quality of isolated RNA was determined spectrophotometrically (A260:A280) and its integrity was verified by 1,2% (w/v) agarose gel electrophoresis staining with ethidium bromide. RNA samples were stored at -70ºC until qRT-PCR analyses were performed.

Primer design and amplification efficiency

The genes used in the qRT-PCR reactions were selected from two skins cattles cDNA libraries infested with ticks (Nascimento CS et al., unpublished results). The primers used for amplification of the all genes and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as endogenous control were designed with the PrimerQuest program provided by Integrated DNA Technologies, Inc. (Coralville, IA, USA). The SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen, Carlsbad, CA, USA) was used to generate the first cDNA strand from total RNA. cDNA was synthesized in a 20 µl volume using 1 µg of total RNA. To confirm the absence of residual genomic DNA contamination, one sample was incubated without Reverse Transcriptase (negative control).

Before real-time quantification (qRT-PCR), the primers concentrations and cDNA quantity that provided the best efficiency of the PCR reaction for both the target genes and the endogenous control were established. Serial cDNA dilution curves were produced to calculate the amplification efficiency for all genes. A graph of threshold cycle (Ct) versus log10 relative copy number of the

sample from a dilution series was produced. The slope of the curve was used to determine the amplification efficiency (Pfaffl, 2001): Efficiency = 10 (-1/slope). At the end of the tests, the amount of 100 ng of cDNA and a primer concentration of 400 nM were standardized for the amplification of all genes. The dissociation curve of each gene was also analyzed.

Analysis of real-time PCR data

The qRT-PCR assays used the SYBR® Green I System (Bio-Rad, Hercules, CA, USA) for detection and were carried out in the SDS ABI PRISM 7000 thermocycler (Applied Biosystems, Foster City, CA, USA). mRNA levels were normalized against cattle GAPDH gene as endogenous control and compared between resistance and susceptible animals. Reactions for the endogenous control and target genes were carried out in duplicate. Each reaction mixture contained 12.5 µL 2X SYBR® Green Supermix, 400 nM of each primer, and 100 ng cDNA in a final volume of 25 µL. The amplification conditions were initial denaturation at 95ºC for 3 min, followed by 40 cycles of denaturation at 95 ºC for 15 s and annealing and extension at 60ºC for 60 s (for all primers). Amplicons were free from DNA contamination as judged by a lack of signal from non-reverse transcribed RNA with all primers sets. A mean cycle threshold (Ct) was established for the target genes and reference gene and ΔCt values were calculated [ΔCt = Ct (target gene) – Ct (GAPDH)]. Individual reactions were performed for each gene in 96-well plate. The relative expression analysis of the target gene was performed using a software REST (Pfaffl et al. 2002) for group-wise comparison in real-time PCR.

RESULTS

The primers used and the randomization tests of qRT-PCR are shown in Table 1 and Table 2, respectively. Efficiency values ranged from 1.65 to 2.23, indicating efficient amplification near the theoretical optimum level of 2 (Wilkening and Bader, 2004).

Table 1. Primer sequences used in the real-time PCR assays

Gene	Primer sequence (5'-3')	Amplicon size (bp)	Acession Number[1]
S100A7	F: TCA GCT TGA GCA GGC CAT TAC AGA R: ACA GGC ACT GAC GAA GTT GAG GAA	130	XM_870600.3
TPT1	F: GGG CTG CAG AAC AAA TCA AGC ACA R: ACA CCA TCC TCA CGG TAG TCC AAT	118	XM_001254326.1
TRPV6	F: ATC ATC GCT GCT TTG CTC ATG CTC R: AAC GTG GCA GCT TCT TCT CTA GCA	142	XM_866519.2
GAPDH	F: GGC GTG AAC CAC GAG AAG TAT AA R: CCC TCC ACG ATG CCA AAG T	99	NM_001034034.1

[1]http://www.ncbi.nlm.nih.gov/GenBank/

In the present study, three genes were analysed in the two groups of animals, resistant and susceptible to tick infestation,in skin tissue collected. Ct values in the two groups ranged from 16.3 to 28.3 for target genes and from 21.9 to 22.3 for GAPDH. The lower Ct value for S100A7 indicates that this gene reaches the detection threshold with less amplification cycles than GAPDH, suggesting that it is more abundant in the skin. On the other hand, the TPT1 and TRPV6 genes shown Ct values high than GAPDH indicant that these genes are less abundant. The coefficient of variation ranged from 0.52 to 3.43 for target genes and from 0.19 to 1.06 for GAPDH across all groups (Table 2).

Table 2: Output of randomization test in two groups and analysis of S100A7, TPT1, TRPV6 and GAPDH in skin tissue from resistant (R) and susceptible (S) animals.

Gene	(Ct) Threshold cycle	Standard deviation	Coeff. variation (%)
S100A7 (R)	17.7	0.43	3.43
S100A7 (S)	16.3	0.06	0.56
TPT1 (R)	27.1	0.10	0.52
TPT1 (S)	25.7	0,31	1.73
TRPV6 (R)	28.3	0.18	0.88
TRPV6 (S)	28.0	0.15	0.73
GAPDH (R)	21.9	0.17	1.06
GAPDH (S)	22.3	0.03	0.19

The expression levels of target genes and control GAPDH were different in resistant and susceptible animals and differentially expressed among groups ($p < 0.001$). REST® analysis showed over expression of S100A7, TPT1 and TRPV6 in skin biopsies from the susceptible group. S100A7 was classified as the more expressed gene of the group, followed by TPT1 and TRPV6. Expression of the S100A7 gene was higher in skin lesions from susceptible group 2.01 (±0.6). The ratios of the all genes expressions in the groups are summarized in Figure 1.

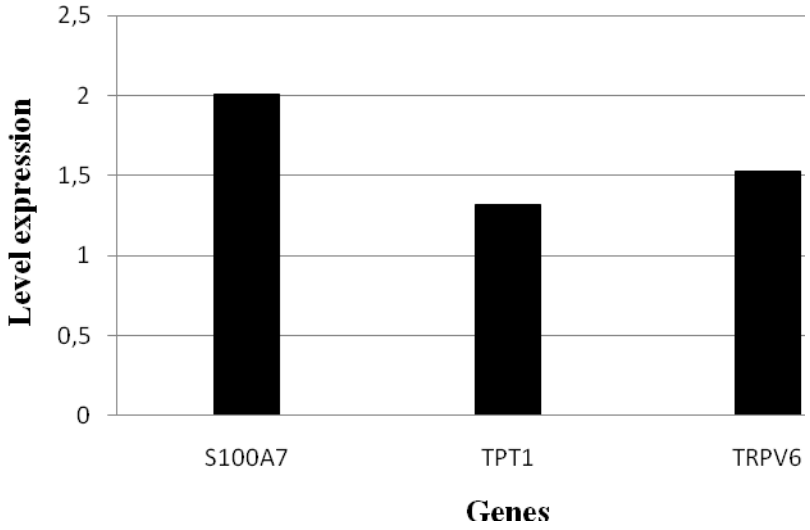


Figure 1. Relative expression measurements of the three genes using qRT-PCR calculated using REST2005 (relative expression software tool).
Relative expression of the S100A7 (2.01±0.6), TPT1 (1.32±0.9 and TRPV6 (1.53±1.2) genes in skin bovine from resistance and susceptible groups (n = 2).

Fold change (y axis) represents the relative expression of the S100A7, TPT1 and TRPV6 genes in comparison to resistance group, normalized by GAPDH reference gene expression. All genes were significant differentially expressed to $p < 0.001$.

DISCUSSION

Expression of TPT1, S100A7 and TRPV6, was detected in skin tissue samples originating from the susceptible and resistant tissues pools animals. These genes were previously identified as differentially expressed in the susceptible library (unpublished). So we performed qRT-PCR to validate the results obtained by analysis of the cDNA libraries.

Our results showed a higher expression of the histamine-releasing protein (TPT1) in skin lesions of susceptible animals. The increase was 1.32 (±0.9) fold to relative resistance group. One possible explanation for this finding might be the presence of antihistamine molecules or histamine-binding proteins in the tick saliva that are competing for histamine receptors, with a consequent reduction in the inflammatory response of the host to the tick bite (Paesen et al., 2000). Studies have demonstrated that the TPT1 gene encodes a calcium-binding protein that is induced in the presence of various intracellular stimuli and possess cytokine-like activities, inducing the production of interleukins by basophils and eosinophils and thus may affect host immune responses in patients with malaria (MacDonald et al., 2001). Degranulation of host mast cells promotes the release of histamine, which results in the dropping off the *Riphicephalus (Boophilus) microplus* tick larva (Kemp and Bourne, 1980), representing an important rejection factor in resistant animals. Schleger et al. (1981a) demonstrated a higher concentration of eosinophils and histamine close to the tick attachment site in highly resistant animals. Allen (1989), who observed higher skin hypersensitivity in the presence of basophils in the affected region, resulting in a possible increase of host irritation. Therefore, it is likely that TPT1 is involved in various inflammatory processes.

The basophils-cutaneous hypersensitivity is a type of immune response very frequent in animals exposed to ticks, characterized by the influx of basophils on the site of the bite with consequent release of histamine. The resistant animals develop this type of response more quickly than susceptible animals (Mattioli et al., 2000) with the detachment of ticks hours after fixing

them stemming the immunological events developed by the host (Ribeiro, 1989). As the animals could develop a later basophils-cutaneous hypersensitivity, the increased expression of gene TPT1 in these animals can be considered normal, whereas the time for collection of tissue for analysis of expression (5 and 12 days after the infestation) could not be ideal for detecting the expression of this genes in the resistant animals because they develop a rapid immune response, what it could recommended the analysis of gene expression in tissue collected hours after tick fixation. However, the time for collection of tissue provided in this work was clever to identify the increased expression of the gene TPT1 in animals susceptible because they have a slower immune response.

The gene S100A7 encodes a low molecular weight calcium-binding protein responsible for the chemotactic migration of CD4+ lymphocytes and neutrophils to the affected region (Jinquan et al., 1996). Our results show that the amplified products of the S100A7 gene started to be detected around cycle 17 before the other genes and even before the endogenous control suggesting that this gene is over expressed in this tissue.

S100A7 overexpression is seen in many epidermal inflammatory diseases and inflammatory lichen sclerosus and atrophicus (Madsen et al, 1991; Algermissen et al, 1996) and S100A7 expression is also increased in invasive skin cancers (Alowami et al, 2003). Di Nuzzo et al. (2000) reported increased expression of S100A7 in the human epidermis coincident with increases in adhesion protein (LFA1/ICAM-1) expression and it is has been associated with increased epidermal accumulation of CD4+ T cells in response to UV treatment. These studies suggest that S100A7 levels increase in response to inflammatory stress and that the S100A7 protein may function as a keratinocyte-derived chemotactic agent for immune cells. We hypothesized that the S100A7 gene can be to promote an increase in leukocyte recruitment to the affected region and thus contribute to proinflammatory activity on susceptible group.

Similarly, our results showed a 1.53 (±1.2) fold increase in the expression of the TRPV6 gene in skin lesions of susceptible animals. The TRPV6 gene encodes a transmembrane protein responsible for the absorption of calcium (Montell et al., 2002). The differential expression of TRPV6 in skin

lesions of animals infested with *R. microplus* agrees with the pattern expected for inflamed tissues, since the stimulus caused by the bite and the action of various substances present in the tick's saliva induce the continuous migration of neutrophils and other defense cells to the affected region, thus characterizing a proinflammatory state. Despite the significant difference between means expressions of the TRPV6 gene, our analysis is characterized by a high variation of the standard error to this gene. Considering that the skins pool analyzed in this study were performed from six animals, the broad variation in standard error could be due both to a composition of pools tissue with animals exhibiting different levels of resistance/susceptibility that could express at different levels the TRPV6 gene.

Our findings show the existence of differential expression in skin lesions of susceptible animals compared to healthy skin of animals highly resistant to the tick *R. microplus*. This is the first study showing the differential expression of genes encoding calcium-binding proteins and a calcium channel protein in cattle. Little is known how these genes participate in the mechanism of the host's response to the parasite. The above results indicated that overexpression of three genes may explain the skin hypersensitivity of susceptible animals, but this hypersensitivity does not seem to be effective in protecting susceptible animals against tick infestation. Other genes might be involved in the host defense mechanism. In addition, parasite-host interactions involve a complex network of gene activation mechanisms and little is known so far of how this mechanism works. Within this line of research, our team continues to investigate new genes and transcripts that are differentially expressed between resistant and susceptible animals. This established the primary foundation for further studies of these bovine genes.

Acknowledgements

This work receive financial support from the FINEP, CNPq, CAPES and FAPEMIG.

References

Algermissen B, Sitzmann J, LeMotte P and Czarnetzki B (1996). Differential expression of CRABP II, psoriasin and cytokeratin 1 mRNA in human skin diseases. *Arch. Dermatol. Res.* 288: 426–430.

Allen JR (1989). Immunology of interactions between ticks and laboratory animals. *Exp. App. Acarology* 7: 5-13.

Alowami S, Qing G, Emberley E, Snell L, et al. (2003). Psoriasin (S100A7) expression is altered during skin tumorigenesis. *BMC Dermatol.* 3: 1-7.

Di Nuzzo S, Sylva-Steenland RM, Koomen CW, de Rie MA, et al. (2000). Exposure to UVB induces accumulation of LFA-1+ T cells and enhanced expression of the chemokine psoriasin in normal human skin. *Photochem. Photobiol*. 72: 374–382.

Jinquan T, Vorum H, Larsen CG, Madsen P, et al. (1996). Psoriasin: a novel chemotactic protein. *J. Invest. Dermatol.* 107: 5-10.

Heiner I, Eisfeld J and Lückhoff A (2004). Role and regulation of TRP channels in neutrophil granulocytes. Cell Calcium 33 (5-6): 533–40.

Kang HS, Lee MJ, Song H, Han SH, et al. (2001). Molecular identification of IgE-dependent histamine-releasing factor as a B cell growth factor. *J. Immun.* 166: 6545–6554.

Kemp DH and Bourne A (1980) Boophilus microplus: the effect of histamine on the attachment of cattle tick larvae – studies in vivo and in vitro. *Parasitology* 80: 487-496.

Madsen P, Rasmussen HH, Leffers H, Honoré B, et al. (1991). Molecular cloning, occurrence, and expression of a novel partially secreted protein "psoriasin" that is highly up-regulated in psoriatic skin. *J. Invest. Dermatol*. 97: 701–712.

Martinez ML, Machado MA, Nascimento CS, Silva MVGB, et al. (2006). Association of BOLA-DRB3.2 alleles with tick (*Boophilus microplus*) resistance in cattle. *Gen. Mol. Res*. 5: 513-524.

MacDonald SM, Bhisutthibhan J, Shapiro TA, Rogerson SJ, et al. (2001). Immune mimicry in malaria: Plasmodium falciparum secretes a functional

histamine-releasing factor homologue in vitro and in vivo. *Proc. Natl. Acad. Sci. USA* 8, 10829–10832.

Mattioli RC, Pandey VS, Murray M and Fitzpatrick JL (2000). Immunogenetic Influences on Tick Resistance in African Cattle with Particular Reference to Trypanotolerant N'Dama (Bos taurus) and Trypanosusceptible Gobra zebu (Bos indicus) *Cattle. Acta Trop*. 75: 263–277.

Montell C, Birnbaumer L and Flockerzi V (2002). The TRP channels, a remarkably functional family. *Cell* 108: 595–598.

Paesen GC, Adams PL, Nuttall PA and Stuart DL (2000). Tick histamine-binding proteins: lipocalins with a second binding cavity. *Biochim. Bioph. Acta* 1482: 92–101.

Pfaffl MW, Horgan GW and Dempfle L (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res*. 30: e36.

Pfaffl MW (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 29(9): e45.

Ribeiro JM (1989). Role of saliva in tick/host interactions. *Exp. App. Acarol.* 7: 15–20.

Rocha UF (1976). *Panoramas da Parasitologia na África e na Austrália*.: Instituto de Biociências Biomédicas da USP. São Paulo, Brazil, pp.150.

Schleger AV, Lincoln DT and Kemp DH (1981a). A putative role for eosinophils in tick rejection. *Experientia* 37: 49-50.

Teodoro RL, Pires MFA, Furlong J, Prata MCA, et al. (2006). Estimativas de parâmetros genéticos e fenotípicos para a resistência a carrapatos e bernes em uma população experimental F_2 Holandês x Gir. In 'Anais da 43th Reunião Anual da Sociedade Brasileira de Zootecnia'. (Ed.SBZ) (CD ROM). (43th Reunião Anual da Sociedade Brasileira de Zootecnia: João Pessoa - PB: Brazil)

Utech KBW, Seifert GW and Wharton RH (1978). Breeding Illawarra Shorthorn cattle for resistance to Boophilus microplus. I. Factors affecting resistance. *Aust. J. Agric. Res*. 29: 411-422.

Villares JB (1941). Climatologia zootecnica III Contribuição ao estudo de resistência e susceptibilidade genética dos bovinos ao Boophilus microplus. In.: *Boletim da Indústria Animal*, Nova Serie, São Paulo 4: 70-86.

Wilkening S and Bader A (2004). Quantitative real-time polymerase chain reaction: Methodical analysis and mathematical model. *J. Biomol. Tech*. 15:107-111.

Internet Resources

PrimerQuest: http://www.idtdna.com/Scitools/Applications/PrimerQuest/.

GenBank: http://www.ncbi.nlm.nih.gov/GenBank/.

Capítulo 5

Differential expression of the cystatin 6-like cysteine proteinase inhibitor gene (CST6) in bovine's skin tissues infested with the tick *Riphicephalus (Boophilus) microplus*

Carlos Souza do Nascimento[1], Marco Antonio Machado[2*], Simone Eliza Facioni Guimarães[1], Marta Fonseca Martins Guimarães[2], Jane de Oliveira Peixoto[1], Paulo Sávio Lopes[1]

ABSTRACT – Cystatins are biochemically well-characterized as strong inhibitors of cysteine proteinases of the papain family, especially cathepsins and also of some lysosomal caspases. Genes encoding cysteine proteinase inhibitors have already been found in the *Ixodes scapularis sialoma* and in *Amblyomma americanum* facilitate the hematophagy behavior suppressing the antigens, processing and immune recognition of molecules of the tick saliva. The objective of the present study was to determine the differential expression profile of the CST6 bovine in skin biopsies from cattle infested with the tick *Riphicephalus (Boophilus) microplus*. Animals were previously phenotyped for resistance to the tick *R.microplus* according to the number of engorged female ticks and classified as resistant (RES) or susceptible (SUS). Two pools were prepared for the extraction of total RNA, one consisting of skin biopsies from resistant animals and the other of skin biopsies from susceptible animals. Real-time polymerase chain reaction (qRT-PCR) was used to detect the level of expression of the cystatin 6-like cysteine proteinase inhibitor (CST6) in skin biopsies of F_2 cattle. The level expression of the CST6 gene was 2.03 fold higher in the susceptible pool. The elevated level of cysteine proteinase inhibitor mRNA in skin lesions from susceptible animals might be explained by the constant turnover of epithelial proteins by endogenous and tick exogenous endopeptidases, promoting regeneration of damaged epithelial tissue

that characterizes the inflammatory state. This result suggests that CST6 gene transcripts can be associated with host suscepbility to *R. microplus* tick.

Keywords: animal breeding, molecular genetics, qRT-PCR

1 - Departmento de Zootecnia, Universidade Federal de Viçosa - UFV, 36570-000, Viçosa, MG, Brazil.
2 - Embrapa Gado de Leite, Rua Eugênio do Nascimento 610, Dom Bosco, 36038-330, Juiz de Fora, MG, Brazil.
*Send correspondence to Marco Antonio Machado. Empresa Brasileira de Pesquisa Agropecuária – Centro Nacional de Pesquisa em Gado de Leite, Rua Eugênio do Nascimento, 610, 36570-000, Dom Bosco, Juiz de Fora – MG, Brasil. Email: machado@cnpgl.embrapa.br

RESUMO – Cistatinas são fortes inibidores de proteinases de cisteína bem caracterizados bioquimicamente pertencentes da família das papaínas, especialmente catepsinas e também de algumas caspases lisossomais. Genes que codificam para inibidores de proteinase de cisteína têm sido reportados em *Ixodes scapularis sialoma* e em *Amblyomma americanum* facilitando a ação hematófaga do parasita suprimindo a apresentação de antígenos, processamento e reconhecimento de moléculas presentes na saliva do carrapato. O objetivo deste estudo foi determinar o nível de expressão gênica do CST6 na de pele de bovinos infestados com o carrapato *Riphicephalus (Boophilus) microplus*. Animais foram fenotipados previamente para o nível de resistência ao carrapato *R. microplus* quanto ao número de teológinas presentes e classificados como resistentes (RES) ou susceptíveis (SUS). Dois pools de tecidos foram preparados para extração de RNA, um consistindo de biópsias de pele de animais resistentes e outro de animais suscetíveis. A metodologia da reação em cadeia da polimerase semi-quantitativa em tempo real (qRT-PCR) foi usada para avaliar o nível de expressão do gene cistatina 6 (CST6) em biópsias de pele de bovino F_2. O nível de expressão dos transcritos do gene CST6 foi 2,3 vezes maior no *pool* SUS em relação ao *pool* RES. O elevado nível de mRNA do gene inibidor de proteinase de cisteína encontrado em lesões de pele dos animais susceptíveis pode ser explicado pelo constante *turnover* de proteínas do epitélio promovido pela ação de endopeptidases do animal e exopeptidases do parasita, promovendo a regeneração do tecido epitelial danificado caracterizado pelo estado inflamatório. Este resultado sugere que os transcritos do gene CST6 pode estar associado à susceptibilidade do hospedeiro ao carrapato *R. microplus*.

Palavras-chave: genética molecular, melhoramento animal, qRT-PCR

Introduction

Riphicephalus (Boophilus) microplus is a cattle ectoparasite found in tropical and subtropical regions worldwide (Willadsen and Jongejan, 1999). In Brazil, this tick represents a great problem for cattle production in different regions and the use of acaricides has been the most common prophylactic control measure against this ectoparasite. The main problems related to this practice are the development of resistant tick strains, the presence of chemical residues in animal products, and environmental pollution (Bullman *et al*., 1996).

Molecules released by the tick *R. microplus* during attachment and feeding of the larva on the host stimulate innate and acquired immune responses. The ability of the host to respond to these molecules will result in different levels of resistance. The tick counterattacks the host response with immunosuppressive molecules present in its saliva. The saliva of *R. microplus* contains substances that act on T lymphocytes, macrophages, neutrophils and natural defense cells, and influence the immunoglobulin class produced in response to the tick antigen. Particularly proteinases, peptidases and their inhibitors have been the focus of studies regarding ectoparasites. A recently published study has speculated regarding the use of proteinases or their inhibitors as a vaccine source, especially against tick serpins (Mulenga *et al*., 2001). Studies have indicated that cystatins exert important regulatory and protective functions against the uncontrolled proteolysis of cysteine proteinases of viral, bacterial and host origin (Bobek and Levine, 1992). In vertebrates, cystatins exert many specific functions include antigen apresentation antígeno (Honey e Rudensky, 2003), immune system development (Lombardi et al. 2005), epidermal homeostase (Reinheckel et al. 2005), extracellular matrix degradation (Serveau-Avesque et al. 2006)

In previous studies, *EST*s were generated from skin tissues of F_2 cattle (unpublished results) and comparative analysis were performed to detected expression profile of the transcripts. The objective of the present study was to validate the differential expression profile of the cystatin 6-like cysteine proteinase inhibitor (CST6) in skin biopsies from cattle infested with the tick *R. microplus*.

Materials and Methods

The animals used in this study belonged to an F_2 population originated from crossing F_1 animals (½ Holstein: ½ Gir) (Martinez et al., 2006). On the basis of the results of previously performed genetic evaluations, animals with extreme breeding values for tick resistance/susceptibility were selected for the experimental groups. Six tick-resistant and six tick-susceptible F_2 animals were selected based on their breeding value for the sample collection.

The selected animals were bathed in acaricides to guarantee that they were free of natural infestation before being submitted to experimental infestation. The animals were kept in a picket free of ticks until full clearance of the chemical acaricides, were inspected visually, and were then submitted to artificial infestation with 10,000 tick larvae. Skin biopsies were obtained from the lesion area provoked by the ticks in susceptible animals and from the healthy region of resistant animals after artificial infestation.

Two pools were prepared for the extraction of total RNA, one consisting of skin biopsies from resistant animals and the other of skin biopsies from susceptible animals. Skin biopsies were removed from RNAlater (Ambion) solution, sliced into several pieces and excess RNAlater removed. Total RNA was extracted using the RNeasy®

Midi kit (Qiagen, Valencia, CA) according to manufacturer instructions. A DNase step (RNase-free DNase I, Invitrogen Life Technologies, Carlsbad, CA) was performed to the total RNA to remove genomic DNA. The quality of isolated RNA was determined spectrophotometrically (A260:A280) and its integrity was verified by 1,2% (w/v) agarose gel electrophoresis staining with ethidium bromide. RNA samples were stored at -70ºC until qRT-PCR analyses were performed.

The genes used in the qRT-PCR reactions were selected from two skins cattles cDNA libraries infested with ticks (unpublished data).The primers used for amplification of the CST6 gene and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as endogenous control were designed with the PrimerQuest program provided by Integrated DNA Technologies, Inc. (Coralville, IA, USA) (Table 1). The SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen, Carlsbad, CA, USA) was used to generate the first cDNA strand from total RNA. cDNA was synthesized in a 20 µl volume using 1 µg of total RNA. To confirm the absence of residual genomic DNA contamination, one sample was incubated without Reverse Transcriptase (negative control).

Before real-time quantification, the primers concentrations and cDNA quantity that provided the best efficiency of the PCR reaction for both the target genes and the endogenous control were established. Serial cDNA dilution curves were produced to calculate the amplification efficiency for all genes. A graph of threshold cycle (Ct) versus log10 relative copy number of the sample from a dilution series was produced. The slope of the curve was used to determine the amplification efficiency (Pfaffl, 2001): Efficiency = 10 (-1/slope). At the end of the tests, the amount of 100 ng of cDNA and a primer concentration of 200 nM were standardized for the amplification of target gene and 400 nM for reference gene. The dissociation curve of each gene was also analyzed.

Table 1. Primer sets, concentration and amplicon sizes used in the real-time PCR assays

Gene symbol	Acession number[1]	Primer sequence (5'-3')	Concentration (nM)	Amplicon size (bp)
CST6	NM_001012764.1	F:AGA AGC TGC GCT GTG ACT TTG AGA	200	84
		R:ACA GGG ACA CAC AGT CGT GCT TTA	200	
GAPDH	NM_002046.3	F:GGC GTG AAC CAC GAG AAG TAT AA	400	99
		R:CCC TCC ACG ATG CCA AAG T	400	

[1]http://www.ncbi.nlm.nih.gov/sites/entrez.

The qRT-PCR assays used the SYBR® Green I System (Bio-Rad, Hercules, CA, USA) for detection and were carried out in the SDS ABI PRISM 7000 thermocycler (Applied Biosystems, Foster City, CA, USA). mRNA levels were normalized against cattle GAPDH gene as endogenous control and compared between resistance and susceptible animals. Reactions for the endogenous control and target genes were carried out in duplicate. Each reaction mixture contained 12.5 µl 2X SYBR® Green Supermix, 400 nM of each primer, and 100 ng cDNA in a final volume of 25 µl. The amplification conditions were initial denaturation at 95ºC for 3 min, followed by 40 cycles of denaturation at 95 ºC for 15 s and annealing and extension at 60ºC for 60 s (for all primers). Amplicons were free from DNA contamination as judged by a lack of signal from non-reverse transcribed RNA with all primers sets. A mean cycle threshold (Ct) was established for the target genes and reference gene and ΔCt values were calculated [ΔCt = Ct (target gene) – Ct (GAPDH)]. Individual reactions were performed for each gene in 96-well plate. The relative expression analysis of the target gene was performed using a software REST® (Pfaffl et al. 2002) for group-wise comparison in real-time

PCR. Efficiency values were 2.01 (GAPDH) and 2.11 (CST6) indicating efficient amplification near the theoretical optimum level of 2 (Wilkening and Bader, 2004).

Results and Discussion

Ct values in the two groups were 22.7 and 21.2 for target gene and 21.9 and 22.3 for GAPDH and the coefficient of variation ranged from 0.12 and 0.90 for target gene and 1.06 and 0.19 for GAPDH, to R and S groups, respectively (Table 2).

Table 2: Output of randomization test in two groups and analysis of CST6 and GAPDH in skin tissue from resistant (R) and susceptible (S) groups.

Gene	(Ct) Threshold cycle	Standard deviation	Coeff. variation (%)
CST (R)	22.7	0.02	0.12
CST6 (S)	21.2	0.13	0.90
GAPDH (R)	21.9	0.17	1.06
GAPDH (S)	22.3	0.03	0.19

The expression levels of target gene was different in resistant and susceptible groups and differentially expressed among groups ($p < 0.001$). Relative quantification by real-time PCR by REST® analysis revealed a 2.03-fold increase in the expression of the cysteine proteinase inhibitor (CST6) gene in lesions from group susceptible compared to healthy skin of resistance group. This gene already had been previously identified as differentially expressed in the susceptible library (unpublished data). So we performed qRT-PCR to validate the results obtained by analysis of the cDNA libraries.

One possible explanation for the high level of cysteine proteinase inhibitor transcripts in skin lesions of susceptible animals might be the constant turnover of epithelial proteins by endopeptidases and regeneration of damaged epithelial tissue that

characterizes the inflammatory state. The control of protein turnover by cystatins has been documented by Turk and Bode (1991). The epithelium serves as the first line of defense between the host and the environment. Disturbances in this barrier can lead to the invasion of microorganisms and pathogenic agents, causing subsequent inflammation. Therefore, the continuous stress to the epithelium induced by microorganisms or an inflammatory stimulus may result in the expression of proteins that protect against pathogenic agents and against the excessive proteolysis of endogenous and exogenous proteinases.

Kotsyfakis *et al*. (2006) recently described a cystatin, sialostatin L, in saliva which affected the proteolytic activity at sites of inflammation, exhibiting anti-inflammatory activity, and inhibited the proliferation of host cytotoxic T lymphocytes, thus contributing to the feeding success of the parasite. The authors concluded that the parasite escapes the host immune system by secreting cystatin, which disrupts antigen processing and presentation. In contrast, in mammals cystatin regulates antigen presentation by MHC class II molecules (Watts, 2001). Using serial analysis of gene expression in human keratinocyte cultures, Jansen *et al*. (2001) found high levels of expression of genes possibly involved in host protection and defense such as proteinase inhibitors and antimicrobial proteins. We therefore suggest that positive regulation of the host cystatin 6 gene transcripts may favor attachment of the tick in susceptible animals by indirectly inhibiting the action of cytotoxic T lymphocytes and, consequently, the production of host defense cytokines. Cystatin 6 probably acts on proteases, controlling the exaggerated proteolysis of proteins during regeneration of the epidermis and, at the same time, contributing to the feeding success of the parasite.

Conclusion

In this study, we were able to identify significant differential expression for the CST6 gene between tick-resistant and susceptible cDNA pools suggesting that this gene could be consider a tick-susceptible candidate gene. However, new expression studies are necessary to confirm these preliminary results and also to understand what could be the role of the CST6 gene on the mechanism of tick susceptibility and how much of the phenotypic variation is explained by this gene.

Acknowledgements

This study was supported by grants from FINEP, CNPq, CAPES and FAPEMIG.

Literature Cited

BOBEK, L.A.; LEVINE, M.J. Cystatins-inhibitors of cysteine proteinases. **Critical Reviews in Oral Biology & Medicine**, v.3, p.307-332, 1992.

BULLMAN, G.M.; MUÑOS ABENAS, M.E.; AMBRÚSTOLO, R.R. El impacto ecológico de las lactonas macrociclicas (endectocidas): una actualizacion compreensiva y comparativa. **Veterinária Argentina**, v.127, p.3-15, 1996.

HONEY, K.; RUDENSKY, A.Y. Lysosomal cysteine proteases regulate antigen presentation. **Nature Reviews Immunology**, v.3, p.472-482, 2003.

JANSEN, B.J.; Van RUISSEN, F.; De JONGH G. et al. Serial analysis of gene expression in differentiated cultures of human epidermal keratinocytes. **The Journal of Investigative Dermatology**, v.116, p.12-22, 2001.

KOTSYFAKIS, M.; SÁ-NUNES, A.; FRANCISCHETTI, I.M.B. et al. Antiinflammatory and immunosuppressive activity of sialostatin L, a salivary cystatin from the tick Ixodes scapularis. **Journal of Biological Chemistry**, v.281, p.26298-26307, 2006.

LOMBARDI, G.; BURZYN, D.; MUNDINANO, J. et al. Cathepsin-L influences the expression of extracellular matrix in lymphoid organs and plays a role in the regulation of thymic output and of peripheral T cell number. **Journal of Immunology**, v.174, p.7022-7032, 2005.

MARTINEZ, M.L.; MACHADO, M.A.; NASCIMENTO, C.S. et al. Association of BOLA-DRB3.2 alleles with tick (Boophilus microplus) resistance in cattle. **Genetics and Molecular Research**, v.5, p.513-524, 2006.

MULENGA, A.; SUGINO, M.; NAKAJIMA, M. et al. Tick-encoded serine proteinase inhibitors (serpins); potential target antigens for tick vaccine development. **Journal of Veterinary Medical Science**, v.63, p.1063-1069, 2001.

PFAFFL, M.W.; HORGAN, G.W.; DEMPFLE, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. **Nucleic Acids Research**, 30: e36, 2002.

PFAFFL, M.W. A new mathematical model for relative quantification in real-time RT-PCR. **Nucleic Acids Research**, 29(9): e45, 2001.

REINHECKEL, T.; HAGEMANN, S.; DOLLWET-MACK, S. et al. The lysosomal cysteine protease cathepsin L regulates keratinocyte proliferation by control of growth factor recycling. **Journal of Cell Science**, v.118, p.3387-3395, 2005.

SERVEAU-AVESQUE, C.; MARTINO, M.F.; HERVE-GREPINET, V. et al. Active cathepsins B, H, K, L and S in human inflammatory bronchoalveolar lavage fluids. **Biology of the Cell**, v.98, p.15-22, 2006.

TURK, V.; BODE, W. The cystatins: Protein inhibitors of cysteine proteinases. FEBS Letters, v.285, p.213–219, 1991.

WATTS, C. Antigen processing in the endocytic compartment. Current Opinion in **Immunology**, v.13, p.26–31, 2001.

WILLADSEN, P.; JONGEJAN, F. Immunology of the tick-host interaction and the control of ticks and tick-borne diseases. **Parasitology Today**, v.7, p.258-62, 1999.

Internet Resources

PrimerQuest program, www.idtdna.com/Scitools/Applications/PrimerQuest/ (December 5, 2006).

National center of Biotechnology Information (Entrez Gene), http://www.ncbi.nlm.nih.gov/ (November 26, 2006).

Capítulo 6

DISCUSSÃO GERAL

A variabilidade presente nas raças zebuínas e a detecção de um *major gene* para resistência ao carrapato em bovino sugerem o emprego de técnicas moleculares para identificação do gene e sua posterior transferência para as raças adaptadas a condições brasileiras, com o posterior aumento da produção e melhoria das condições gerais da pecuária. Esses genes podem ser manipulados para melhorar o desempenho e o bem-estar do animal. A genômica funcional tem o potencial de revelar novos genes candidatos visando melhorar os programas de seleção. Além disso, estudos funcionais contribuem para integrar informações fisiológicas e nutricionais com dados genéticos, com eventual ganho adicional no processo de seleção.

Sequências de *EST* têm sido adicionadas continuamente em bancos de dados especializados. Esses conjuntos de *EST* têm sido ampliados em larga escala e mais bibliotecas representando estádios de saúde ou de desenvolvimento são amostradas para identificar genes e verificar as vias metabólicas nas quais estão envolvidos. O volume de informações disponível para humanos e camundongos representa aproximadamente 21 vezes a informação existente para o genoma bovino.

Em bovinos, sequências derivadas dos transcritos têm sido organizadas no DFCI Cattle Gene Index (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=cattle) para prover informações úteis no desenho ou interpretação dos experimentos de genômica funcional. No entanto, mais *EST* são necessárias que as atuais ≈1.051.687 sequências para que seja possível identificar, anotar e classificar os genes específicos da espécie bovina. Maior número de informações sobre os transcritos é fundamental para interpretar resultados gerados em experimentos de genômica funcional, especialmente em virtude da pouca informação disponível sobre as sequências gênicas dos zebuínos (*B. indicus*). Essas raças dominam o mercado de produção bovina nas regiões tropicais e a geração de *EST* provenientes desses animais é importante para construir ferramentas de estudos genômicos que possam ser

utilizadas para caracterizar a diversidade genética entre os animais de origem taurina e zebuína, em relação à saúde e produção.

Para bovinos, o BodyMap-Xs disponibiliza informações de 1.331.319 *ESTs* oriundas de 399 bibliotecas de diferentes tecidos e, dessas bibliotecas, apenas duas são de pele e representam 35.211 transcritos. Desses transcritos, 29.518 foram originados de pele de embrião e apenas 5.693 *EST* foram geradas a partir de *pool* de pele de animais Hereford e Shorthorn infestados com o carrapato *Riphicephalus (Boophilus) microplus* (Wang et al., 2008).

No capítulo 2, os resultados da construção de duas bibliotecas de cDNA foram apresentadas. No total 4.070 clones foram produzidos, gerando 2.557 *EST* de qualidade. Essas sequências representam acréscimo de 7,3% ao total de transcritos de pele organizados BodyMap-Xs. Quando consideradas as *EST* derivadas de experimentos semelhantes com infestação com carrapatos, esse número representa 44,9% (5.693). Portanto, a geração dessas duas bibliotecas aumentará significativamente o número de sequências *EST* derivadas de pele de bovinos no bancos de dados.

As buscas de homologia em bases de dados revelaram 129 genes expressos de bovino e 661 transcritos tiveram homologia com outras espécies. Esses resultados indicam número reduzido de genes conhecidos em bovinos e foram obtidos pela atribuição automática de homologia em bases de dados. No entanto, esses genes precisam ser confirmados pela anotação manual. Atribuição automática de homologia pode estar sujeita a erros se parâmetros de buscas bem definidos não foram usados.

A atribuição funcional dos transcritos na base Gene Ontology (GO) revelou um padrão similar de atividade gênica em ambas as bibliotecas. As sequências únicas foram definidas em termos de processo biológico, componentes celulares e processo biológico. Por exemplo, a maioria dos transcritos foi associada a atividades catalíticas e de ligação. Esses resultados evidenciam que, para ambos os grupos de animais, resistentes e suscetíveis, processos fisiológicos semelhantes podem estar ocorrendo. O desafio com carrapatos pode não ter sido efetivo em discriminar genes envolvidos na resposta ao hospedeiro, o que pode ser exemplificado pelo reduzido número de transcritos envolvidos com resposta imune, consequentemente, genes pouco expressos ou raros podem não ter sido amostrados.

Os resultados para cobertura do número de genes amostrados indicam redundância de 49% na biblioteca RES e 40% na SUS. Ou seja, para cada 100 transcritos gerados, 49 (RES) e 40 (SUS) deles representam genes comuns. Essas elevadas taxas de redundância indicam que protocolos mais eficientes que favoreçam a identificação de genes raros devem ser usados. Sugerem ainda que a construção de bibliotecas normalizadas pode ser mais efetiva em detectar genes raros, considerando o experimento realizado.

No capítulo 3, analisaram-se a cobertura gênica e o número esperado de novos genes em uma amostragem futura. A qualidade dos agrupamentos das *EST* feitos pelo programa CAP3 foi avaliada com estimativas paramétricas, como descrito por Susko & Roger (2004), e comparadas a *EST* imadores não-paramétricos. As estimativas para cobertura prediz maior redundância em RES para menor número de *EST* geradas em relação a SUS.

Erros podem ser atribuídos a diferentes fontes, por exemplo, qualidade do sequenciamento e contaminação com organismos simbiontes presentes em amostras tomadas do ambiente. Foi investigada a possibilidade de ter ocorrido contaminação durante as amostras de pele em ambos os grupos de dados. Esse tipo de contaminação pode ocorrer em casos de existência de organismos que vivem relações simbiontes e que poderiam estar presentes na área de pele coletada. Consequentemente, sequências indesejadas desses organismos poderiam ser amostradas junto com as *EST* de bovinos. Para isso, analisou-se a contribuição de cada espécie nas análises de BlastN contra a base nr, conforme descrito no capítulo 2. Para RES, *Pseudomonas fluorescens* foi a principal espécie a contribuir para os genes anotados na base nr, seguida de *Bos taurus*. As análises do BlastN revelaram 161 genes (46 contigs e 115 singletons) com homologia com genoma de *Pseudomonas fluorescens* em RES. Para SUS, não foram observadas sequências contaminantes para essa espécie. Diante desses resultados, presume-se que a presença de sequências para essa espécie pode ter ocorrido durante o procedimento de coleta de tecido de pele nos animais resistentes, portanto, novos procedimentos de coleta devem ser considerados na coleta de amostras de tecido quando exposto a condições do ambiente.

No capítulo 4, objetivou-se validar a expressão diferencial de genes ligantes de cálcio (S100A7 e TPT1) e do canal de cálcio (TRP6). Esses genes mostraram ser mais expressos nos animais suscetíveis. Esses genes exemplificam categorias funcionais identificadas na atribuição funcional feita ao *Gene Ontology* (Captítulo 2). Termos funcionais foram altamente representados para a categoria ligante (Função molecular-S100A7). Os genes S100A7 e TPT1 são representantes de grupos de genes com alta representação para proteínas ligantes (função molecular).

As proteínas S100 agem como mediadoras do cálcio associado a transdução de sinal celular. Esse gene codifica para uma proteína de baixo peso molecular ligante de cálcio e responsável pela migração quimiostática de linfócitos CD4+ para a região afetada (Jinquan et al., 1996). Estudos funcionais têm mostrado que o gene TPT1 codifica para proteínas ligantes de cálcio que são induzidas na presença de vários estímulos dentro das células. O gene S100A7 age como mediador do cálcio associado com transdução de sinal e relaciona-se a mudanças celulares em resposta a estímulos extracelulares. Além disso, as proteínas TPT1 possuem atividades semelhantes à das citocinas, podendo induzir a produção de interleucinas dos basófilos e dos eosinófilos bem como ser induzidas por certas citocinas, agindo como fator de crescimento para as células B. Portanto, é provável que as TPT/HRF estejam envolvidas em vários processos inflamatórios.

Outro gene representativo da categoria funcional de atividade de transporte pela membrana (função molecular) foi TRPV6, recentemente renomeado ECaC2 (*Epithelial calcium channels 2*), que codifica para uma proteína transmembrana responsável pela absorção de cálcio pelo intestino (Montell et al., 2002). Essa categoria funcional foi amostrada unicamente no grupo SUS, o que sugere que mecanismos de transportes transmembrana são mais expressos em animais suscetíveis.

Em neutrófilos granulócitos, importantes respostas celulares são mediadas ou essencialmente reguladas pela concentração de Ca2+ livre, incluindo a produção e liberação de ácido araquidônico, degranulação dos neutrófilos e o estresse respiratório (liberação de ânions superperóxidos e outros radicais livres). Além disso, Ca2+ pode ser importante para a

quimiotaxia, sobretudo para ativação das integrinas, por permitir a firme adesão ao endotélio dos vasos sanguíneos e a reciclagem das integrinas na migração dos neutrófilos. Segundo Heiner et al. (2003), o Ca2+ contribui essencialmente para a função dos granulócitos durante sua defesa contra infecções por fungos e bactérias, promovendo migração das células de defesa para a região afetada e, por conseguinte, permitindo maior comunicação celular mediada por cálcio.

No capítulo 5, objetivou-se avaliar o perfil de expressão do gene inibidor de cisteíno proteinase semelhante à Cistatina 6 (CST6) presente em biópsias de pele de bovinos infestados por carrapato. Esse gene mostrou ser diferencialmente expresso no grupo de animais suscetíveis. Previamente, no capítulo 2, foi identificada representação diferencial de genes presentes na categoria processos metabólicos celulares (Processo Biológico) na biblioteca SUS (21) em relação ao grupo RES (10), o que pode indicar a existência de proteínas associadas ao *turnover* de moléculas causadas por apoptose celular. Isso pode ser atribuído à atividade hematófoga do parasita, que, ao se alimentar do sangue do hospedeiro, libera substâncias que provocam resposta imune do bovino, desencadeando uma cascata de eventos fisiológicos que culminam com a morte celular. Um estudo publicado recentemente tem especulado sobre o uso de proteinases ou seus inibidores como fonte de vacina, em especial para a serpina do carrapato (Mulenga et al., 2001). Estudos têm indicado que as cistatinas exercem importantes funções regulatórias e de proteção contra a proteólise não-controlada de cisteíno proteinases de origem viral, de bactérias e do próprio hospedeiro (Bobek e Levine, 1992).

CONSIDERAÇÕES FINAIS

O desenvolvimento deste projeto possibilitou a identificação de 4.070 *ESTs* (2.700 sequências únicas, que correspondem a 1.235 *ESTs* da biblioteca RES (700 sequências únicas) e 1.465 *ESTs* da biblioteca SUS (592 sequências únicas);

Os *singlets* e *contigs*, relativos às bibliotecas de cDNA de pele, foram submetidos à análise de similaridade em bancos de dados (BlastX – nr) e 790 apresentaram similaridade com outras proteínas já depositadas: 300 foram confirmadas pelo Swissprot e, destas, 11 sequências únicas tiveram similaridade com proteínas envolvidas no sistema imune;

Observou-se distribuição similar das *EST* das bibliotecas RES e SUS entre os termos do Gene Ontology (GO);

Na análise de expressão gênica, foram identificadas 54 sequências como transcritos diferencialmente expressos em ambas as bibliotecas, em relação ao genoma de *Bos taurus*;

Quatro genes (S100A7, TPT1, TRPV6 e CST6) foram analisados por Real Time – PCR e confirmados como diferencialmente expressos na biblioteca SUS em relação a RES;

Os dados gerados neste estudo poderão ser empregados em análises comparativas entre espécies.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDREOTTI, R.; GOMES, A.; MALAVAZI-PIZA, K.C.; SASAKI, S.D.; SAMPAIO, C.A.M.; TANAKA, A.S. BmTI antigens induce a bovine protective immune response against Boophilus microplus tick. *International Immuno-pharmacology* v.2, p.557-63, 2002.

BARKER, S.C.; MURRELL, A. Phylogeny, evolution and historical zoogeography of ticks: a review of recent progress. **Experimental Applied Acarology**, v.28, n.1-4, p.55-68, 2002.

BYFORD, I.; COLDITZ, P.; SIBBICK, R. A tick resistant dairy breed. **Queensland Agriclutural Journal**, v.102, n.1, p.11-15, 1976.

BOBEK, L.A.; Levine, M.J. Cystatins-inhibitors of cysteine proteinases. **Critical Reviews in Oral Biology & Medicine**, v.3, p.307-332, 1992.

BULLMAN, G.M.; MUÑOS ABENAS, M.E.; AMBRÚSTOLO, R.R. El impacto ecológico de las lactonas macrociclicas (endectocidas): una actualizacion compreensiva y comparativa. **Veterinária Argentina**, v.127, p.3-15, 1996.

BRASIL: Ministério da Agricultura, Pecuária e Abastecimento - Secretaria de Política Agrícola. *Estatísticas agrícolas*. Disponível em: <http://www.agricultura.gov.br/spa/pagespa/ch03/3_1.xls>. Acesso em: 07/02/2003

FRISCH, J.E.; O'NEILL, C.J.; KELLY, M.J. Using genetics to control cattle parasites - the Rockhampton experience. **International Journal Parasitology**, v.30, p.253-264, 2000.

FRISCH, J.E. Towards a permanent solution for controlling cattle ticks. **International Journal of Parasitology***,* v.29, n.1, p.57-71, 1999.

FRISCH, J.E. Identification of a major gene for resistance to cattle ticks. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVOCK PRODUCTION, 4., 1994, Guelf, Ontario. **Proceedings…** Guelf, 1994, p.293-295.

GRISI, L.; MASSARD, C.L.; BORJA, M.G.E.; PEREIRA, J.B. Impacto econômico das principais ectoparasitoses em bovinos no Brasil. **A Hora Veterinária**, v.21, n.125, p.8-10, 2002.

HAMMOND, K. Small Holder Milk Recording for Developing Countries. 29th Session. *International Committee for Animal Recording*. July, 31, August, 6, Ottawa, Canada, 1994.

HEINER, I.; EISFELD, J.; LÜCKHOFF, A. Role and regulation of TRP channels in neutrophil granulocytes. **Cell Calcium***,* v.33, n.6, p.533–40, 2004.

HEWETSON, R.W. Resistance of cattle to cattle tick *Boophilus microplus*. II. The inheritance of resistance to experimental infestation. **Journal of Agricultural Research**, v.19, p.495-505, 1968.

HORN, S.C.; ARTECHE, C.C.P. Situação parasitária da pecuária no Brasil. **A Hora Veterinária**, v.23, p.12-32, 1995.

HORN, S. C. Programa nacional de controle de parasitoses. In: *Curso de Parasitologia Animal*, v.2, p. 21-42, 1988.

JINQUAN, T.; VORUM, H.; LARSEN, C.G.; MADSEN, P.; RASMUSSEN, H. Psoriasin: a novel chemotactic protein. **Journal of Investigative Dermatology***,* v.107, p.5-10, 1996.

KASHINO, S.S.; RESENDE, J.; SACCO, A.M.S.; ROCHA, C.; PROENÇA, L. et al. Boophilus microplus: The pattern of bovine immunoglobulin isotype responses to high and low tick infestation. **Experimental Parasitology**, v.110, p.12-21, 2005

KERR, R.J.; FRISCH, J.E.; KINGHORN, B.P. Evidence for a major gene for tick resistance in cattle. *Proc 5th WCGALP,* **Guelph, Ontario**, v.20, p.265-268, 1994.

LEMOS, A.M.; TEODORO, R.L.; OLIVEIRA, G.P.; MADALENA, F.E. Compartive performance of six Holstein-Friesian x Guzera grades in Brazil. 3. Burdens of Boophilus microplus under field conditions. **Animal Production**, v.41, n.2, p.187-191, 1985.

MADALENA, F.E.; TEODORO, R.H.; LEMOS, A.M.; OLIVEIRA, G.P. Causes of variation of field burdens of cattle ticks (*B. microplus*). **Revista Brasileira de Genética**, v.8, n.2, p.361-375,1985.

MATTIOLI, R.C.; PANDEY, V.S.; MURRAY, M.; FITZPATRICK, J.L. Immunogenetic influences on tick resistance in african cattle with particular reference to trypanotolerant N'Dama (Bos taurus) and trypanosusceptible Gobra zebu (Bos indicus). **Acta Tropica**, v.75, n.3, p. 263-277, 2000.

MARTINEZ, M.L.; MACHADO, M.A.; NASCIMENTO, C.S.; SILVA, M.V.G.B.; TEODORO, R.L. et al. Association of BOLA-DRB3.2 alleles with tick (*Boophilus* microplus) resistance in cattle. **Genetics and Molecular Research**, v.5, p.513-524, 2006.

MARTINEZ, M.L.; SILVA, M.V.G.B.; MACHADO, M.A.; TEODORA, R.L.; VERNEQUE, R.S. A biologia molecular como aliada no combate aos carrapatos. In: SIMPÓSIO DA SOCIEDADE BRASILEIRA DE MELHORAMENTO ANIMAL, 5., 2004, Pirassununga, SP. **Anais...** Pirassununga, 2004. p.1-3.

MIYOSHI, T.; TSUJI, N.; ISLAM, K.M.; KAMIO, T.; FUJISAKI, K. Enzymatic characterization of a cubilin-related serine proteinase from the hard tick Haemaphysalis longicornis. **Journal of Veterinary Medical Science**, v.66, p.1195–1198, 2004.

MONTELL, C.; BIRNBAUMER, L.; FLOCKERZI, V. THE TRP channels, a remarkably functional family. **Cell**, v.108, p.595–598, 2002.

MULENGA, A.; MACALUSO, K.R.; SIMBER, J.A.; AZAD, A.F. The American dog tick encodes a functional histamine release factor homolog. **Insect biochemistry and insect molecular biology**, v.33, p.911–919, 2003.

MULENGA, A.; SUGINO, M.; NAKAJIMA, M.; SUGIMOTO, C.; ONUMA, M. Tick-encoded serine proteinase inhibitors (serpins): potential target antigens for tick vaccine development. **Journal of Veterinary Medical Science**, v.63, p.1063–1069, 2001.

OGASAWARA, O.; OTSUJI, M.; WATANABE, K.; IIZUKA, T.; TAMURA, T.; et al. BodyMap-Xs: Anatomical breakdown of 17 million animal *EST*s for cross-species comparison of gene expression. **Nucleic Acids Research**, v.34, p.D628–D631, 2006.

O'NEILL, C.J.; WELDON, G.J.; HILL, R.A.; THOMAS, J.E. Adaptaur Association of Australia; A breed association based on the performance recording of Adaptaur cattle. **Animal Production in Australian**, v.22, p.229-232, 1998.

RENDEL, J. Breeding for milk products in the tropics. In: INTERNATIONAL DAIRY CONGRESS. n.2, 1971. *Proceed*., 1971. 452p.

RIBEIRO, J.M.C.; MAKOUL, G.T.; LEVINE, J.; ROBINSON, D.R.; SPIELMAN, A. Antihemostatic, anti-inflamatory and immunosuppresive properties of the saliva of tick, Ixodess dammini. **Journal of Experimmental Medical**, v.161, p.332-344, 1985.

RIBEIRO, J.M.; WEIS, J.J.; TELFORD, S.R. Saliva of the tick Ixodes damminis neutrophil function. *Experimmental Parasitology*, v.74, p.382-388, 1990

SAUER, J.R.; MCSWAIN, J.L.; BOWMAN, A.S.; ESSENBERG, R.C. Tick salivary gland physiology. **Annu. Ver. Entomol.**, v.40, p.245-247, 1995.

SEIFERT, G.W.; SPRINGELL, P.H.; TATCHELL, R.J. Radioactive studies on the feeding of larvae, nymphs, and adult of the cattle tick, *Boophilus microplus* (Canestrini). **Parasitology***.* v.58, p.415-30, 1968.

SEIXAS, A.; DOS SANTOS, P.C.; VELLOSO, F.F.; DA SILVA VAZ, I.J.R.; MASUDA, A. et al. A Boophilus microplus vitellin-degrading cysteine endopeptidase. **Parasitology***,* v.126, p.155–163, 2003.

SUTHERST, R.W.; JONES, R.J.; SCHNITZERLING, H.J. Tropical legumes of the genus *Stylosanthes* immobilize and kill cattle ticks. **Nature**, v.295, p.320-1, 1982.

SUSKO, E.; ROGER, A.J. Estimating and comparing the rates of gene discovery and expressed sequence tag (*EST*) frequencies in *EST* surveys. **Bioinformatics**, v.20, p.2279–2287, 2004.

TEODORO, R.L.; LEMOS, A.M.; MOREIRA, D.P.; MADALENA, F.E. Resistência genética dos bovinos ao carrapato (Boophilus microplus). VII. Resistência de touros miços sob infestação artificial. In: *REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA*, 21, Belo Horizonte, MG, **Anais...** 1984c, p.54.

UTECH, K.B.W.; SEIFERT, G.W.; WHARTON, R.H. Breeding Australian Illawarra Shorthorn Cattle for Resistance to Boophiplus microplus. I. Factors Affecting Resistance. **Australian Journal Agricultural Research**, v.29, n.2, p.411-422, 1978.

VERÍSSIMO, C.J.; SILVA, R.G.; OLIVEIRA, AAD.; RIBEIRO, WR.; ROCHA, UF. Resistência e suscetibilidade de bovinos leiteiros mestiços ao carrapato *Boophilus microplus*. Boletim Indústria Animal, v.54, n.2, p.1-10, 1997.

VILLARES, J.B. Climatologia zootecnica III. Contribuição ao estudo de resistência e susceptibilidade genética dos bovinos ao Boophilus microplus. *In.: Boletim da Indústria Animal*, Nova Serie, São Paulo, n.1, v.4, p.70-86, 1941.

WANG, Y.H.; REVERTER, A.; KEMP, D.; MCWILLIAM, S.M.; INGHAM, A. et al. Gene expression profiling of Hereford Shorthorn cattle following challenge with Boophilus microplus tick larvae. **Australia Journal Experimental Agricultural**, v.47, p.1397-1407, 2008.

WANG, J.P.Z.; LINDSAY, B.G.; LEEBENSMACK, J.; CUI, L.; WALL, P.K.; WEBB, C.M.; DE PAMPHILIS C.W. *EST* clustering error evaluation and correction. **Bioinformatics**, v.20, p.2973-2984, 2004.

WILLADSEN, P.; JONGEJAN, F. Immunology of the tick-host interaction and the control of ticks and tick-borne diseases. **Parasitology Today**, v.7, p.258-262, 1999.

WHARTON, R.H.; UTECH, K.B.W.;TURNER, H.G. Resistance to the cattle tick, *Boophilus microplus*, in a herd of Australian Illawarra Shorthorn cattle: its assessment and heritability. **Aust. J. Agric. Res.**, v.21, p.163-181, 1970.

Endereços Eletrônicos

http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=cattle

http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html.

http://lifesciencedb.jp/bodymap-xs.

www.hgsc.bcm.tmc.edu/projects/bovine/