

Identificación inicial de genes en *Babesia bigemina* mediante análisis de Etiquetas de Secuencia Expresadas en el estadio intraeritrocítico del parásito

Babesia bigemina: Initial gene identification by Expressed Sequence Tags (EST) analysis of the intraerythrocytic stage

José Javier Pérez de la Rosa^{a,b}, Patricia Vargas Ureostegui^a, J. Antonio Álvarez Martínez^a, Carmen Rojas Martínez^a, Victor Manuel González Zuñiga^c, Julio V. Figueroa Millán^a

RESUMEN

En este estudio se realizó el análisis de Etiquetas de Secuencias Expresadas (EST) obtenidas a partir de 2208 clonas de *Escherichia coli*, con plásmidos recombinantes conteniendo insertos de cDNA de *Babesia bigemina*. Las secuencias se analizaron mediante búsqueda de homología en las bases de datos de genes. El análisis de homología en secuencia permitió identificar 470 clonas (agrupadas en 267 clusters) conteniendo EST con similitud de secuencia estadísticamente no significativa con algún gen de *Babesia* spp o de otro organismo Apicomplexa, sugiriendo la presencia de genes nuevos de *B. bigemina*; Se identificaron 21 clonas con EST correspondientes a 6 secuencias de genes previamente reportados para *B. bigemina*; además de 1285 clonas (conformando 159 clusters de genes distintos) de identidad significativa con proteínas hipotéticas o correspondientes a genes ya reportados en el genoma secuenciado de *Babesia bovis*; 32 clonas con EST homólogas a 16 genes distintos de *Theileria* spp; 51 clonas (26 genes distintos) con similitud en secuencia a genes de *Plasmodium* spp; 25 clonas con EST de moderada similitud con 13 genes distintos genes de *Toxoplasma gondii*; y 4 clonas con EST de mayor identidad con 4 genes diferentes de *Cryptosporidium* spp. Los resultados obtenidos permiten elaborar una base de datos sobre EST del estadio intraeritrocítico de *Babesia bigemina*, información básica esencial para la caracterización molecular del parásito, que permite llevar a cabo la identificación y regulación de nuevas regiones génicas codificadoras y, eventualmente el establecimiento de nuevas estrategias de control de la babesiosis bovina causada por *B. bigemina*.

PALABRAS CLAVE: *Babesia bigemina*, Etiquetas de Secuencia Expresadas (EST).

ABSTRACT

In this study, Expressed Sequence Tags (EST) were obtained and analyzed from 2208 randomly selected clones containing plasmids with cDNA inserts derived from a *Babesia bigemina* library. The obtained sequences were extracted and subject to Blast homology search in the Genbank databases. Sequence homology analysis resulted in the identification of 470 clones (grouped in 267 distinct clusters) which contained EST with no significant sequence identity with *Babesia* sp genes or other Apicomplexan parasites. Presumably, these EST would correspond either to new, unreported *B. bigemina* transcribed genes present in the erythrocyte stages of the parasite, or to non-translated sequences of the putative genes. 21 clones were identified which contained EST corresponding to 6 genes coding for *B. bigemina* antigens already reported in the literature; 1285 clones (grouped in 159 clusters of distinct sequences) had significant sequence identity with genes coding for hypothetical proteins previously identified in the *Babesia bovis* genome. Moreover, 32 clones had EST corresponding to 16 different *Theileria* sp. genes; 51 clones (26 distinct sequences) showed EST with sequence similarity to genes of *Plasmodium* sp., 25 EST had low identity with 13 different *Toxoplasma gondii* genes; and 4 clones with EST for 4 different *Cryptosporidium* sp genes. The results obtained, in addition to EST analysis of a larger number of *B. bigemina* cDNA clones, will allow the characterization and, eventually, the manipulation of gene coding regions, essential for the establishment of improved control strategies for cattle babesiosis caused by *B. bigemina*.

KEY WORDS: *Babesia bigemina*, Expressed Sequence Tags (EST) analysis.

INTRODUCCIÓN

Babesia bigemina es un protozooario intraeritrocítico causante de la babesiosis bovina, transmitida por garrapatas del género *Boophilus*⁽¹⁾ que afecta al ganado vacuno en regiones tropicales

INTRODUCTION

Babesia bigemina is an intraerythrocytic protozoan transmitted by ticks of the genus *Boophilus*⁽¹⁾, producing bovine babesiosis, which affects bovine cattle in subtropical and tropical areas worldwide,

^a Laboratorio de Babesia, Centro Nacional de Investigación Disciplinaria en Parasitología Veterinaria / INIFAP Km 11.5 Carretera Cuernavaca-Cuautla, Col. Progreso, Jiutepec, Morelos, C.P. 62550, MÉXICO. Tel +52 (777) 319-2850 ext 139. figueroa.julio@inifap.gob.mx Correspondencia al último autor.

^b Dirección Actual: Centro Nacional de Servicios de Constatación en Salud Animal / SENASICA, Km 11.5 Carretera Cuernavaca-Cuautla, Col. Progreso, Jiutepec, Morelos, C. P. 62550

^c Centro de Ciencias Genómicas, UNAM Campus Morelos. Av. Universidad s/n, Col. Chamilpa, Cuernavaca, Morelos, C.P. 62210, México

y subtropicales del mundo, con elevadas pérdidas económicas en la industria ganadera⁽²⁾. En México, los datos hasta 1980 mostraron que las pérdidas económicas por esta enfermedad ascendían a cerca de \$ 3,000 millones de pesos⁽³⁾. Además, se ha referido que hasta un 75 % de los bovinos del hato nacional estimado en 32 millones de cabezas contabilizados en el 2009⁽⁴⁾, estaban en riesgo de contraer babesiosis por estar en zonas endémicas donde se localiza la garrapata vector. Con la finalidad de controlar la babesiosis bovina se han instrumentado diversas estrategias de control entre las que destacan: el empleo de fármacos, baños garrapaticidas, control biológico contra el vector, y el diseño de vacunas utilizando regiones inmunogénicas del parásito a fin de conferir resistencia a la enfermedad.

La capacidad de los animales para recuperarse de una infección por *Babesia* y la efectividad de las vacunas vivas contra desafíos con aislados homólogos y heterólogos, demuestran que se puede alcanzar una inmunidad protectora en los animales en riesgo⁽⁵⁾. Los mecanismos de la respuesta inmune están dirigidos en mayor parte, contra antígenos parasitarios presentes en los merozoitos y los eritrocitos infectados, y se han identificado algunos antígenos de estos estadios mediante el uso de anticuerpos monoclonales^(6,7). Sin embargo, varios estudios han demostrado que cuando son utilizados como inmunógeno (en forma nativa o recombinante), estos componentes subunitarios inducen sólo una respuesta protectora parcial contra el desarrollo de la enfermedad^(8,9,10). Esto es debido, en gran parte, al reducido número de antígenos de los parásitos reconocidos y caracterizados hasta ahora, los cuales pudieran no ser los inmunógenos más adecuados.

A pesar de la aparente simplicidad para la propagación *in vitro* de *Babesia bigemina*⁽¹¹⁾, y que el sistema ha permitido obtener poblaciones de reducida virulencia o atenuadas⁽¹²⁾, hasta la fecha no se han realizado estudios de genética clásica o molecular en estas poblaciones del parásito. Esto se debe, en gran medida, a que existe un vacío de información en cuanto al número de secuencias génicas del parásito apropiadas para caracterización y manipulación. Una búsqueda reciente realizada en la base de datos del banco de genes "Genbank" (<http://www.ncbi.nlm.nih.gov>), demostró que éste contenía información de 185

and a source of serious losses in cattle rearing operations⁽²⁾. In México, data up to 1980 show that losses due to this disease added up to almost 3 billion pesos⁽³⁾. Besides, almost three quarters of the domestic herd, estimated at 32 million head in 2009⁽⁴⁾ were at risk of contracting this disease in tick endemic areas. Several strategies have been put in place for controlling babesiosis, among them, specific drugs, tick dipping vats, biological control of the vector and use of vaccines making use of immunogenic components of the parasite for conferring resistance to this illness.

Capacity of animals to recover from a *Babesia sp.* infection and effectiveness of live vaccines against challenges with homologous and heterologous isolates show that protective immunity can be achieved in animals at risk⁽⁵⁾. Immune response mechanisms are focused mainly against parasitic antigens present in infected merozoites and erythrocytes, and some of these antigens have been identified at these stages through monoclonal antibodies^(6,7). However, several studies show that when used as immunogens (either native or recombinant) these subunit components induce only a partial protective response against disease development^(8,9,10). This, to a great extent, is due to the comparatively low number of recognized and characterized antigens, which could not necessarily be the most adequate immunogens.

Despite the apparent simplicity for *Babesia bigemina in vitro* propagation⁽¹¹⁾ and to the fact that this method has allowed obtaining populations with low or attenuated virulence⁽¹²⁾, neither classical nor molecular genetic studies have been carried out on this species. This is due to a great extent to the nonexistence of satisfactory data on the number of gene sequences of the parasite for characterization and manipulation. A recent search in the "Genbank" (<http://www.ncbi.nlm.nih.gov>) shows that data on 185 *Babesia bigemina* sequences are available, of which 87 correspond to ribosomal RNA genes of several isolates, as well as 10 genes that codify for *Babesia bigemina* proteins, highlighting the presence of 56 coding sequences of proteins associated to rhoptries (RAP-1) from several *Babesia bigemina* geographical isolates⁽¹³⁾.

One of the most common approaches in DNA sequence databases have been those related to cDNA molecule sequencing for EST⁽¹⁴⁾ generation. Even when sequences are incomplete, it has been

secuencias de *B. bigemina*, de las cuales 87 correspondieron a los genes del RNA ribosomal de diversos aislados, así como secuencias de 10 genes que codifican por proteínas de *B. bigemina*, resaltando la presencia de 56 secuencias codificantes de las proteínas asociadas a las roptrias (denominadas RAP-1) de distintos aislados geográficos de *B. bigemina*⁽¹³⁾.

Uno de los abordajes más frecuentes en las bases de datos de secuencias de DNA, ha sido el relacionado con la secuenciación de moléculas de cDNA para generar EST⁽¹⁴⁾. Aun cuando las secuencias son incompletas, se argumenta que la generación de EST provee un método muy poderoso para el descubrimiento de nuevos genes, así como para la producción de sondas a utilizarse en estudios de mapeo. Los algoritmos actualmente disponibles y usados para comparar secuencias génicas y su correspondiente secuencia de aminoácidos, tales como Blast (búsqueda de alineamiento local básico, por sus siglas en inglés, <http://www.ncbi.nlm.nih.gov>) han permitido la identificación de un número significativo de secuencias homólogas, al comparar EST con secuencias completas disponibles en las bases de datos genómicos. Tales comparaciones han sido usadas, por ejemplo, para identificar secuencias presentes en diversos Phyla y que representan regiones de una proteína con estructuras y funciones altamente conservadas. Una área de estudio donde la secuenciación de EST probablemente tenga su mayor impacto, es en el descubrimiento de genes de organismos filogénicamente distantes. Tal es la situación de muchos de los organismos parasitarios, los cuales, a pesar de su importancia médica veterinaria, se encuentran un tanto rezagados en cuanto a la aplicación de la genética para resolución de problemas biológicos importantes. Se ha demostrado la factibilidad del análisis de EST para el descubrimiento de genes en organismos Apicomplexa tales como *Toxoplasma gondii*, *Plasmodium falciparum*, *Eimeria tenella*, *Sarcocystis neurona*, *Neospora caninum*, y más recientemente *Babesia bovis*⁽¹⁵⁻¹⁸⁾. Utilizando el programa Blast para realizar la anotación de secuencias mediante una comparación con las bases de datos existentes, 21, 20 y 32 % de las EST generadas para *B. bovis*, *T. gondii* y *P. falciparum*, respectivamente, pudieron

argued that EST provides a powerful method for discovering new genes, as well as for developing probes useful in mapping studies. Algorithms available at present and used for gene sequencing comparison and their corresponding amino acid sequence, such as Blast (Basic Local Alignment Search Tool, <http://www.ncbi.nlm.nih.gov>) have allowed identification of a significant number of homologous sequences when comparing ESTs to available complete sequences in gene banks. These comparisons, which represent highly conserved and structured areas of a protein, have been used for identifying sequences present in diverse Phyla. A study area where EST sequencing probably shows a great impact is gene discovery in organisms phylogenetically distant. This situation is common to many parasites wherein, despite their importance in veterinary medicine, the use of genetics for solving important biological problems has lagged. Feasibility of EST sequencing for gene discovery in Apicomplexa organisms, such as *Toxoplasma gondii*, *Plasmodium falciparum*, *Eimeria tenella*, *Sarcocystis neurona*, *Neospora caninum* and more recently, *Babesia bovis*, has been extensively confirmed⁽¹⁵⁻¹⁸⁾. By means of the Blast program for annotating sequences through comparison with current gene databases, 21, 20 and 32 % of EST generated for *B. bovis*, *T. gondii* and *P. falciparum*, respectively, were identified as homologous sequences with genes present in the genome of the three species⁽¹⁸⁾.

EST generation and analysis provide an option for identifying a great number of expressed genes of an organism at different stages of its life cycle. Strategies based on EST have been useful in parasitology for discovery of genes that code for biologically important proteins in parasites, including *B. bovis*^(17,18,19), thus allowing discovery of new drugs that have an effect on their life cycle. Besides, it has made possible identification of several antigens, which have become promising candidates to be included in vaccines against these parasites, providing significant results. Based on the above, the purpose of the present study was to start generating a gene database which should allow looking for and acquiring new expressed genes in the intraerythrocytic stage of *Babesia bigemina*, which in turn could help set up new control strategies for babesiosis, based on expressed gene sequences of the parasite.

identificarse como secuencias homólogas con genes conocidos en el genoma de las tres especies de parásitos Apicomplexa⁽¹⁸⁾.

La generación y análisis de EST proporcionan una opción para la identificación de un gran número de genes expresados de un organismo en los diferentes estadios de su ciclo de vida. Las estrategias basadas en EST han sido de utilidad en parasitología para el descubrimiento de genes que codifican proteínas de importancia biológica en los parásitos, incluyendo *B. bovis*^(17,18,19), lo que ha permitido el descubrimiento de nuevas drogas que afecten los mecanismos del ciclo de vida. Asimismo, han facilitado la identificación de varios antígenos que son candidatos para utilizarse como vacunas con resultados significativos a partir de EST. Basado en lo anterior, el objetivo de este trabajo es iniciar la generación de una base de datos que permita la búsqueda y conocimiento de nuevos genes expresados en el estadio intra-eritrocítico de *B. bigemina*, que permita el establecimiento de nuevas estrategias de control en la babesiosis bovina, basadas en el conocimiento de secuencias génicas expresadas por el parásito.

MATERIALES Y MÉTODOS

Crecimiento e infección de células hospederas Escherichia coli XL1-Blue

A partir de una genoteca de cDNA de *Babesia bigemina* construida en el bacteriófago λ ZAP⁽²⁰⁾, y con el objeto de obtener los fagémidos (en su forma de plásmido pBluescript) recombinantes insertados en el fago vector, se realizó una escisión masiva de las clonas recombinantes utilizando el fago auxiliar Exassist (Stratagene) y siguiendo la metodología descrita previamente⁽²¹⁾. Se utilizó una alícuota de la genoteca de cDNA, con un título de 1×10^4 unidades formadoras de placa, para realizar la infección de las células hospederas *E. coli* XL1-Blue en presencia de fago auxiliar, y propiciar la liberación del fagémido contenido dentro de cada fago λ ZAP recombinante. Posterior a una incubación por 15 min a 37 °C para permitir la adhesión de los fagos a las células hospederas, se adicionó medio de cultivo LB líquido conteniendo Isopropil-1-thio- β -D-galactopiranosido (IPTG) para inducir la expresión del gene de la β -galactosidasa, y la molécula reportera 5-bromo-4-cloro-3-indolil- β -D-galactopiranosido (X-gal) como substrato

MATERIALS AND METHODS

Cultivation and infection of Escherichia coli XL1-Blue host cells

From a *Babesia bigemina* cDNA gene library constructed in the λ ZAP bacteriophage⁽²⁰⁾ with the objective of obtaining recombinant phagemids (as pBluescript plasmids) inserted in the phage vector, a massive excision of recombinant clones by means of the Exassist (Stratagene) helper phage was performed following the previously described method⁽²¹⁾. An aliquot of the cDNA library, with 1×10^4 plaque forming units, was used to infect *Escherichia coli* XL1-Blue host cells in presence of the helper phage, to allow the liberation of the phagemids inside each λ ZAP recombinant phage. After incubation for 15 minutes at 37 °C for allowing attachment of phages to host cells, an LB liquid culture media containing isopropyl-1-thio- β -D-galactopiranoside (IPTG, to induce expression of the β -galactosidase gene) and the reporter molecule 5-bromo-4-chloro-indolyl- β -D-galactopyranoside (X-gal, as chromogenic substrate) was added. This mixture was sown in LB plates with ampicillin and incubated at 37 °C overnight.

Clone selection and plasmid purification for sequencing

The LB plates containing clones with recombinant plasmids (white colonies) were collected using sterile sticks and placed in 96-deep well plates containing 1 ml LB liquid media and incubated in presence of ampicillin at 37 °C and shaken overnight. The recombinant plasmids contained in each cultivated clone were purified through an alkaline lysis protocol using a commercial system (SV Plasmid DNA Purification Kit, Promega Co., Madison, WI, USA). The purified plasmid DNA was quantified through routine spectrophotometry⁽²²⁾.

Sequencing reaction

The plasmids obtained through the above mentioned method were used as template for the cycle sequencing method with terminal fluorescent di-deoxynucleotides (Applied Biosystems, Foster City, CA, USA), using T7 as sequencing primer. After the non incorporated di-deoxynucleotides purification process, capillary electrophoresis was carried out in an automatic sequencing analyzer (3130XL Applied Biosystems Sequencing

cromogénico. La mezcla se sembró en placas de LB con ampicilina e incubó a 37 °C durante la noche.

Selección de clonas y purificación de plásmidos para secuenciación

Las placas de LB agar conteniendo clonas con plásmidos recombinantes (colonias de color blanco) fueron colectadas con palillos estériles y depositadas en placas de 96 pozos elongados y conteniendo 1 ml de medio líquido LB e incubaron a 37 °C durante la noche con agitación y en presencia de ampicilina. Se purificaron los plásmidos recombinantes contenidos en cada una de las clonas cultivadas mediante un protocolo de lisis alcalina y utilizando un sistema comercial (SV Plasmid DNA Purification kit, Promega Co., Madison USA). El DNA plasmídico purificado fue cuantificado mediante técnicas espectrofotométricas de rutina⁽²²⁾.

Reacción de secuenciación

Los plásmidos obtenidos fueron utilizados como molde para el método de secuenciación en ciclo con terminación de didesoxinucleótidos fluoromarcados (Applied Biosystems, Foster City, CA USA), usando T7 como oligonucleótido iniciador. Posterior a la purificación de didesoxinucleótidos no incorporados se procedió a realizar la electroforesis capilar en un secuenciador automático (3130 XL Applied Biosystems Sequencing Analyzer).

Análisis de las secuencias

Las secuencias obtenidas fueron inicial e individualmente procesadas para evaluar su calidad mediante la aplicación Phred del software Sequencing Analysis versión 5.2 (Applied Biosystems), para posteriormente ser extraídas y analizadas mediante una búsqueda de homologías en secuencias con las bases de datos existentes en el *National Center for Biotechnology Information* (<http://www.ncbi.nlm.nih.gov/blast/>) para identificación de contaminación con secuencias de vector (vecscreen); secuencias del genoma de bovino (*Bos taurus*); y para encontrar identidades o similitudes con secuencias nucleotídicas (megablast) y con secuencias de proteínas (blastx), utilizando la aplicación blast⁽²³⁾.

RESULTADOS Y DISCUSIÓN

Con el propósito de obtener un número suficiente de clonas para analizar las secuencias de los

Analyzer).

Sequence analysis

Acquired sequences were initially and individually processed for assessing quality through the Phred Sequencing Analysis v. 5.2 software (Applied Biosystems), and later extracted and analyzed through a Blast search in the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/blast/>) for identification of contamination with vector sequences (vecscreen), bovine (*Bos taurus*) genome sequences and finding identities or similarities with nucleotide sequences (megablast) and protein sequences (blastx)⁽²³⁾.

RESULTS AND DISCUSSION

With the purpose of obtaining an adequate number of clones for analyzing sequences of *Babesia bigemina* cDNA inserts contained in pBluescript plasmids, a massive excision of recombinant phages contained in the available cDNA library was carried out, through recombinant/non recombinant colony screening in *Escherichia coli* XL-1 Blue host cells infected with the λZAP phage (pBluescript carrier) and propagated in a selective medium added with IPTG and X-gal. Screen analysis disclosed the presence of one white colony (recombinant) to every six blue (non recombinant) colonies. PCR analysis with T3 and T7 primers in 100 recombinant colonies selected at random, revealed the presence of cDNA inserts in the 250 - >3000 base pairs (bp) range, indicating heterogeneity in the size of transcripts copied to cDNA (not shown). However, and relative to the outcome of the present study, only readings of sequences derived from a single sequencing pass with primer T7 were used, an attribute that defines an EST⁽¹⁴⁾.

A grand total of 2,208 clones were processed for plasmid sequencing and generation of an EST database. The homology analysis performed on sequences acquired through both Megablast and Blastx shows that 320 clones were discarded because they contained plasmids belonging to either the vector phage (pBluescript), or *Escherichia coli* XL-1 Blue host cells, or did not provide enough plasmids for acquiring sequences of quality. A second group comprising 470 clones grouped in 267 clusters of differing genes contained EST with

insertos de cDNA de *B. bigemina* contenidos en los plásmidos pBluescript, se realizó una escisión masiva de fagos recombinantes contenidos en la genoteca de cDNA disponible, mediante el esquema de selección de colonias recombinantes/no recombinantes en células hospederas *E. coli* XL-1 Blue infectadas con el fago λ ZAP (portador de pBluescript), y sembradas en medio selectivo adicionado de IPTG y X-gal. El análisis de escrutinio reveló la presencia de una colonia blanca (recombinante) por cada 6 colonias azules (no recombinantes). El análisis por PCR con iniciadores T3 y T7 de 100 colonias recombinantes seleccionadas al azar, reveló la presencia de insertos de cDNA con tamaños dentro de un intervalo de 250 a >3000 pb, indicando la heterogeneidad en tamaño de los transcritos copiados a cDNA (no mostrado). Sin embargo, y para efectos de este estudio, sólo se utilizaron las lecturas de secuencia derivadas de un solo pase de secuenciación con T7 como iniciador, una característica que define a una EST⁽¹⁴⁾.

En total, se procesaron 2,208 clonas para secuenciación de sus plásmidos y generar la base de datos de EST. El análisis de homología en secuencias obtenidas mediante las aplicaciones Megablast y Blastx mostró un grupo de 320 clonas que fue descartado debido a que contenían secuencias correspondientes al fagémido vector (pBluescript), a la célula hospedera *E. coli* (cepa XL-1 Blue), o no proporcionaron suficiente cantidad de plásmido para obtener secuencias de calidad. Un segundo grupo comprendiendo 470 clonas (agrupadas en 267 *clusters* de genes distintos) contenían EST con tamaño promedio de 395 pb (intervalo de 57 a 1205 pb) con similitud de secuencia estadísticamente no significativa ($e \leq 1 \times 10^{-3}$) con algún gen de *Babesia* spp o de otro Apicomplexa (*Theileria*, *Plasmodium*, *Toxoplasma*, *Cryptosporidium*, etc.) y las cuales, presumiblemente, corresponderían a nuevos genes, aún no reportados, para *B. bigemina*. Un tercer grupo de 1,418 clonas (agrupadas en 224 *clusters* de genes distintos) contenían EST con tamaño promedio de 450 pb (intervalo de 95 a 1,239 pb) con similitud de secuencia estadísticamente significativa ($e \geq 1 \times 10^{-3}$) con genes de organismos Apicomplexa. Se destaca la identificación de 21 clonas que resultaron con EST correspondientes a 6 secuencias de genes previamente reportados para

a 395 bp average size (ranging between 57 and 1,205 bp), with non significant sequence similarity ($e \leq 1 \times 10^{-3}$) with genes of *Babesia* spp or other Apicomplexa organisms, which presumably belong to non reported genes. A third group comprising 1,418 clones grouped in 224 clusters of different genes, contained EST of 450 bp average size (95 - 1,239 bp) with significant ($e \geq 1 \times 10^{-3}$) sequence similarity with genes of Apicomplexa organisms. Identification of 6 clones showing EST corresponding to 6 genes previously reported for *Babesia bigemina* should be highlighted, and also of a large group made up by 1,285 clones, comprising 159 clusters of different genes which show greater and significant identity with hypothetical proteins and/or corresponding genes reported in the *B. bovis* genome⁽²⁴⁾. Besides, 32 clones with EST corresponding to 16 *Theileria* spp genes, 51 clones (corresponding to 26 different genes) with greater sequencing similarity to *Plasmodium* spp genes, a group of 25 clones with EST of moderate identity with 13 *Toxoplasma gondii* genes and a group of 4 clones with greater identity with 4 different *Cryptosporidium* spp genes were identified (Table 1).

Cuadro 1. Resumen del análisis de secuencias realizado a partir de 2,208 clonas

Table 1. Summary of a Sequence Analysis carried out from 2,208 clones

Identificación	No. de clonas (secuencias únicas)	Total (secuencias únicas)
Sin secuencia de DNA	92	
Sin inserto de cDNA	142	
Con inserto, DNA de <i>E. coli</i>	71	
Con inserto, DNA de fago vector	15	
EST con similitud de secuencia no significativa	470 (267)	470 (267)
EST con similitud significativa con genes de: <i>B. bigemina</i>	21 (6)	1306 (165)
<i>B. bovis</i>	1285 (159)	
<i>Theileria</i> spp	32 (16)	
<i>Plasmodium</i> spp	51 (26)	112 (59)
<i>Toxoplasma gondii</i>	25 (13)	
<i>Cryptosporidium</i> spp	4 (4)	
	2208	1888 (491)

B. bigemina; un gran grupo de 1,285 clonas (conformando 159 *clusters* de genes distintos) tuvieron una mayor y significativa identidad con proteínas hipotéticas o correspondientes a genes ya reportados en el genoma secuenciado de *Babesia bovis* (24). Se identificaron, además, 32 clonas con EST correspondientes a 16 distintos genes de *Theileria* spp., 51 clonas (correspondientes a 26 genes distintos) con mayor similitud en secuencia a genes de *Plasmodium* spp.; un grupo de 25 clonas con EST de moderada identidad con 13 distintos genes de *Toxoplasma gondii*; y un grupo de 4 clonas con EST de mayor identidad con 4 genes diferentes de *Cryptosporidium* spp (Cuadro 1).

En el Anexo 1 se resumen los genes identificados en orden alfabético, derivados del análisis de homología de secuencias utilizando el algoritmo Blastx y en el que se incluye información relacionada con: EST representativa de secuencia en el *cluster* o secuencia única; longitud de la EST; valor del resultado “S” (del inglés, *Score*) de alineamiento; valor esperado “E” (del inglés, *Expected*), que representa la significancia estadística del alineamiento; Porcentaje de identidad de secuencia; Identificación (anotación) del gene homólogo; y número de acceso para dicho gen en el banco de genes. La última columna indica el número de clonas cuya EST es idéntica o muy similar entre ellas, y que permite identificar un clúster de secuencias únicas o “hermanas” del gen identificado por homología. La herramienta Blast encuentra regiones de similitud local entre secuencias; de esta manera se comparan secuencias nucleotídicas o de proteínas y se calcula la significancia estadística de las similitudes o identidades de secuencia, lo cual puede ser utilizado para inferir relaciones funcionales y evolutivas entre secuencias, así como para ayudar a identificar los miembros de familias de genes. En general, entre más elevado sea el valor “S”, el alineamiento de secuencias es de mejor calidad, y entre más bajo sea el valor esperado “E”, la identidad de secuencia es más significativa, reflejando una probabilidad estadística elevada de que la EST analizada sea la correspondiente a un gen homólogo presente en otra especie u organismo. Vale la pena destacar que se presenta únicamente la información relativa a la mayor identidad en secuencia identificada para cada

In Attachment 1 appears a list in alphabetical order of identified genes, derived from a blastx homology search, showing data related to EST representative of sequences in clusters or of unique sequences, EST length, “S” score alignment, expected “E” value (representing alignment statistical significance), sequence identity percentage, annotation of homologous gene and access number for a gene in the Genbank. The last column shows the number of clones, whose EST is identical to or very close to others and which allows identification of unique sequences or “sisters” of a gene identified through homology. The Blast tool is able to identify areas of local similarity among sequences, and therefore nucleotide or protein sequences can be compared and the statistical significance of their similarity or identity can be estimated, which can be applied to infer both functional and evolutionary relationships between sequences, in addition to helping identify gene family members. As a rule, a higher “S” denotes greater quality sequence alignment, and a lower “E” stands for more significant sequence identity, indicating a greater probability that the EST being analyzed corresponds to a homologous gene present in another species or organism. It should be emphasized that only data relative to the highest identity in the identified sequence for each cluster or unique gene is shown. However, every analyzed unique gene or cluster shows homology, even with lower significance, with gene sequences present in *Cryptosporidium* spp, *Theileria* spp, *Plasmodium* spp and *Toxoplasma* spp, confirming sequence conservation among homologous genes (properly named as orthologous) of organisms belonging to the Phylum Apicomplexa.

Of the 224 clusters or EST groups identified through a Blast homology search as orthologous genes present in the intraerythrocytic stage of *Babesia bigemina*, 105 represent sequences corresponding to hypothetical genes, that is to say sequences which contain an open reading frame (ORF) and therefore a locus with its corresponding access number in the gene bank is given, but that has not been proven, either experimentally or functionally, that these gene regions effectively codify for a given protein. Of these, most of them (67 sequences) correspond to hypothetical genes identified in *B. bovis*, while the remaining 38 correspond to hypothetical genes of *Plasmodium*

cluster o gen único. No obstante, cada *cluster* o gen único analizado presenta homología, aun cuando presente menor significancia, con secuencias génicas correspondientes a los organismos *Cryptosporidium* spp., *Plasmodium* spp., *Theileria* spp., y *Toxoplasma* spp., confirmando la conservación de secuencias entre genes homólogos (propriadamente denominados como ortólogos) de los organismos pertenecientes al Phylum Apicomplexa.

Se destaca que de los 224 clústeres o grupos de EST que pudieron ser identificadas mediante el análisis Blast como genes ortólogos presentes en el estadio intraeritrocítico de *Babesia bigemina*, 105 de ellas representan secuencias correspondientes a genes hipotéticos, es decir secuencias que contienen un marco abierto de lectura y por lo tanto se les asigna un locus con su correspondiente número de acceso en el banco de genes, pero que no se ha comprobado, experimental ni funcionalmente, que efectivamente sean regiones de genes que codifican para una supuesta proteína. De éstas, la mayoría (67 secuencias) corresponden a genes hipotéticos identificados en *Babesia bovis*, mientras que las 38 EST restantes corresponden a genes hipotéticos de *Plasmodium* spp (18 EST), *Toxoplasma gondii* (10 EST), *Theileria* spp (8 EST) y *Cryptosporidium* spp (2 EST), especies pertenecientes al Phylum Apicomplexa, y por lo tanto se permite apreciar la conservación de genes entre organismos similares. Por otro lado, y a diferencia con un proyecto de secuenciación del genoma de un organismo, una colección de EST obtenida por secuenciación de forma aleatoria provee información funcional importante con respecto al nivel de expresión de un gen en particular⁽¹⁸⁾, ya que se ha demostrado que el número de EST presentes en un clúster corresponde proporcionalmente a los niveles de transcripción para un gen en particular, tales como los obtenidos mediante técnicas de hibridación de ácidos nucleicos en formato de micro-arreglos. Así, se resalta la elevada representatividad (redundancia) de EST cuyas secuencias en *B. bigemina* tienen mayor identidad con genes que codifican para proteínas identificadas en el genoma de *B. bovis*⁽²⁴⁾, como BBOV_III002570 (hipotética); BBOV_III004000 (hipotética); BBOV_IV003330 (subunidad 2 del factor 3 de inicio de la traducción); BBOV_I004210 (proteína

spp (18 EST), *Toxoplasma gondii* (10 EST), *Theileria* spp (8 EST) and *Cryptosporidium* spp (2 EST), species belonging to the Phylum Apicomplexa and thus allow to observe gene conservation among similar organisms. On the other hand, and in contrast to a genome sequencing project for a given organism, an EST collection obtained through random sequencing provides important functional data relative to the expression rate of a particular gene⁽¹⁸⁾, as it has been shown that the number of EST present in a cluster correspond proportionately to the transcription level for a particular gene, as those obtained through hybridization techniques of nucleic acids in micro-array format. Like so, the high representativeness (redundancy) of EST whose sequences in *B. bigemina* have greater identity with genes that codify for proteins identified in the *B. bovis* genome⁽²⁴⁾, such as BBOV_III002570 (hypothetical), BBOV_III004000 (hypothetical), BBOV_IV003330 (translation initiation factor 3 subunit 2), BBOV_I004210 (spherical body protein-3), BBOV_II002630 (hypothetical), BBOV_I002450 (hypothetical), BBOV_IV006360 (senescence-associated protein), BBOV_III000550 (60S ribosomal protein L15), BBOV_III10780 (hypothetical), BBOV_III010250 (eukaryotic translation initiation factor 4A), BBOV_IV003220 (HAD superfamily hydrolase), BBOV_III006650 (hypothetical), BBOV_II005360 (hypothetical), BBOV_III006700 (hypothetical), BBOV_007250 (tubuline, gamma chain), conforming groups of 113, 95, 88, 77, 67, 55, 41, 34, 33, 29, 22, 22, 21, 20, and 20 redundant EST for every cluster, respectively.

The cluster of greater abundance found (113 EST, 6 % of total analyzed EST) and identified as a group with predicted homologous sequences to hypothetical protein BBOV_III002570 was the result of greater similarity by the blastx analysis. Nevertheless, these same ESTs had a second match with a *Babesia bigemina* antigen known as Bbg 1.1 showing high statistical significance [$S=111$, $E=1 \times 10^{-5}$, identity=24/56 (42 %) and similarity=35/56 (62 %)], for which 12 EST were also specifically identified with high identity to antigen Bbg 1.1 (and with less similitude to Bbg 1.2).

Both antigens, Bbg 1.1 and Bbg 1.2, are apparently found in the intraerythrocytic stage of *B. bigemina*,

de cuerpo esférico-3); BBOV_II002630 (hipotética); BBOV_I002450 (hipotética); BBOV_IV006360 (proteína asociada a senectud); BBOV_III000550 (proteína ribosomal L15 de la unidad 60S); BBOV_III010780 (hipotética); BBOV_III010250 (factor 4A de inicio de la traducción); BBOV_IV003220 (hidrolasa de la superfamilia HAD); BBOV_III006650 (hipotética); BBOV_II005360 (hipotética); BBOV_III006700 (hipotética); BBOV_IV007250 (tubulina, cadena gamma); conformando grupos de 113, 95, 88, 77, 67, 55, 41, 34, 33, 29, 22, 22, 21, 20 y 20 EST redundantes por clúster, respectivamente.

El *cluster* de mayor abundancia encontrado (113 EST, 6% del total de EST analizadas) e identificado como grupo con secuencias predichas homólogas a la proteína hipotética BBOV_III002570, fue resultado de la mayor similitud arrojada por el análisis BLASTX. Sin embargo, estas mismas EST tuvieron un segundo encuentro de elevada significancia estadística [$S=111$, $E=1 \times 10^{-5}$, identidad=24/56 (42 %), y similitud=35/56 (62 %)], con el antígeno de *Babesia bigemina* denominado Bbg 1.1 y para el cual, también se identificaron más específicamente 12 EST con elevada identidad al antígeno denominado Bbg 1.1 (y con menor similitud a Bbg 1.2) de *B. bigemina*.

Los antígenos Bbg1.1 y Bbg1.2 aparentemente se encuentran en cantidades mucho más elevadas en el estadio intraeritrocítico de *B. bigemina*, como pudo evidenciarse por análisis de homología de secuencias, encontrándose hasta 125 EST (considerando las 113 EST identificadas como BBOV_I0002470). Sin embargo, esto se tendrá que demostrar de forma experimental en estudios posteriores. El antígeno de *B. bigemina* Bbg 1.1 (una proteína de ~34 KDa) ha sido clonado y caracterizado en trabajos previos⁽²⁵⁾, partiendo del tamizado de una genoteca de cDNA con antisuero policlonal de bovino preparado contra antígenos de *Babesia bigemina* separados por cromatografía. La secuenciación de DNA de dos clonas representativas permitió la identificación de marcos abiertos de lecturas, codificando polipéptidos que describen la región carboxi-terminal de dos diferentes antígenos, Bbg 1.1 y Bbg 1.2. Ambos polipéptidos contienen un motivo central en tándem de repetidos flanqueados por una región carboxi-terminal altamente conservada,

as evidenciado por análisis de secuencia homóloga, encontrando hasta 125 EST (tomando en cuenta las 113 identificadas como BBOV_I0002470). Sin embargo, esto tiene que ser confirmado en estudios posteriores. El antígeno Bbg 1.1 (una proteína de ~34 KDa) ha sido clonado y caracterizado en estudios previos⁽²⁵⁾, por el tamizado de una genoteca de cDNA con antisuero policlonal de bovino preparado contra antígenos de *Babesia bigemina* separados por cromatografía. El análisis de secuencia de dos clones representativos permitió la identificación de marcos abiertos de lecturas, codificando polipéptidos que describen la región carboxi-terminal de los dos antígenos, Bbg 1.1 y Bbg 1.2. Ambos polipéptidos contienen un motivo central de repetidos flanqueados por una región carboxi-terminal altamente conservada, pero las secuencias de los repetidos no poseen una gran homología. Los análisis de hibridación y de enzimas de restricción de clones de cDNA indican que estos antígenos pertenecen a una familia de al menos nueve genes relacionados, pero no idénticos, que conforman una familia cuyos antígenos son parte de la superficie del merozoito y participan en la respuesta inmune además de intervenir en la evasión inmune. Así, probablemente, las EST expresadas en altos rangos identificadas en el presente estudio corresponden a al menos 2 de los genes descritos para la familia de antígenos presente en la superficie del merozoito. Por lo tanto, y derivado de su alta antigenicidad y su abundante expresión en el estadio intraeritrocítico de *B. bigemina*, el antígeno Bbg 1.1 (y eventualmente la familia de antígenos) resulta ser el candidato ideal para caracterizar tanto inmunológicamente como molecularmente en términos de su conservación a nivel de secuencia de nucleótidos y de epítopos predictivos en diferentes aislamientos de *B. bigemina* de diferentes ubicaciones geográficas en México y para ser considerado como candidato para ser evaluado en estudios de protección inmune.

El segundo *cluster* en orden de abundancia de expresión identificado en el presente estudio como BBOV_III004000 (95 EST, 5% de las EST analizadas) también muestra una alta identidad de secuencia [$S=724$, $E=1 \times 10^{-76}$, identidad=133/213 (62 %), Similaridad=165/213 (77 %)] con una proteína altamente conservada presente en *Theileria parva* (XP_763039.1) y otros parásitos del complejo Apicomplexa (*Theileria annulata*, *Cryptosporidium parvum*, *Plasmodium falciparum*, *Toxoplasma gondii*, etc.).

pero las secuencias rio arriba de los repetidos no tienen elevada homología. Ensayos de hibridación y análisis mediante enzimas de restricción de clones de cDNA indicaron que estas proteínas antigénicas pertenecen a una familia de por lo menos nueve genes relacionados, pero no idénticos, conformando una familia de antígenos que son parte de la superficie del merozoito y que participan en la respuesta inmunológica, además de intervenir en la evasión de la respuesta inmune. Por ello, es probable que las EST expresadas en elevados niveles e identificadas en este estudio correspondan a por lo menos dos de los genes descritos para la familia de antígenos presentes en la superficie del merozoito. Por lo tanto, derivado de su naturaleza altamente antigénica y dada su abundante expresión en el estadio intraeritrocítico de *B. bigemina*, el antígeno Bbg 1.1 (y eventualmente la familia correspondiente) resulta candidato ideal para caracterizar molecular e inmunológicamente en términos de su conservación, a nivel secuencia de nucleótidos y epítopes predictivos, en diferentes aislados de *B. bigemina* de distinto origen geográfico en México y ser considerado como candidato a evaluarse en estudios de inmunoprotección.

El segundo *cluster* expresado más abundantemente e identificado en este trabajo como BBOV_III004000 (95 EST, 5 % del total de EST analizadas), presenta también encuentros de elevada identidad de secuencia [$S=724$, $E=1 \times 10^{-76}$, Identidad= $133/213$ (62 %), Similitud= $165/213$ (77 %)], con una proteína altamente conservada en *Theileria parva* (XP_763039.1) y otros parásitos Apicomplexa (*Theileria annulata*, *Cryptosporidium parvum*, *Plasmodium falciparum*, *Toxoplasma gondii*, etc). La proteína de 281 aminoácidos ha sido anotada como una metil-transferasa dependiente de S-adenosil-metionina (AdoMet-MTasa), una clase de enzimas clasificadas con base en la especificidad de su sustrato (pequeñas moléculas, lípidos ácidos nucleicos, etc.) que transfiere grupos metilo de un compuesto a otro e involucrada en un gran número de procesos celulares (biosíntesis, transducción de señales, reparación de proteínas, regulación de la cromatina y silenciamiento de genes).

El tercer *cluster* más abundante (88 EST, 4.6 % del total de EST analizadas), es el representado por la proteína anotada como subunidad 2 del

This 281 amino acid protein has been annotated as adenosyl-methionine dependent methyl-transferase, (AdoMet-Mtase), a class of enzymes classified on the specificity of their substrate (small molecules, lipids, nucleic acids, etc.) that transfer methyl groups from one compound to another in a number of cellular processes (biosynthesis, signal transduction, protein repair, chromatin regulation and gene silencing).

The third cluster in EST abundance (88 EST, 4.6 % of all analyzed ESTs) is represented by the protein annotated as translation initiation factor 3 subunit 2 (TIF3s2), a 340 amino acid long protein [BBOV_IV003330: $S=428$, $E=1 \times 10^{-41}$, identity= $79/90$ (87 %)]. Like in the second cluster, this protein is highly conserved in Apicomplexa parasites and involved in protein biosynthesis from mRNA molecules. These subunits are necessary at initiation of translation by promoting bonding of the tertiary complex of initiation factor 2 to ribosomal subunit 40S. Protein TIF3s2 presents in addition a region containing a WD40 domain, which is also found in a number of proteins performing a wide variety of functions which include signal transduction, pre-RNA processing and assembly of cell skeleton.

The fourth most abundant EST cluster (77 EST, 4.0 % of total analyzed EST) is the one identified as homologous in *B. bigemina* of *B. bovis* spherical body protein-3 [BBOV_I004210: $S=700$, $E=1 \times 10^{-71}$, identity= $128/291$ (43 %)]. Unlike the previous 2 clusters, but akin the first cluster, this seems to be a specific protein in *Babesia sp.*, as the blast search did not cast any statistically significant match with sequences of other Apicomplexa organisms. Known as SBP in *B. bovis*, they are expressed in the spherical bodies of merozoites, organelles located in the apical region of intraerythrocytic parasites. It has been mentioned that SBPs take part in merozoite binding during erythrocyte invasion and modify the membrane of the infected cell. Three molecules whose cell location lies in the spherical bodies have been described, SBP-1 having an 87 kDa molecular weight, apparently involved in the erythrocyte infection process by acting on the parasitophorous vacuole⁽²⁶⁾. SBP-2 and SBP-3, 225 kDa and 135 kDa proteins, respectively, are also involved in the erythrocyte invasion process by inserting themselves on the red cell membrane^(28,27). Each of

factor 3 de inicio de la traducción (TIF3s2), una proteína de 340 aminoácidos de longitud [BBOV_IV003330: S=428, E=1x7⁻⁴¹, Identidad=79/90 (87 %)]. De forma similar a la anterior, la proteína se encuentra altamente conservada en parásitos Apicomplexa e involucrada en la biosíntesis de proteínas a partir de moléculas de mRNA. Estas subunidades son necesarias en la fase de inicio de la traducción, al promover la unión del complejo ternario del factor de iniciación 2 a la subunidad 40S ribosomal. La proteína TIF3s2 presenta además una región que contiene un dominio WD40. Dicho dominio se encuentra en un gran número de proteínas que realizan una amplia variedad de funciones que incluyen la transducción de señales, procesamiento de pre-RNA y el ensamblaje del citoesqueleto.

El cuarto *cluster* más abundante (77 EST, 4.0 % del total de EST analizadas) es el identificado como homólogo en *B. bigemina* de la proteína de cuerpo esférico 3 de *B. bovis* [BBOV_I004210: S=700, E=1x7⁻⁷¹, Identidad=128/291 (43 %)]. A diferencia de los dos *clusters* anteriores, pero similar al primer *cluster*, ésta parece ser una proteína específica en especies de *Babesia*, ya que el análisis Blast no arrojó encuentro de significancia estadística con secuencias de otro organismo Apicomplexa. Denominadas como SBP en *B. bovis*, se expresan en los cuerpos esféricos de los merozoitos, organelos localizados en la región apical de los parásitos intra-eritrocíticos. Se ha reportado que las SBPs participan en la unión del merozoito durante la invasión al eritrocito, y causa modificación de la membrana del eritrocito infectado. Se han descrito tres moléculas cuya localización celular radica en los cuerpos esféricos: SBP-1 con un peso molecular relativo de 87 KDa, aparentemente implicada en el proceso de infección al eritrocito al actuar sobre la vacuola parasitófora (26); SBP-2 y SBP-3, de 225 KDa y 135 KDa, respectivamente, que están implicadas en el proceso de invasión al insertarse sobre la membrana de los glóbulos rojos^(27,28). Estas proteínas están codificadas cada una por un gen unicopia, y son altamente conservadas entre cepas o aislados de *B. bovis*. La función y capacidad protectora de estas proteínas como antígeno vacunal aún no se ha determinado. Las EST identificadas en este estudio mostraron un porcentaje de identidad de ≥ 66 % con la Proteína

these proteins is coded by a single copy gene and is highly conserved among *B. bovis* strains and isolates. Protective functionality and capacity of these proteins as vaccine antigens has not yet been verified. EST identified in the present study show ≥ 63 % identity with spherical body protein 3 (SBP3) reported for *Babesia bovis*. This 135 kDa protein possesses an antigenic region or epitope conserved in *Babesia bovis* strains in Texas, Australia and México. This protein most probably is involved in *Babesia bovis* intracellular survival, growth and development. The gene that codes for SBP3 has been identified and sequenced by screening a *Babesia bovis* cDNA library with monoclonal antibodies⁽²⁷⁾. Because in the present study abundant transcription of the orthologous gene present in *Babesia bigemina* is reported for the first time, the coded antigen should be taken into account with the purpose of investigating and studying its antigenic functionality and assessing the immune mechanisms developed in the bovine host, because apparently, antigens of this type are involved in the binding and invasion process of merozoites to bovine erythrocytes, and hence should be considered as viable candidates to further studies in *Babesia bigemina*.

In subsequent clusters with abundant EST representation in *B. bigemina* (hypothetical proteins BBOV_II002630; BBOV_I002450; BBOV_III010780; BBOV_III006650; BBOV_II005360; BBOV_III006700; BBOV_III006200 and BBOV_IV000390) no statistically significant similarity with other Apicomplexa parasites was found, from which a molecular function or structure could be inferred. Analogous domain search or with sequences of other prokaryotes and eukaryotes could shed more light on the possible structure and function of proteins identified as hypothetical.

Within the group of ESTs of *B. bigemina* present in moderate abundance the following stand out: BBOV_IV006360 (41 EST), annotated as a hypothetical senescence associated protein, also conserved in *Cryptosporidium parvum*, *Plasmodium yoelli* and *Toxoplasma gondii*, not yet functionally characterized; BBOV_III000550 (34 ESTs), 60S ribosomal protein L15; BBOV_III010250 (29 EST), translation initiation factor 4A, an ATP dependent protein, involved in RNA splicing, necessary for a variety of cellular

de Cuerpo Esférico 3 (SBP3) reportada para *Babesia bovis*. Esta proteína de 135 KDa tiene una región antigénica determinante o epítipo conservado en cepas de *Babesia bovis* de Texas, México y Australia. La proteína probablemente se encuentra implicada en la supervivencia intracelular, crecimiento y desarrollo de *Babesia bovis*. El gen que codifica para la proteína SBP3 ha sido ya identificado y secuenciado mediante tamizado de una genoteca de cDNA de *B. bovis* con anticuerpos monoclonales⁽²⁷⁾. Dado que en este estudio se reporta por primera vez la abundante transcripción del gen ortólogo presente en *B. bigemina*, el antígeno codificado debería ser considerado para investigar y estudiar su funcionalidad antigénica, con la finalidad de evaluar los mecanismos de inmunidad desarrollados en el hospedero bovino, ya que aparentemente este tipo de antígenos están involucrados en el proceso de unión e invasión de los merozoitos a los eritrocitos del bovino, por lo que se consideran candidatos viables a ser estudiados más detalladamente en *Babesia bigemina*.

De los subsecuentes *clusters* con abundante representación de EST en *B. bigemina* (proteínas hipotéticas BBOV_II002630; BBOV_I002450; BBOV_III010780; BBOV_III006650; BBOV_II005360; BBOV_III006700; BBOV_III006200 y BBOV_IV000390) no se encontró similitud estadísticamente significativa con secuencias de otros parásitos Apicomplexa, a partir de la cual se pudiera inferir una función o estructura molecular. La búsqueda de analogía por dominio o con secuencias de otros organismos procariontes o eucariontes pudiera dar mayor luz sobre la posible estructura/función del grupo de proteínas identificadas como hipotéticas.

Dentro del grupo de EST de *B. bigemina* presentes en moderada abundancia resaltan: BBOV_IV006360 (41 EST), anotada como una supuesta proteína asociada a senectud, también conservada en *Cryptosporidium parvum*, *Plasmodium yoelli* y *Toxoplasma gondii*, pero aún no caracterizada funcionalmente; BBOV_III000550 con 34 EST (proteína ribosomal L15 de la unidad 60S); BBOV_III010250 de 29 EST (factor 4A de inicio de la traducción, proteína dependiente de ATP involucrada en el desdoblamiento de RNA necesario en una variedad de procesos celulares

processes which include intron elimination, ribosomal biogenesis and RNA degradation; BBOV_IV007250 (22 EST), a hydrolase pertaining to superfamily HAD, which plays a role in xenobiotic detoxification or of products of metabolism; BBOV_IV007250 (20 ESTs), a structural, gamma chain, tubuline protein, essential for building cytoskeletons; BBOV_IV011580 (18 EST), helicase, with capacity to unthread nucleic acid duplexes with a directional polarity, using energy freed by triphosphate nucleotide hydrolysis for propelling itself along the DNA chain, unthreading the duplex during the process; BBOV_II006040 (16 EST), a nucleolar protein known as NOP5, involved in ribosomal biogenesis, ribosomal translation and structure; XP_001612931.1 (15 EST), a precursor protein of *Plasmodium vivax* synthetase III, an enzyme that initiates elongation of fatty acids; BBOV_II005900 (14 EST), factor MCM2, a triphosphate nucleoside hydrolase involved in a variety of processes including DNA replication; BBOV_II001650 (13 EST), 40S ribosomal protein S3, which forms part of the front region of subunit 40S and interacts with mRNA; BBOV_IV003810 (13 ESTs) an adenosine kinase associated to regulation of extracellular adenosine levels and intracellular adenylate level preservation, besides taking part in phosphorylation during metabolism of purines in *B. bovis* via the alternate pathway; gb|AF053002.1 (13 EST), a gene corresponding to apocytochrome b, which forms part of the enzyme complex involved in mitochondrial respiratory chain; BBOV_III011660 and BBOV_II005480, *B. bovis* hypothetical proteins; gb|AAA27790.1 (12 ESTs), Bbg 1.1 antigen of *B. bigemina*; BBOV_III002090 (12 EST), 40S ribosomal protein S30; BBOV_III008720 (12 EST) a *B. bovis* membrane protein and BBOV_III05740 (12 EST), a protein containing a kinase domain.

Other ESTs with antigenic relevance identified in *Babesia bigemina*

An EST whose sequence showed high homology with Apical Membrane Antigen 1 (AMA-1) was identified, which was originally reported as an antigen found in the asexual phase of the intraerythrocytic stage of *Plasmodium falciparum*. AMA-1 is a microneme protein secreted to the surface of *Plasmodium spp* merozoite and

que incluyen la eliminación de intrones, biogénesis ribosomal y degradación de RNA); BBOV_IV003220 de 22 EST (una hidrolasa de la superfamilia HAD, participa en la desintoxicación de xenobióticos o productos generados por el metabolismo); BBOV_IV007250 con 20 EST (tubulina, cadena gamma, proteína estructural fundamental en la conformación del citoesqueleto); BBOV_IV011580 de 18 EST (Helicasa, con la capacidad de deshebrar los duplexes de ácidos nucleicos con una polaridad direccional y utilizando la energía libre de la hidrólisis de nucleótido trifosfatos para impulsar su traslación lo largo del DNA, deshebrando la doble hélice durante el proceso); BBOV_II006040 de 16 EST, una proteína nucleolar denominada NOP5, implicada en biogénesis ribosomal (traducción y estructura ribosomal); XP_001612931.1 de 15 EST, una proteína precursora de la sintetasa III de *Plasmodium vivax* (enzima que inicia la elongación de ácidos grasos); BBOV_II005900 de 14 EST, factor MCM2, una nucleósido trifosfato hidrolasa involucrada en una variedad de procesos incluyendo la replicación de DNA; BBOV_II001650 de 13 EST, proteína ribosomal S3 de la subunidad 40S (S3 forma parte de la región delantera de la subunidad ribosomal 40S e interactúa con el mRNA); BBOV_IV003810 con 13 EST, una adenosina quinasa (asociada con la regulación de los niveles de adenosina extracelular y la preservación de los niveles intracelulares de adenilato. Participa además, en la fosforilación durante el metabolismo de las purinas en *B. bovis* en la vía alterna); gb|AF053002.1 de 13 EST, un gene correspondiente al apocitocromo b (que forman parte del complejo enzimático implicado en la cadena respiratoria mitocondrial); BBOV_III011660 y BBOV_II005480, proteínas hipotéticas de *B. bovis*; gb|AAA27790.1 con 12 EST, el antígeno Bbg 1.1 de *B. bigemina*; BBOV_III002090 de 12 EST, la proteína ribosomal S30 de la subunidad 40S; BBOV_III008720 de 12 EST, una proteína membranal de *B. bovis*; y BBOV_III005740 de 12 EST, una proteína que contiene un dominio de enzima quinasa.

Otras EST identificadas con relevancia antigénica en *B. bigemina*

Se identificó una EST cuya secuencia mostró una elevada homología con el denominado Antígeno 1

Toxoplasma gondii tachyzoite. Afterwards the orthologous gene was identified in an Israeli *B. bovis* isolate⁽²⁹⁾, and later the presence of AMA-1 was identified and annotated in the sequenced genome of the *B. bovis* Texas isolate⁽²⁴⁾; later, this gene was sequenced in five isolates of different geographical origins in Brazil⁽³⁰⁾, indicative of its high conservation. More recently, the sequences corresponding to *B. bigemina* AMA-1 in isolates from Italy and Argentina were added to the gene bank^(31,32). Immunofluorescence experiments carried out with AMA-1 anti-peptide antibodies show that the antigen is located in the apical half of the *B. bovis* merozoite. However, immunoelectrotransference and metabolic marker tests of the parasite show that this *B. bovis* AMA-1 82 kDa antigen is found at very low quantities in the merozoite⁽²⁹⁾. There exists the probability of a very low expression level of *B. bigemina* orthologous AMA-1 transcripts, as evidenced by the limited EST number identified in the present study. On the other hand, expression level of transcripts is not directly and necessarily related to the amount of translated product. For example, in genes that codify for metabolic enzymes, transcripts are widely found in genomes, while enzyme concentration in protein extracts is low, but there is high enzymatic activity. It remains unknown if *Babesia spp* AMA-1 forms part of a specific enzyme family.

It was also identified an EST, whose identity corresponds to antigen p200 of 200 kDa, an immunodominant protein conserved in *B. bigemina* isolates of diverse geographical origins in Africa and México⁽³³⁾. It has been proposed that such antigen participates in immune response elucidation and its existence as a structural cytoskeleton protein remains a matter of speculation. Apparently, the transcript of this protein is found at very low levels in *B. bigemina*, as only one EST came up in the analysis.

An EST with high homology to the *B. bigemina* RAP-1 (Rhoptry Associated Protein 1) antigen was identified. Previous studies⁽³⁴⁾ report that RAP-1 induces a protective immune response in bovine cattle. Nevertheless, several RAP-1 variants have been identified, showing sequence polymorphism in both amino-terminal and carboxy-terminal regions, which suggests that variation in regions with dimorphism in RAP-1 sequence are

de la Membrana Apical (AMA-1). Originalmente reportado como antígeno encontrado en la fase asexual del estadio intraeritrocítico de *Plasmodium falciparum*, AMA-1 es una proteína de micronema que se secreta a la superficie del merozoito de *Plasmodium* spp y taquizoito de *Toxoplasma gondii*. Más tarde, fue identificado el gen ortólogo en un aislado israelita de *B. bovis*⁽²⁹⁾, posteriormente se identificó y anotó la presencia del antígeno AMA-1 en el genoma secuenciado del aislado Texas de *B. bovis*⁽²⁴⁾ y más tarde fue secuenciado el gen a partir de cinco aislados de *B. bovis* con diferente origen geográfico en Brasil⁽³⁰⁾, demostrando su elevada conservación. Más recientemente, la secuencia correspondiente al gen AMA-1 de *B. bigemina* (Aislados argentino e italianos) fueron agregadas en el banco de genes^(31,32). Experimentos de inmunofluorescencia con anticuerpos anti péptidos de AMA-1 demostraron que el antígeno fue localizado en la mitad apical del merozoito de *B. bovis*. Sin embargo, experimentos de inmunoelectrotransferencia y marcado metabólico del parásito mostraron que el antígeno AMA-1 de *B. bovis* de 82 KDa se encuentra presente en muy pequeñas cantidades en el merozoito⁽²⁹⁾. Es probable que exista un bajo nivel de expresión de transcritos del ortólogo AMA-1 de *B. bigemina*, como se puede evidenciar por el escaso número de EST identificadas en este estudio. No obstante, el nivel de expresión de los transcritos no está directa y necesariamente relacionado con la cantidad de producto traducido. Por ejemplo, en genes que codifican para enzimas metabólicas los transcritos son ampliamente encontrados en los genomas, mientras que la concentración de enzima en un extracto proteico es baja, pero con una actividad enzimática elevada. Se desconoce si AMA-1 de *Babesia* spp pertenece a alguna familia particular de enzimas.

Se identificó una EST cuya identidad correspondió con el antígeno p200 de 200 KDa. Una proteína inmunodominante conservada en aislados de *B. bigemina* de diferente origen geográfico, África y México⁽³³⁾. Se propone que este antígeno participa en la elucidación de una respuesta inmune, y se especula que existe como una proteína estructural del citoesqueleto. Aparentemente, el transcrito de esta proteína se encuentra en bajos niveles en *B. bigemina*, dado que sólo se presentó una EST en el

immunologically relevant, as antibodies targeted on epitopes present in an amino-terminal variant do not recognize the antigen in a second variant, and T CD4+ cells that recognize epitopes of a carboxy-terminal variant do not cross-react with a second carboxy-terminal variant. Analysis of the RAP-1 locus in the *B. bigemina* genome showed that *rap-1* genes are distributed as tandem repeats, but with a different number of gene copies and arrangements depending on the strain and isolate being analyzed. Gene conversion and unequal recombination events contribute to generate new *rap-1* variants⁽¹³⁾.

An EST with high homology with a *B. bovis* anonymous antigen was identified. This molecule is apparently a membrane protein, described as homologous to BBOV_III008720 found in the *B. bovis* genome. In the present study, a Blast homology search showed the presence of 12 ESTs with 42 % identity in amino acidic sequence. More experimentation with this antigen is required for establishing its biological importance.

An EST with homology to a new antigen found in *Babesia* sp., WA1 (of human origin in the state of Washington, USA), whose biological function is described as unknown, and corresponds to a type II putative membrane protein, containing a von Willebrand factor type A domain.

The importance of identifying in this analysis 4 clones with *B. bigemina* EST showing similarity to antigens of the *B. bovis* variable erythrocyte surface antigens (VESA) family and involved in the immune escape mechanism warrants its experimental analysis in later studies. ESTs identify *ves* gene sequences that comprise a large family in the *B. bovis* genome⁽²⁴⁾. VESA-1 is a 300 kDa heterodimeric protein made up by VESA-1 alpha and VESA-1 beta^(35,36,37). These proteins are synthesized by the *B. bovis* merozoite and exported later to the erythrocyte surface⁽³⁵⁾. VESA-1 goes through a quick antigenic variation, which is involved in both immune response evasion and cytoadhesion, factors that seem to play a vital role in parasite persistence and in its pathogenesis. VESA 1 shows functional homology with protein PfEMP, coded by the *var* gene family in *Plasmodium falciparum*⁽³⁶⁾. In fact, 3 additional EST were identified in the present study that had non statistically significant match with *var* proteins present in surfaces of erythrocytes infected by

análisis realizado.

Se identificó una EST con elevada homología al antígeno de *B. bigemina* denominado Proteína Asociada a la Roptria-1 (RAP-1). Estudios previamente realizados reportan que RAP-1 induce una respuesta inmune protectora en bovinos⁽³⁴⁾. Sin embargo, se han identificado variantes de RAP-1, con polimorfismo de secuencia en las regiones amino-terminal y carboxi-terminal, lo que sugiere que la variación en las regiones con dimorfismo en secuencia de RAP-1 son inmunológicamente significativas, ya que anticuerpos dirigidos a epítopes presentes en una variante amino-terminal no reconocen el antígeno en una segunda variante; y células T CD4+ que reconocen epítopes en una variante carboxi-terminal no reaccionan en forma cruzada con una segunda variante carboxi-terminal. Un análisis del locus RAP-1 en el genoma de *B. bigemina* demostró que los genes *rap-1* se encuentran distribuidos como repetidos en tándem, pero con diferente número de copias del gen y arreglos dependiendo de la cepa o aislado analizado. Eventos de conversión génica y desigual recombinación contribuyen a la generación de nuevas variantes *rap-1*⁽¹³⁾.

Una EST con homología para con un antígeno anónimo de *B. bovis* fue identificada. La molécula aparentemente es una proteína de membrana, descrita como homóloga a la proteína BBOV_III008720 encontrada en el genoma de *B. bovis*. En este estudio el análisis Blast mostró la presencia de 12 EST con una identidad de 42 % en secuencia aminoacídica. Se requiere de mayor experimentación con este antígeno para determinar su importancia biológica.

Se encontró una EST con homología a un nuevo antígeno encontrado en *Babesia* sp. WA1 (de origen humano en el estado de Washington, EE.UU.) y cuya función biológica se describe como desconocida hasta ahora. Sólo se establece que corresponde a una supuesta proteína de membrana tipo II, conteniendo un dominio del factor von Willebrand tipo A.

Importante destacar en este análisis, el hecho de identificar 4 clonas con EST de *B. bigemina* con similitud a los antígenos de la familia de antígenos variables de la superficie del eritrocito (VESA) de *B. bovis*, implicados en el mecanismo de escape inmune, y que garantiza su análisis experimental

Plasmodium falciparum. It is a known fact that VESA-1 is involved in the cytoadhesion process of erythrocytes infected by *B. bovis* that causes erythrocyte sequestration in brain capillaries and consequently development of a more severe illness. Although no erythrocyte sequestration has been described in *B. bigemina* infections, it can be surmised that the function of VESA1 orthologous proteins present in *B. bigemina* play an important role in the antigenic variation of the parasite once erythrocytes are infected. Alternatively, antigens similar to *B. bigemina* VESA-1 could take part in antibody-mediated aggregation/agglutination of infected erythrocytes, a process exemplified in *in vitro* experiments with *B. bigemina* infected erythrocytes and immune bovine serum containing antibodies that recognize *B. bigemina* infected erythrocyte surface antigens⁽³⁸⁾. Whether the antigens identified in *B. bigemina* erythrocyte surfaces correspond to *B. bovis* VESA antigens, then it should be demonstrated through experimentation.

The amount of EST acquired and analyzed in the present study and graded as either physiologically or immunologically important for the parasite and illness development, is low (491 single genes), only 14 % of almost 3,500 genes present in the *B. bovis* genome, a phylogenetically similar parasite⁽²⁴⁾. This could be due to a low quantity of recombinant phagemids acquired after excision of the λZAP gene library. There is a possibility that a high percentage of cDNAs contained in the gene library could have been lost during the pBluescript plasmid acquiring process in addition to the fact that the constructed gene library was not normalized on purpose.

However, *B. bigemina* gene regions identified in the EST analysis performed in the present study could provide guidelines to carry out research conducive to characterization of expressed proteins considered as functionally significant in the molecular machinery of the parasite. For example, those involved in metabolism or in the signaling mechanisms developed by the parasite when invading red cells that would allow envisioning the development of experiments designed to test blocking of these proteins with new or existing chemotherapeutic drugs not yet tested on *B. bigemina*. Nonetheless, it should be of importance to study and characterize *B. bigemina* specific

en posteriores estudios. Las EST identifican secuencias de genes *ves* que comprenden una gran familia en el genoma de *B. bovis*⁽²⁴⁾. VESA 1 es una proteína heterodimérica de 300 KDa compuesta de VESA 1 alfa y VESA 1 beta^(35, 36,37). Las proteínas son sintetizadas por el merozoito de *B. bovis* y exportada posteriormente a la superficie del eritrocito⁽³⁵⁾. VESA 1 sufre una rápida variación antigénica, la cual está implicada en la evasión de la respuesta inmune y cito-adhesión; ambos factores parecen jugar un papel vital en la persistencia del parásito y su patogénesis. VESA 1 muestra una homología funcional con la proteína PfEMP1, codificado por la familia de genes *var* en *P. falciparum*⁽³⁶⁾. De hecho, en este estudio se identificaron 3 EST adicionales que tuvieron encuentro, si bien no estadísticamente significativo, con proteínas tipo *var* presentes en la superficie del eritrocito infectado con *P. falciparum*. Se sabe que VESA-1 está implicada en el proceso de cito-adherencia de eritrocitos infectados con *B. bovis*, que ocasiona un secuestro de eritrocitos en los capilares del cerebro y como consecuencia un desarrollo de la enfermedad mucho más grave. Si bien en casos de infección por *B. bigemina* no se ha demostrado secuestro de eritrocitos durante el proceso de la enfermedad, se puede especular que la función de las proteínas ortólogas a VESA-1 presentes en *B. bigemina* juegan un papel importante en la variación antigénica del parásito una vez que infectan al eritrocito. Alternativamente, antígenos similares a VESA-1 de *B. bigemina* pudieran participar en la agregación/aglutinación de eritrocitos infectados, mediada por anticuerpos, proceso el cual se ha demostrado *in vitro* en ensayos con eritrocitos infectados con *B. bigemina* y suero de bovino inmune conteniendo anticuerpos que reconocen antígenos expuestos en la superficie del eritrocito infectado con *B. bigemina*⁽³⁸⁾. Si los antígenos identificados en la superficie de eritrocitos infectados con *B. bigemina* corresponden a los antígenos VESA de *B. bovis*, requiere todavía ser demostrado experimentalmente.

En general, la cantidad de EST obtenidas y analizadas en este estudio y clasificadas como de importancia fisiológica o inmunológica para el parásito y el desarrollo de la enfermedad es baja (491 genes únicos), ya que sólo representa el 14 % de los aproximadamente 3,500 genes presentes en

proteínas in detail in order to determine the structural differences between parasite and counterpart bovine host proteins, with the aim of not affecting the physiological functions of the bovine host.

Another type of research should imply studying proteins expressed in the parasite, containing conserved regions among different isolates with capacity to generate a protective immune response in the bovine host against *B. bigemina*. Probably, starting from one of these conserved genic regions it could be possible to build hybrid proteins (multiepitope, multivalent) for generating new generation recombinant vaccines that could help control this parasite.

CONCLUSIONS

Generation of databases with EST of a microorganism constitutes a powerful tool for identification and selection of regions that may become candidates for establishing control strategies and genetic regulation. Besides, it allows discovery of new genes whose importance for the parasite is such that they could eventually be set as targets for developing new and more effective chemotherapeutic agents or else, of new generation vaccines. In the present study, 267 *B. bigemina* sequences were identified, which did not show any kind of similitude with either *B. bovis* or any other Apicomplexa parasite. Presumably, these sequences correspond to new genes, whose functions assigned through similitude have not been identified and not yet reported for *B. bigemina*. However, they could also correspond to sequences of untranslated gene regions that the Blast search does not allow to identify, because this algorithm performs sequence identity search in function of a conceptual translation of a blank sequence and compares it to the amino acid sequence saved in databases. Besides, it is worth mentioning that even when the analysis was performed at a relatively low scale, the fact of identifying a *B. bigemina* EST showing similitude to *B. bovis* VESA family antigens allegedly involved in the immune escape mechanism, contributes to guarantee its experimental analysis in further studies. Other identified sequences that codify for relevant proteins are those found showing high homology with Spherical Body proteins (SBP) in *B. bovis*, expressed in merozoite spherical bodies, organelles

el genoma de *B. bovis*, un parásito filogenéticamente similar⁽²⁴⁾. Esto puede ser resultado directo de la baja cantidad de fagémidos recombinantes obtenidos en la escisión de la genoteca en λZAP, existiendo la posibilidad de que un elevado porcentaje de los cDNAs contenidos en la genoteca se haya perdido durante el proceso de obtención de los plásmidos pBluescript, además de que la genoteca construida no fue normalizada expresos.

No obstante, las regiones génicas identificadas en *Babesia bigemina* durante el análisis de EST desarrollado en este trabajo podrá dar la pauta para realizar investigación conducente a la caracterización de las proteínas expresadas consideradas como importantes, funcionalmente, en la maquinaria molecular del parásito. Por ejemplo, las involucradas en el metabolismo o en los mecanismos de señalización que desarrolla el parásito al momento de invadir los glóbulos rojos, permitiría vislumbrar el desarrollo de ensayos de bloqueo de estas proteínas con nuevas drogas quimioterapéuticas (o drogas ya existentes, pero aún no probadas en *B. bigemina*). Sin embargo, será importante estudiar y caracterizar detalladamente las proteínas específicas de *B. bigemina* para determinar las diferencias estructurales existentes entre proteínas del parásito y la contraparte del hospedador bovino, con el objeto de no afectar las funciones fisiológicas del hospedero bovino.

Otro tipo de investigaciones implicaría el estudio de proteínas que se expresan en el parásito conteniendo regiones conservadas entre distintos aislados, y que tengan la capacidad de generar una respuesta inmune protectora por parte del bovino contra *B. bigemina*. Es probable que partiendo de estas regiones génicas conservadas se puedan construir proteínas híbridas (multiepitópicas, multivalentes) para generar un tipo de vacuna recombinante de nueva generación para coadyuvar en el control de este parásito.

CONCLUSIONES

La generación de bancos de datos con EST de un microorganismo es una herramienta poderosa para la identificación y selección de regiones candidatas que permiten establecer estrategias de control y regulación genética. Asimismo, permiten descubrir nuevos genes cuya importancia biológica para el

located in the apical region of the intraerythrocytic parasites. Given their importance in the invasion process and to lack of information on homologous proteins corresponding to *B. bigemina*, those identified in the present study assure their inclusion in further studies. Similarly, antigens identified as Bbg 1.1 and Bbg 1.2 should be revisited in terms of its immunobiologically molecular characterization. If further studies show that they are highly conserved in Mexican isolates, it should be of particular interest to assess both their immunogenic and immunoprotective potential.

Results obtained in the present study, together with analysis of a larger number of *B. bigemina* EST would allow characterization and manipulation of coding gene regions of the intraerythrocytic stage, essential for developing new bovine babesiosis control strategies.

ACKNOWLEDGMENTS

The present study was funded by SEP-CONACYT Fondo Sectorial, Project # 2004-C01-47739. The authors wish to thank most sincerely the anonymous reviewers for their observations and suggestions, which undoubtedly improved this paper.

End of English version

parásito es tal, que eventualmente pueden servir como blancos para el desarrollo de nuevos y más efectivos agentes quimioterapéuticos, o para el desarrollo de vacunas de nueva generación. En este trabajo se identificaron 267 secuencias de *B. bigemina* que no mostraron similitud alguna con *B. bovis* ni con ningún otro parásito del *Phylum* Apicomplexa. Presumiblemente, estas secuencias corresponden a genes nuevos sin identificación de función asignada por similitud y aún no reportados para *B. bigemina*. Sin embargo, pudieran corresponder también, a secuencias de regiones no traducidas de genes que el análisis Blast no permite identificar, dado que este algoritmo realiza la búsqueda de identidad de secuencia en función de la traducción conceptual de la secuencia blanco, y la compara con la secuencia de aminoácidos depositadas en los bancos de datos.

Asimismo, es importante destacar que aun cuando el análisis fue realizado a relativamente baja escala, el hecho de identificar EST en *B. bigemina* con similitud a los antígenos de la familia VESA de *B. bovis* supuestamente implicados en el mecanismo de escape inmune, es una contribución que garantiza su análisis experimental en posteriores estudios. Otras secuencias identificadas que codifican para proteínas de importancia son las encontradas con una elevada homología con Proteínas de Cuerpo Esférico, denominadas como SBP en *B. bovis*, que se expresan en los cuerpos esféricos de los merozoitos, organelos localizados en la región apical de los parásitos intra-eritrocíticos. Dada su importancia en el proceso de invasión y la carencia de información con respecto a los homólogos correspondientes a *B. bigemina*, las identificadas en este estudio garantizan su inclusión en estudios de caracterización posterior.

LITERATURA CITADA

- Friedhoff KT. Transmission of *Babesia*, In: Ristic M editor, Babesiosis of Domestic Animals and Man, 1st ed. Boca Raton, Florida, USA: CRC Press; 1988:23-52.
- Bock R, Jackson L, de Vos A, Jorgensen W. Babesiosis of cattle. Parasitol 2004; 129:S247-S269.
- FAO, Delegación Mexicana. Estimación de pérdidas económicas por enfermedades en la ganadería Mexicana durante el año 1980. Bull Off Int Epiz 1981;93:903-915.
- Servicio de Información Agroalimentaria y Pesquera, SAGARPA, México. Ganadería, Bovino Carne y Leche. Población ganadera 2000 - 2009, [en línea]: <http://www.siea.sagarpa.gob.mx/>. Consultado Junio 2011.
- Pipano E. Live vaccines against hemoparasitic disease in livestock. Vet Parasitol 1995;57:213-231.
- McElwain TF, Perryman LE, Davis WC, McGuire TC. Antibodies define multiple proteins with epitopes exposed on the surface of live *Babesia bigemina* merozoites. J Immunol 1987;138(7):2298-2304.
- Figueroa JV, Buening GM, Kinden DA, Green TJ. Identification of common surface antigens among *Babesia bigemina* isolates by using monoclonal antibodies. Parasitol 1990;100(2):161-175.
- Wright IG, Casu R, Commins MA, Dalrymple BP, *et al.* The development of a recombinant *Babesia* vaccine. Vet Parasitol 1992;44 3-13.
- Brown WC, Palmer GH. Designing blood-stage vaccine against *Babesia bovis* and *Babesia bigemina*. Parasitol Today 1999;15:275-281.
- Brown WC, Norimine J, Goff WL, Suarez CE, McElwain TF. Prospects for recombinant vaccines against *Babesia bovis* and related parasites. Parasite Immunol 2006;28:315-327.
- Vega CA, Buening GM, Rodriguez SD, Carson CA. In vitro cultivation of *Babesia bigemina*. Am J Vet Res 1985;46(2):416-420.
- Figueroa-Millán JV, Cantó-Alarcón GJ, Álvarez-Martínez JA, Lona-Galván R, Ramos-Aragón JA, Vega y Murguía CA. Capacidad protectora en bovinos de una cepa de *Babesia bigemina* derivada de cultivo *in vitro*. Téc Pecu Méx 1998;36: 95-108.
- Hotzel I, Suarez CE, McElwain TF, Palmer GH. Genetic variation in the dimorphic regions of RAP-1 genes and rap-1 loci of *Babesia bigemina*. Mol Biochem Parasitol 1997;90(2):479-489.
- Adams MD, Kelley JM, Gocayne JD, *et al.* Complementary DNA sequencing: expressed sequence tags and the human genome project. Science 1991; 252:1651-1656.
- Black MW, Boothroyd JC. Lytic cycle of *Toxoplasma gondii*. Microbiol Mol Biol 2000;64:607-623.
- Chakrabarti D, Reddy GR, Dame JB, Almira EC, Lapis PJ, Ferl RJ, Yang TP, Rowe TC, Schuster SM. Analysis of expressed sequence tags from *Plasmodium falciparum*. Mol Biochem Parasitol 1994 66:97-104.
- Li L, Brunk BP, Kissinger JC, Pape D. *et al.* Gene discovery in the Apicomplexa as revealed by EST sequencing and assembly of a comparative gene database. Genome Res 2003;13:443-454.
- de Vries E, Corton C, Harris B, Cornelissen AW, Berriman M. Expressed sequence tag (EST) analysis of the erythrocytic stage of *Babesia bovis*. Vet Parasitol 2006;138:61-74.
- Boguski MS. The turning point in genome research. Trends Biochem Sci 1995; 20:295-296.
- Figueroa JV, Buening GM, Mishra V, McElwain TF. Screening of a *Babesia bigemina* cDNA library with monoclonal antibodies directed to surface antigens. Annals New York Acad Sci 1992;653:122-130.

IDENTIFICACIÓN DE GENES EN *B. bigemina* MEDIANTE ETIQUETAS DE SECUENCIA EXPRESADAS

21. Short JM, Fernández JM, Sorge JA, Huse WD. λZAP: A bacteriophage λ expression vector with in vivo excision properties. *Nucleic Acids Res* 1988; 16:7583-7600.
22. Sambrook J, Fritsch E, Maniatis T. *Molecular cloning a Laboratory Manual*. 2nd ed. Cold Spring Harbor, New York. 1989.
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; 215(3):403-410.
24. Brayton KA, Lau AOT, Herndon DR. *et al*, Genome Sequence of *Babesia bovis* and Comparative Analysis of Apicomplexan Hemoprotozoa. *PLoS Pathog* 2007; 3(10): e148. doi:10.1371/journal.ppat.0030148.
25. Kung'u MW, Dalrymple BP, Wright IG, Peters JM. Cloning and characterisation of members of a family of *Babesia bigemina* antigen genes containing repeated sequences. *Mol Biochem Parasitol* 1992;55:29-38.
26. Hines SA, Palmer GH, Brown WC, McElwain TF, Suarez CE, Vidotto O, Rice-Ficht AC. Genetic and antigenic characterization of *Babesia bovis* merozoite spherical body protein Bp-1. *Mol Biochem Parasitol* 1995;69:149-159.
27. Dowling SC, Perryman LE, Jasmer DP. A *Babesia bovis* 225-Kilodalton spherical-body protein: localization to the cytoplasmic face of infected erythrocytes after merozoite invasion. *Infect Immun* 1996;64:2618-2626.
28. Ruef BJ, Dowling SC, Conley PG, Perryman LE, Brown WC, Jasmer DP, Rice-Ficht AC. A unique *Babesia bovis* spherical body protein is conserved among geographic isolates and localizes to the infected erythrocyte membrane. *Mol Biochem Parasitol* 2000;105:1-12.
29. Gaffar FR, Yatsuda AP, Franssen FF, de Vries E. Erythrocyte invasion by *Babesia bovis* merozoites is inhibited by polyclonal antisera directed against peptides derived from a homologue of *Plasmodium falciparum* Apical Membrane Antigen 1. *Infect Immun* 2004;72(5):2947-2955.
30. Ramos CAN, Araujo FR, Souza IIF, Oliveira RHM, Soares CO, Alves LC. Molecular characterization of Brazilian isolates of *Babesia bovis*. Genbank: FJ588024.1. [on line]
31. AbouLaila MR, Yokoyama N, Igarashi I. Molecular cloning and phylogenetic analysis of *Babesia bigemina* apical membrane antigen-1(AMA-1). Genbank:AB481200.2. [on line] <http://www.ncbi.nlm.nih.gov/nuccore/222101623>. Accessed September 2009.
32. Torina A, Agnone A, Sireci G, Mosqueda JJ, Blanda V, Albanese I, La Farina M, Cerrone A, Cusumanu F, Caracappa S. Characterization of the apical membrane antigen 1 in Italian strains of *Babesia bigemina*. *Trans Emerg Dis* 2010;57(1-2): 52-56.
33. Tebele N, Skilton RA, Katende J, Wells CW, Nene V, McElwain T, Morzaria SP, Musoke AJ. Cloning, characterization, and expression of a 200-kilodalton diagnostic antigen of *Babesia bigemina*. *J Clin Microbiol* 2000;38(6):2240-2247.
34. McElwain TF, Perryman LE, Musoke AJ, McGuire TC. Molecular characterization and immunogenicity of neutralization-sensitive *Babesia bigemina* merozoite surface protein. *Mol Biochem Parasitol* 1991;47:213-222.
35. O'Connor RM, Lane TJ, Stroup SE, Allred DR. Characterization of a variant erythrocyte surface antigen (VESA1) expressed by *Babesia bovis* during antigenic variation. *Mol Biochem Parasitol* 1997;89:259-270.
36. Allred DR, Cinque DM, Lane TJ, Ahrens KP. Antigenic variation of parasite-derived antigens on the surface of *Babesia bovis*-infected erythrocytes. *Infect Immun* 1994;62:91-98.
37. Allred RD, Carlton JM, Satcher RI, Long JA, Brown WC, *et al*. The ves multigene family of *Babesia bovis* encodes components of rapid antigenic variation at the infected erythrocytes surface. *Mol Cel* 2000;5:153-162.
38. Shompole S, McElwain TF, Jasmer DP, Hines SA, Katende J, Musoke A, Rurangirwa FR, McGuire TC. Identification of *Babesia bigemina* infected erythrocyte surface antigens containing epitopes conserved among strains. *Parasite Immunol* 1994;16:119-127.

Anexo 1. Análisis de Etiquetas de Secuencia Expresadas (EST) en *Babesia bigemina*
Attachment 1. Expressed Sequence Tags (EST) analysis in *Babesia bigemina*

EST	L	S	E	Identidad (%)	Anotación	Gen ID	No. EST
1	1177	1222	4e-132	272/301 (90%)	<i>B. bigemina</i> apical membrane antigen-1	dbj BAH227706.1	1
2	159	298	2e-69	151/152 (99%)	<i>B. bigemina</i> 18S ribosomal RNA gene	gb DQ785311.1	1
3	250	204	1e-16	40/46 (86%)	<i>B. bigemina</i> 200KdA Antigen p200	AAF63787.1	1
4	1095	616	7e-66	129/196 (65%)	<i>B. bigemina</i> Bbg 1.1 antigen	gb AAA27790.1	12
5	1191	521	9e-51	124/253 (49%)	<i>B. bigemina</i> Bbg 2.1 antigen	gb AAA27792.1	5
6	192	185	1e-12	36/39 (92%)	<i>B. bigemina</i> RAP-1 alpha-1	gb AAG14907.1	1
7	611	263	8e-69	131/197 (66%)	<i>B. bovis</i> RNA 3' phosphate cyclase [BBOV_IV009590]	XP_001610647.1	1
8	191	249	4e-20	43/53 (81%)	<i>B. bovis</i> 40S ribosomal protein S14p/S29e [BBOV_III003810]	XP_001611512.1	1
9	302	92.8	8E-18	43/51 (84%)	<i>B. bovis</i> 40S ribosomal protein S17, [BBOV_III002940]	XP_001611260.1	1
10	457	404	4e-38	85/119 (71%)	<i>B. bovis</i> 40S ribosomal protein S20 [BBOV_II004020]	XP_001609925.1	9
11	296	329	2e-29	63/77 (81%)	<i>B. bovis</i> 40S ribosomal protein S25 [BBOV_IV010200]	XP_001610942.1	5
12	381	196	6e-49	99/107 (92%)	<i>B. bovis</i> 40S ribosomal protein S26e [BBOV_III002940]	XP_001611426.1	1
13	643	416	3e-39	109/148 (73%)	<i>B. bovis</i> 40S ribosomal protein S27a protein [BBOV_III008500]	XP_001611978.1	10
14	664	1032	1e-110	202/217 (93%)	<i>B. bovis</i> 40S ribosomal protein S3 protein [BBOV_II001650]	XP_001609690.1	13
15	78	96	1e-03	16/22 (72%)	<i>B. bovis</i> 40S ribosomal protein S30 [BBOV_III002090]	XP_001611344.1	12
16	98	117	7e-05	22/26 (84%)	<i>B. bovis</i> 40S ribosomal protein S7e protein [BBOV_III009350]	XP_001612059.1	5
17	496	811	3e-85	157/165 (95%)	<i>B. bovis</i> 40S ribosomal subunit protein S9 [BBOV_II002670]	XP_001609790.1	7
18	231	253	1e-20	50/68 (73%)	<i>B. bovis</i> 60S ribosomal L27e family protein [BBOV_III010680]	XP_001612188.1	1
19	284	61.2	2e-08	28/34 (82%)	<i>B. bovis</i> 60S ribosomal protein L18p/L5e [BBOV_III010590]	XP_001612181.1	2
20	189	75.5	1e-03	26/24 (76%)	<i>B. bovis</i> 60S ribosomal protein L10 [BBOV_III010500]	XP_001612172.1	1
21	667	979	2e-104	196/204 (96%)	<i>B. bovis</i> 60S ribosomal protein L15 [BBOV_III000550]	XP_001611190.1	34
22	127	224	3e-17	42/42 (100%)	<i>B. bovis</i> 60S ribosomal protein L2 [BBOV_III011800]	XP_001612300.1	1
23	210	126	5e-28	13/13 (100%)	<i>B. bovis</i> 60S ribosomal protein L23 [BBOV_III005180]	XP_001611293.1	1
24	212	124	4e-22	148/205 (72%)	<i>B. bovis</i> 60S ribosomal protein L24e [BBOV_III003570]	XP_001611489.1	2
25	813	216	1e-54	128/183 (69%)	<i>B. bovis</i> 60S ribosomal protein L3 [BBOV_I002850]	XP_001608935.1	2
26	158	89	1e-16	43/48 (89%)	<i>B. bovis</i> 60S ribosomal protein L35 [BBOV_IV008090]	XP_001610731.1	1
27	159	247	7e-20	50/51 (98%)	<i>B. bovis</i> 60S ribosomal protein L36 [BBOV_II003460]	XP_001609869.1	7
28	369	73.4	7E-10	14/37 (37%)	<i>B. bovis</i> 60S ribosomal protein L44 [BBOV_II001060]	XP_001609633.1	1
29	191	86	7e-12	96/131 (73%)	<i>B. bovis</i> 60S ribosomal protein L6e [BBOV_II000880]	XP_001609566.1	3

IDENTIFICACIÓN DE GENES EN *B. bigemina* MEDIANTE ETIQUETAS DE SECUENCIA EXPRESADAS

EST	L	S	E	Identidad (%)	Anotación	Gen ID	No. EST
30	192	104	1e-03	20/24 (83%)	<i>B. bovis</i> acetyltransferase, GNAT protein [BBOV_III007230]	XP_001611852.1	6
31	759	872	7e-92	162/211 (76%)	<i>B. bovis</i> aconitate hydratase 1 family protein [BBOV_III011790]	XP_001612299.1	10
32	154	49.3	1e-04	27/45 (60%)	<i>B. bovis</i> actin [BBOV_I000300]	XP_001608692.1	1
33	131	147	3e-08	26/42 (61%)	<i>B. bovis</i> acylphosphatase [BBOV_IV011650]	XP_001611085.1	2
34	573	223	7e-17	44/119 (36%)	<i>B. bovis</i> adenosine kinase [BBOV_IV003810]	XP_001610310.1	13
35	573	217	3e-16	47/119 (39%)	<i>B. bovis</i> adenosine kinase [BBOV_IV003860]	XP_001610310.1	2
36	311	55	0.001	15/37 (40%)	<i>B. bovis</i> peptide chain release factor 1 [BBOB_IV006160]	XP_001610544.1	1
37	245	124	1e-05	27/63 (42%)	<i>B. bovis</i> anonymous antigen-3	gb ABC25606.1	1
38	1063	122	8e-21	166/234 (70%)	<i>B. bovis</i> apical membrane antigen 1 [BBOV_IV011230]	gb AY486101.1	1
39	161	187	3e-16	39/48 (81%)	<i>B. bovis</i> apocytochrome b gene	gb AF053002.1	13
40	486	552	3e-55	104/149 (69%)	<i>B. bovis</i> ATPase, AAA family protein [BBOV_III001380]	XP_001611273.1	1
41	148	128	4e-06	24/26 (92%),	<i>B. bovis</i> casein kinase II, alpha chain (CK II) [BBOV_III007790]	XP_001611907.1	2
42	259	245	1e-19	44/55 (80%)	<i>B. bovis</i> citrate synthase [BBOV_II006920]	XP_001610210.1	5
43	216	161	6e-10	33/53 (62%)	<i>B. bovis</i> coatomer beta subunit [BBOV_IV005890]	XP_001610518.1	2
44	1214	244	5e-33	50/113 (44%)	<i>B. bovis</i> conserved hypothetical protein [BBOV_II003760]	XP_001609899.1	9
45	1024	1051	1e-116	192/213 (90%)	<i>B. bovis</i> conserved hypothetical protein [BBOV_III004000]	XP_001611531.1	95
46	739	728	2e-95	135/168 (80%)	<i>B. bovis</i> CRAL/TRIO domain containing protein [BBOV_III011520]	XP_001612272.1	1
47	776	1063	1e-117	192/214 (89%)	<i>B. bovis</i> deoxyhypusine synthase [BBOV_III010890]	XP_001612209.1	4
48	841	1074	3e-121	200/243 (82%)	<i>B. bovis</i> DNA replication licensing factor MCM2 [BBOV_II005900]	XP_001610108.1	14
49	549	102	1e-03	26/54 (48%)	<i>B. bovis</i> Dullard-like phosphatase domain protein [BBOV_III004040]	XP_001611535.1	3
50	133	102	2e-04	19/23 (83%)	<i>B. bovis</i> elongation factor 1-alpha [BBOV_IV010620]	XP_001610983.1	2
51	140	40	7e-11	57/65 (87%)	<i>B. bovis</i> elongation factor 1-alpha [BBOV_IV010630]	XM_001610934.1	1
52	889	671	2e-70	124/197 (62%)	<i>B. bovis</i> elongation factor Tu GTP binding protein [BBOV_II003130]	XP_001609836.1	7
53	1167	420	2e-125	79/120 (65%)	<i>B. bovis</i> endonuclease/exonuclease/phosphatase [BBOV_IV011570]	XM_001611027.1	2
54	275	183	2e-12	42/72 (58%)	<i>B. bovis</i> enolase [BBOV_III007950]	XP_001611923.1	3
55	244	268	1e-21	50/53 (94%)	<i>B. bovis</i> eukaryotic translation initiation factor 4A [BBOV_III010250]	XP_001612149.1	29
56	1010	127	2e-27	76/91 (83%)	<i>B. bovis</i> fructose 1,6 - biphosphate aldolase [BBOV_I000300]	XP_001609245.1	1

EST	L	S	E	Identidad (%)	Anotación	Gen ID	No. EST
57	138	132	1e-06	23/34 (67%)	<i>B. bovis</i> glucosamine-fructose-aminotransferase [BOV IV00250]	XP_001609193.1	7
58	1138	262	7e-21	56/170 (32%)	<i>B. bovis</i> HAD superfamily hydrolase [BBOV IV003220]	XP_001609486.1	22
59	186	66	1e-03	14/20 (70%)	<i>B. bovis</i> HAD superfamily hydrolase [BBOV IV003240]	XP_001609488.1	1
60	193	124	1e-05	22/55 (40%)	<i>B. bovis</i> HAD superfamily hydrolase [BBOV IV003310]	XP_001609495.1	1
61	1043	620	2e-62	117/204 (57%)	<i>B. bovis</i> helicase [BBOV IV011580]	XP_001611078.1	18
62	189	121	2e-26	57/61 (93%)	<i>B. bovis</i> hypothetical cytochrome C1 protein [BBOV II002550]	XP_001609778.1	1
63	689	253	9e-66	101/138 (72%)	<i>B. bovis</i> hypothetical importin beta subunit [BBOV IV000520]	XP_001609220.1	1
64	492	201	2e-14	64/109 (58%)	<i>B. bovis</i> hypothetical protein [BBOV I001040]	XP_001608765.1	7
65	1180	230	4e-17	52/180 (28%)	<i>B. bovis</i> hypothetical protein [BBOV I001400]	XP_001608792.1	2
66	228	136	5e-07	31/73 (42%)	<i>B. bovis</i> hypothetical protein [BBOV I002120]	XP_001608863.1	7
67	211	117	8e-05	23/33 (69%)	<i>B. bovis</i> hypothetical protein [BBOV I002450]	XP_001608895.1	55
68	166	85	1e-03	18/30 (60%)	<i>B. bovis</i> hypothetical protein [BBOV I004240]	XP_001609073.1	5
69	1180	212	5e-15	48/178 (26%)	<i>B. bovis</i> hypothetical protein [BBOV II000090]	XP_001609538.1	2
70	971	79.3	4e-13	44/95 (46%)	<i>B. bovis</i> hypothetical protein [BBOV II000630]	XP_001609591.1	1
71	171	135	6e-07	25/34 (73%)	<i>B. bovis</i> hypothetical protein [BBOV II000930]	XP_001609621.1	1
72	767	361	1e-32	87/261 (33%)	<i>B. bovis</i> hypothetical protein [BBOV II001180]	XP_001609645.1	5
73	203	93	1e-03	20/37 (54%)	<i>B. bovis</i> hypothetical protein [BBOV II001760]	XP_001609701.1	2
74	499	186	9e-13	46/126 (36%)	<i>B. bovis</i> hypothetical protein [BBOV II002060]	XP_001609730.1	7
75	337	94	1e-03	21/53 (39%)	<i>B. bovis</i> hypothetical protein [BBOV II002630]	XP_001609786.1	67
76	336	98	1e-03	24/77 (31%)	<i>B. bovis</i> hypothetical protein [BBOV II002870]	XP_001609810.1	4
77	494	614	2e-62	102/155 (65%)	<i>B. bovis</i> hypothetical protein [BBOV II003230]	XP_001609846.1	1
78	684	248	1e-19	52/115 (45%)	<i>B. bovis</i> hypothetical protein [BBOV II003760]	XP_001609899.1	9
79	151	126	7e-26	24/50 (48%)	<i>B. bovis</i> hypothetical protein [BBOV II004230]	XP_001609946.1	1
80	245	287	2e-24	55/82 (67%)	<i>B. bovis</i> hypothetical protein [BBOV II005360]	XP_001610055.1	21
81	1172	569	3e-60	131/275 (47%)	<i>B. bovis</i> hypothetical protein [BBOV II005480]	XP_001610067.1	12
82	130	110	5e-04	18/35 (51%)	<i>B. bovis</i> hypothetical protein [BBOV II006250]	XP_001610143.1	2
83	364	147	2e-34	58/78 (74%)	<i>B. bovis</i> hypothetical protein [BBOV II006940]	XP_001610212.1	1
84	1180	215	2e-15	49/178 (27%)	<i>B. bovis</i> hypothetical protein [BBOV III000080]	XP_001611145.1	2
85	420	62	1e-03	11/26 (42%)	<i>B. bovis</i> hypothetical protein [BBOV III000330]	XP_001611168.1	2
86	1163	119	3e-04	36/124 (29%)	<i>B. bovis</i> hypothetical protein [BBOV III000920]	XP_001611227.1	6
87	680	148	5e-08	26/77 (33%)	<i>B. bovis</i> hypothetical protein [BBOV III002570]	XP_001611390.1	113
88	673	137	9e-07	34/105 (32%)	<i>B. bovis</i> hypothetical protein [BBOV III002580]	XP_001611391.1	3

IDENTIFICACIÓN DE GENES EN *B. bigemina* MEDIANTE ETIQUETAS DE SECUENCIA EXPRESADAS

EST	L	S	E	Identidad (%)	Anotación	Gen ID	No. EST
89	126	80	1e-03	17/28 (60%)	<i>B. bovis</i> hypothetical protein [BBOV III002660]	XP_001611398.1	1
90	654	160	9e-38	81/167 (48%)	<i>B. bovis</i> hypothetical protein [BBOV III003290]	XP_001611461.1	1
91	1117	687	1e-79	137/220 (62%)	<i>B. bovis</i> hypothetical protein [BBOV III003960]	XP_001611527.1	8
92	1183	1083	5e-116	226/365 (61%)	<i>B. bovis</i> hypothetical protein [BBOV III004710]	XP_001611602.1	9
93	1185	1153	6e-126	220/299 (73%)	<i>B. bovis</i> hypothetical protein [BBOV III005370]	XP_001611668.1	4
94	403	99	1e-19	51/122 (41%)	<i>B. bovis</i> hypothetical protein [BBOV III005730]	XP_001611704.1	1
95	645	321	4e-28	60/125 (48%)	<i>B. bovis</i> hypothetical protein [BBOV III006200]	XP_001611749.1	16
96	476	390	2e-36	79/138 (57%)	<i>B. bovis</i> hypothetical protein [BBOV III006650]	XP_001611794.1	22
97	732	649	5e-66	115/184 (62%)	<i>B. bovis</i> hypothetical protein [BBOV III006700]	XP_001611799.1	20
98	617	340	2e-30	84/195 (43%)	<i>B. bovis</i> hypothetical protein [BBOV III006760]	XP_001611805.1	1
99	305	154	2e-36	68/88 (77%)	<i>B. bovis</i> hypothetical protein [BBOV III007050]	XP_001611834.1	1
100	250	124	1e-05	21/36 (58%)	<i>B. bovis</i> hypothetical protein [BBOV III007870]	XP_001611915.1	3
101	282	75.9	1e-12	42/84 (50%)	<i>B. bovis</i> hypothetical protein [BBOV III009700]	XP_001612094.1	2
102	822	115	5e-52	48/95 (50%)	<i>B. bovis</i> hypothetical protein [BBOV III010150]	XP_001612139.1	1
103	413	258	4e-21	52/107 (48%)	<i>B. bovis</i> hypothetical protein [BBOV III010780]	XP_001612198.1	33
104	373	193	2e-13	39/60 (65%)	<i>B. bovis</i> hypothetical protein [BBOV III011120]	XP_001612232.1	3
105	523	127	4e-28	54/136 (39%)	<i>B. bovis</i> hypothetical protein [BBOV III011230]	XP_001612243.1	1
106	178	91	1e-03	16/35 (45%)	<i>B. bovis</i> hypothetical protein [BBOV III011620]	XP_001612282.1	1
107	1128	576	3e-57	113/217 (52%)	<i>B. bovis</i> hypothetical protein [BBOV III011660]	XP_001612286.1	13
108	1180	210	8e-15	50/175 (28%)	<i>B. bovis</i> hypothetical protein [BBOV III011910]	XP_001612311.1	2
109	1177	848	8e-89	192/299 (64%)	<i>B. bovis</i> hypothetical protein [BBOV IV000390]	XP_001609207.1	16
110	983	681	1e-79	161/243 (66%)	<i>B. bovis</i> hypothetical protein [BBOV IV001930]	XP_001609358.1	3
111	627	129	2e-28	65/122 (53%)	<i>B. bovis</i> hypothetical protein [BBOV IV002360]	XP_001609401.1	1
112	1209	309	3e-26	115/381 (30%)	<i>B. bovis</i> hypothetical protein [BBOV IV002700]	XP_001609435.1	3
113	186	66	1e-03	14/20 (70%)	<i>B. bovis</i> hypothetical protein [BBOV IV003250]	XM_001609439.1	1
114	1070	390	9e-36	88/189 (46%)	<i>B. bovis</i> hypothetical protein [BBOV IV003920]	XP_001610321.1	3
115	309	69	1e-03	14/38 (36%)	<i>B. bovis</i> hypothetical protein [BBOV IV004440]	XP_001610373.1	1
116	281	325	6e-29	56/68 (82%)	<i>B. bovis</i> hypothetical protein [BBOV IV006610]	XP_001610585.1	5
117	629	163	7e-10	31/108 (28%)	<i>B. bovis</i> hypothetical protein [BBOV IV007590]	XP_001610681.1	3
118	1239	404	3e-37	88/173 (50%)	<i>B. bovis</i> hypothetical protein [BBOV IV008820]	XP_001610804.1	1
119	1239	540	6e-53	120/173 (69%)	<i>B. bovis</i> hypothetical protein [BBOV IV008850]	XP_001610807.1	1
120	1201	202	2e-42	340/492 (69%)	<i>B. bovis</i> hypothetical protein [BBOV IV008860]	XM_001610758.1	1
121	714	305	2e-81	160/194 (82%)	<i>B. bovis</i> hypothetical protein [BBOV IV009120]	XP_001610834.1	2
122	226	80	1e-03	14/28 (50%)	<i>B. bovis</i> hypothetical protein [BBOV IV011180]	XP_001611038.1	10
123	613	151	2e-64	74/86 (86%)	<i>B. bovis</i> hypothetical Proteolipid subunit C	XP_001609797.1	1

EST	L	S	E	Identidad (%)	Anotación	Gen ID	No. EST
					[BBOV_II002740]		
124	705	265	2e-69	135/232 (58%)	<i>B. bovis</i> hypothetical P-Type ATPase 4 [BBOV_010020]	XP_001610924.1	1
125	745	182	3e-44	95/188 (50%)	<i>B. bovis</i> hypothetical ubiquitin fusion degradation protein [BBOV_II005920]	XP_001610110.1	1
126	243	38.9	1e-03	9/11 (81%)	<i>B. bovis</i> hypothetical vacuolar sorting protein 28 [BBOV_IV010370]	XP_001610958.1	1
127	174	80	3e-10	46/50 (92%)	<i>B. bovis</i> iron-dependent superoxide dismutase [BBOV_IV011480]	XP_001611068.1	2
128	1180	493	2e-91	100/160 (62%)	<i>B. bovis</i> long-chain acyl-CoA synthetase [BBOV_III010400]	XP_001612163.1	3
129	232	41.2	1e-03	7/15 (46%)	<i>B. bovis</i> Mak16 family protein [BBOV_III001370]	XP_001611272.1	2
130	245	124	1e-05	27/63 (42%)	<i>B. bovis</i> membrane protein [BBOV_III008720]	XP_001611998.1	12
131	906	770	7e-80	161/300 (53%)	<i>B. bovis</i> NADH dehydrogenase [BBOV_I004980]	XP_001609147.1	2
132	449	615	1e-62	118/136 (86%)	<i>B. bovis</i> nucleolar protein NOP5 [BBOV_II006040]	XP_001610122.1	16
133	1064	120	2e-04	49/229 (21%)	<i>B. bovis</i> p18 protein [BBOV_II002890]	XP_001609812.1	1
134	214	74	1e-03	19/51 (37%)	<i>B. bovis</i> papain family cysteine protease protein [BBOV_III010070]	XP_001612131.1	1
135	379	397	3e-37	73/108 (67%)	<i>B. bovis</i> phosphoglycerate mutase 1 family protein [BBOV_III007860]	XP_001611914.1	3
136	404	167	1e-10	48/128 (37%)	<i>B. bovis</i> protein kinase domain containing protein [BBOV_III010080]	XP_001612132.1	1
137	650	178	3e-43	58/124 (46%)	<i>B. bovis</i> protein kinase domain containing protein [BBOV_I001950]	XP_001608847.1	1
138	527	841	1e-88	158/172 (91%)	<i>B. bovis</i> protein kinase domain containing protein [BBOV_III005470]	XP_001611678.1	1
139	1153	1267	2e-138	232/286 (81%)	<i>B. bovis</i> protein kinase domain containing protein [BBOV_III005740]	XP_001611705.1	12
140	950	102	6e-20	21/27 (77%)	<i>B. bovis</i> pyruvate kinase family protein [BBOV_III010130]	XP_001612137.1	1
141	156	168	1e-10	28/34 (82%)	<i>B. bovis</i> RNA recognition motif - containing protein [BBOV_IV004150]	XP_001610344.1	1
142	1069	1077	2e-115	223/318 (70%)	<i>B. bovis</i> RNB-like domain containing protein [BBOV_I003540]	XP_001609004.1	9
143	819	524	1e-65	100/176 (56%)	<i>B. bovis</i> root hair defective 3 GTP binding protein [BBOV_II001090]	XP_001609636.1	3
144	295	314	1e-27	92/98 (93%)	<i>B. bovis</i> rRNA pseudouridine synthase protein [BBOV_II005520]	XP_001610071.1	1

IDENTIFICACIÓN DE GENES EN *B. bigemina* MEDIANTE ETIQUETAS DE SECUENCIA EXPRESADAS

EST	L	S	E	Identidad (%)	Anotación	Gen ID	No. EST
145	1077	427	6e-40	94/177 (53%)	<i>B. bovis</i> Sec 24 protein transport protein [BBOV_IV000740]	XP_001609240.1	9
146	569	380	8e-35	73/82 (89%)	<i>B. bovis</i> senescence-associated protein [BBOV_IV006360]	XP_001610563.1	41
147	875	106	1e-03	29/74 (39%)	<i>B. bovis</i> serine/threonine protein kinase [BBOV_IV008960]	XP_001610818.1	2
148	128	169	8e-11	30/42 (71%)	<i>B. bovis</i> SET domain containing protein [BBOV_IV004680]	XP_001610397.1	4
149	1000	700	1e-71	128/291 (43%)	<i>B. bovis</i> spherical body protein 3 [BBOV_I004210]	XP_001609070.1	77
150	760	41	1e-03	27/29 (93%)	<i>B. bovis</i> tetratricopeptide repeat domain protein [BBOV_IV000430]	XM_001609161.1	1
151	327	68.9	3e-11	36/80 (45%)	<i>B. bovis</i> translation initiation factor 3 subunit 10 [BBOV_IV005090]	XP_001610438.1	1
152	466	428	7e-41	79/90 (87%)	<i>B. bovis</i> translation initiation factor 3 subunit 2 [BBOV_IV003330]	XP_001609497.1	88
153	1198	1286	2e-139	262/311 (89%)	<i>B. bovis</i> tubulin gamma chain [BBOV_IV007250]	XP_001610649.1	20
154	177	89	1e-03	19/37 (51%)	<i>B. bovis</i> U3 small nucleolar ribonucleoprotein [BBOV_IV001330]	XP_001609298.1	1
155	777	1060	1e-113	192/258 (74%)	<i>B. bovis</i> ubiquitin fusion degradation protein [BBOV_II005920]	XP_001610110.1	8
156	236	89	0.14	20/53 (37%)	<i>B. bovis</i> vacuolar sorting protein 28 [BBOV_IV010370]	XP_001610958.1	2
157	524	61	8.5	14/43 (32%)	<i>B. bovis</i> VESA-1, alpha subunit [BBOV_I001430]	XM_001608745.1	1
158	219	34.3	3.3	10/27 (37%)	<i>B. bovis</i> VESA 1 alpha subunit [BBOV_I005300]	XP_001608620.1	1
159	538	40	0.081	9/22 (40%)	<i>B. bovis</i> VESA 1 alpha subunit [BBOV_II000370]	XP_001609565.1	1
160	293	38.5	0.18	17/31 (54%)	<i>B. bovis</i> VESA 1 alpha subunit [BBOV_III002320]	XP_001611366.1	1
161	872	79	0.001	26/84 (32%)	<i>B. bovis</i> VESA-1 family protein [BBOV_III003060]	XP_001611438.1	2
162	153	210	1e-15	39/48 (81%)	<i>B. bovis</i> vesicle-associated membrane protein [BBOV_II001250]	XP_001609652.1	2
163	427	201	2e-50	67/105 (63%)	<i>B. bovis</i> WD domain G-beta domain protein [BBOV_III009950]	XP_001612119.1	2
164	1184	660	7e-67	123/186 (66%)	<i>B. bovis</i> ZPR1 zinc-finger domain protein [BBOV_IV004120]	XP_001610341.1	3
165	405	90	0.11	15/30 (50%)	<i>Babesia</i> sp. WA1 unknown	gb AAY83301.1	7
166	291	63	3	14/47 (29%)	<i>C. muris</i> RN66 hypothetical protein	XP_002139854.1	1
167	405	62	4.8	13/29 (44%)	<i>C. muris</i> RN66EF hand family protein	XP_002142486.1	1
168	153	60	6.6	12/19 (63%)	<i>C. parvum</i> hypothetical protein	XP_626156.1	1
169	387	87	0.74	27/78 (34%)	<i>C. parvum</i> translation initiation factor eif-2b epsilon subunit	emb AM269894.1	1
170	269	42.4	0.015	30/34 (88%)	<i>P. berghei</i> hypothetical protein	XP_669154.1	2
171	378	83	0.014	17/23 (73%)	<i>P. berghei</i> hypothetical protein	XP_671654.1	1

EST	L	S	E	Identidad (%)	Anotación	Gen ID	No. EST
172	192	80	1.5	16/23 (69%)	<i>P. berghei</i> hypothetical protein PB000102.02.0	XP_674265.1	1
173	238	77	3.2	15/19 (78%)	<i>P. berghei</i> hypothetical protein PB000724.01.0	XP_669154.1	1
174	515	108	0.001	18/18 (100%)	<i>P. berghei</i> hypothetical protein PB102643.00.0	XP_668857.1	1
175	1174	194	1e-47	90/91% (98%)	<i>P. chabaudi chabaudi</i> hypothetical protein	XP_731877.1	3
176	203	62	3.9	14/45 (31%)	<i>P. chabaudi chabaudi</i> hypothetical protein	XP_745544.1	1
177	281	63	3	20/40 (50%)	<i>P. chabaudi chabaudi</i> hypothetical protein	XP_741065.1	1
178	64	70	4.4	13/17 (76%)	<i>P. falciparum</i> 3D7 chromosome 14 hypothetical protein	gb AE014187.1	1
179	234	66	1.4	15/61 (24%)	<i>P. falciparum</i> 3D7 DNA-dependent RNA polymerase	XP_001347935.1	1
180	579	44	6.3	30/34 (88%)	<i>P. falciparum</i> 3D7 hypothetical protein	XM_001350366.1	6
181	518	68	2e-06	52/64 (81%)	<i>P. falciparum</i> 3D7, phosphatase 1 regulatory subunit	emb AL929352.1	1
182	219	61	5	12/42 (28%)	<i>P. falciparum</i> erythrocyte membrane protein 1 Gb174var7	gb AAN86297.1	1
183	500	44	5.4	22/22 (100%)	<i>P. falciparum</i> erythrocyte membrane protein 1 type k (var) gene	gb AF084584.1	1
184	752	44	8.3	25/27 (92%)	<i>P. falciparum</i> var13.1, var13.2 and partial rif genes	emb AM411451.1	1
185	117	119	4e-05	22/24 (91%)	<i>P. knowlesi</i> 40S ribosomal protein S16 putative	emb CAQ42027.1	1
186	226	65	0.55	16/37 (43%)	<i>P. knowlesi</i> hypothetical protein	XM_002260970.1	1
187	391	69	0.64	15/43 (34%)	<i>P. knowlesi</i> hypothetical protein, conserved	XP_002261416.1	1
188	455	73	9.8	18/51 (35%)	<i>P. vivax</i> beta-ketoacyl-acyl carrier protein synthase III precursor	XP_001612931.1	15
189	786	77	0.35	17/42 (40%)	<i>P. vivax</i> hypothetical protein	XP_001615813.1	1
190	602	61	4.9	11/16 (68%)	<i>P. vivax</i> SaI-1 hypothetical protein	XM_001614448.1	1
191	418	157	2e-09	30/34 (88%)	<i>P. vivax</i> Senescence-associated protein	XP_001614866.1	2
192	425	127	5e-06	24/29 (82%)	<i>P. yoelii yoelii</i> hypothetical protein PY04651	XP_724980.1	2
193	418	120	4e-05	23/27 (85%)	<i>P. yoelii yoelii</i> str. 17XNL hypothetical protein	XP_727233.1	1
194	1105	54	0.024	32/35 (91%)	<i>P. yoelii yoelii</i> str. 17XNL hypothetical protein (PY00343)	XM_726344.1	1
195	418	124	1e-05	24/29 (82%)	<i>P. yoelii yoelii</i> str. 17XNL senescence-associated protein	XP_729762.1	2
196	494	731	6e-76	142/162 (87%)	<i>Th. annulata</i> 40S ribosomal protein S9	XP_952982.1	1
197	517	68.9	3e-14	35/89 (39%)	<i>Th. annulata</i> DNA binding protein, similar to Myb	XP_953716.1	1
198	732	450	5e-43	77/178 (43%)	<i>Th. annulata</i> eukaryotic translation initiation factor 3 subunit II	XP_952417.1	1
199	1169	216	2e-15	43/21 (35%)	<i>Th. annulata</i> histone acetyltransferase, GCN5-like	XP_952392.1	4
200	157	59	8.6	9/17 (52%)	<i>Th. annulata</i> hypothetical protein	XP_955051.1	1
201	636	58.2	4e-07	12/22 (54%)	<i>Th. annulata</i> hypothetical protein	XP_951819.1	1
202	222	62	1.4	10/14 (71%)	<i>Th. annulata</i> hypothetical protein	XM_948878.1	2
203	127	63	3	14/38 (36%)	<i>Th. annulata</i> hypothetical protein	XP_954205.1	1

IDENTIFICACIÓN DE GENES EN *B. bigemina* MEDIANTE ETIQUETAS DE SECUENCIA EXPRESADAS

EST	L	S	E	Identidad (%)	Anotación	Gen ID	No. EST
204	977	442	2e-49	86/161 (53%)	<i>Th. annulata</i> hypothetical protein TA16775	XP 001612286.1	1
205	176	98	6e-04	18%41 (43%)	<i>Th. annulata</i> hypothetical protein TA18015	XM 950338.1	1
206	561	128	0.001	25/73 (34%)	<i>Th. annulata</i> molecular chaperone	XP 952378.1	1
207	426	596	7e-150	335/357 (93%)	<i>Th. parva</i> 28S ribosomal RNA gene	gb AF218825.1	9
208	457	133	4e-30	56/131 (42%)	<i>Th. parva</i> heat shock protein DnaJ	XP 764324.1	1
209	752	87	1e-15	46/92 (50%)	<i>Th. parva</i> hypothetical protein, RNA helicase	XP 766669.1	1
210	578	292	3e-150	334/353 (94%)	<i>Th. parva</i> large subunit ribosomal RNA gene	gb AF013419.1	5
211	277	69	0.61	14/36 (38%)	<i>Th. parva</i> vacuolar sorting protein 35	XP 766315.1	1
212	115	66	1.4	15/34 (44%)	<i>T. gondii</i> conserved hypothetical protein	gb EEE24881.1	1
213	203	61	5.1	20/60 (33%)	<i>T. gondii</i> conserved hypothetical protein	gb EEE34889.1	1
214	952	66	8.7	16/47 (34%)	<i>T. gondii</i> conserved hypothetical protein	gb EEE34069.1	1
215	124	60	0.76	11/23 (47%)	<i>T. gondii</i> ME49 DEAD/DEAH box helicase	XM 002364494.1	6
216	481	35.8	1.2	21/59 (35%)	<i>T. gondii</i> ME49 glycerol-3-phosphate acyltransferase	gb EEA98616.1	1
217	375	66	1.3	19/50 (38%)	<i>T. gondii</i> ME49 hypothetical protein	XP 002369653.1	1
218	525	75	7.6	20/55 (36%)	<i>T. gondii</i> ME49 hypothetical protein TGME49 059230	gb EEA97996.1	2
219	987	91	0.39	26/79 (32%)	<i>T. gondii</i> ME49 hypothetical protein TGME49 110600	gb EEA97165.1	3
220	1039	84	2.7	25/78 (32%)	<i>T. gondii</i> ME49 hypothetical protein TGME49 113850	gb EEA97454.1	3
221	439	74	7.5	20/63 (31%)	<i>T. gondii</i> ME49 hypothetical protein, conserved	gb EEB04537.1	2
222	466	37.4	0.39	27/70 (38%)	<i>T. gondii</i> putative ubiquitin transferase	CAJ20599.1	1
223	402	68	0.63	14/27 (51%)	<i>T. gondii</i> structural maintenance chromosomes protein	XM 002371159.1	1
224	376	83	0.7	33/117 (28%)	<i>T. gondii</i> SWIM zinc finger domain-containing protein	gb EEB04757.1	2

EST= Etiqueta de Secuencia Expresada representativa de secuencia en el clúster o secuencia única / *Expressed Sequence Tag, representative of a sequence in a cluster or a single copy sequence.*

L = Longitud de secuencia en el clúster o secuencia única, expresada en número de bases / *Sequence length in a cluster or single copy sequence, expressed as base number.*

S = *Score*, valor de alineamiento / *alignment value.*

E = Valor Esperado (significancia estadística del alineamiento) / *“Expected” value (alignment statistical significance).*

Gen ID = Número de acceso del Genbank (<http://www.ncbi.nlm.nih.gov/blast/>) / *Genbank Access number.*