Korean J Vet Res(1999) 39(4) : 797~805

Analysis of partial cDNA sequence from *Theileria sergenti*

Jin-ho Park, Joon-seok Chae*, Dae-hyuk Kim**, Yong-suk Jang**, Oh-deog Kwon, Joo-mook Lee

College of Veterinary Medicine, Chonbuk National University
*School of Veterinary Medicine, UC Davis, USA**
*Faculty of Biological Sciences, Chonbuk National University, Korea***
(Received Jul 26, 1999)

Abstract : *T sergenti* cDNA library were constructed to get a more broad information about the structural, functional or antigenic properties of the proteins, and analyzed for their partial cDNA sequences and expression sequences tags(ESTg). The mRNA were purified from *T sergenti* isolates to identify the information of antigen gene, then first and second strand cDNA was synthesized. *EcoR* I adaptor ligation and *Xho* I enzyme restriction were used to the synthesized cDNA, and ligated into a Uni-ZAP XR vector. *T sergenti* cDNA library was constructed with packaging and amplification *in vitro*. Antibody screening was performed with constructed *T sergenti* cDNA library using antisera against *T sergenti*. Among those clones, eight phagemids were rescued from the recombinant *in vivo* excision with f1 helper phage. Using the analysis of endonuclease restriction and PCR, the recombinant cDNA were proved having a 0.5-3.0kb of inserts. The eight of partial cDNA clones' sequences were obtained and examined for their homology using BLASTN and BLASTX. The eight of sequenced clones were classified into three groups according to the basis of database searches. A total 3,045bp of partial cDNA sequence were determined from six clones. The putatively identified clones contain a cytochrome c gene, a heat shock protein gene, a cyclophilin gene, and a ribosomal protein gene. The unidentified clones have a homology to ATP-binding protein(mtrA) gene of *S argillaceus*, DNA-binding protein(DBP) gene of *Pseudorabies virus*, 85kDa merozoite protein gene of *B bovis*, mRNA spm1 protein of *T annulata* and glycine-rich RNA-binding protein mRNA of *O sativa* etc.

Key words : *Theileria sergenti*, cDNA library, expression sequences tags(ESTg).

This study was financially supported in part by a Research Grant from Bio-Safery Research Institute, Chonbuk National University in 1999(CNU-BSRI, No. 99-03A).

Address reprint requests to Dr. Joo-mook Lee, College of Veterinary Medicine, Chonbuk National University, 561-756, Chonju, Republic of Korea. E-mail : jmlee@moak.chonbuk.ac.kr

Introduction

Cloning, sequencing, and expression of genes encoding antigenic proteins have allowed examination of polypeptide function and detailed analysis of epitope conservation in light of genetic polymorphism. Many of researchers have focused their attention in this area, because analysis of cDNA libraries plays an important role in genome project[1]. Partial sequencing of randomly selected cDNA clones generates many expressed sequence tags(ESTs) which can be used for genome mapping and those of sequence tagged sites(STSs) are also becoming standard markers in genome mapping[2,3].

ESTs can provide the opportunity to analyze expression levels of functional genes in different tissue and under various conditions. Many informations of antigenic protein gene on *Theileria annulata*, *T buffeli*, and *Babesia bovis* were identified, but few in *T sergenti*. EST for *T sergenti* was conducted to generate expression profiles which can give us molecular information on the putative antigenic protein genes.

By the way, *T sergenti* membrane proteins(MP) have been considered as major recombinant vaccine candidates, since they have been identified as antigenic proteins and inducers for specific antibodies. The information of major immunoreactive membrane protein gene such as 32/34kDa gene in Japan's ikeda or shitose stock types were already reported, but the similar information is not available yet in this country's stock types.

In this study, the cDNA library of *T sergenti* was constructed for EST analysis which render us to have antigenic protein genes for the establishment of an effective control of theileriosis. Constructed *T sergenti* cDNA library were screened using antisera against *T sergenti*. Among the strongly reacted clones, plasmid DNA of recombinant phagemid were analysed using the endonuclease restriction and PCR. The predicted amino acid structures of inserted *T sergenti* mRNA were analyzed by nucleic acid sequencing with the computer-aid DNAsis program and searched for homology sequence in the pre-existing data library.

Materials and Methods

Purification of poly(A)$^+$ RNA : Total RNA was extracted from isolated *T sergenti* by the guanidinium isothiocyanate phenol method[4]. Poly(A)$^+$ RNA were purified by chromatography on oligo-dT cellulose column with mRNA purification kit(Pharmacia Biotech, USA)[5]. Briefly, the prepared total RNA were applied to the oligo-dT cellulose spun column and washed with high-salt buffer(10mM Tris-HCl pH 7.4, 1mM EDTA, 0.5M NaCl) once and low-salt buffer(10mM Tris-HCl pH 7.4, 1mM EDTA, 0.1M NaCl) twice, and then the remained poly(A)$^+$ RNA in column were eluted with elution buffer(10mM Tris-HCl pH 7.4, 1mM EDTA). The eluted poly(A)$^+$ RNA were stored at -70℃.

Construction of *T sergenti* cDNA library : A cDNA library was constructed with poly(A)$^+$ RNA using the ZAP-cDNA synthesis kit(Stratagene, USA) by the manufacturer's instruction. First and second cDNA were synthesized using M-MuLV RTase and oligo-dT linker primer. The synthesized cDNA were ligated with *EcoR* I adaptor and digested with *Xho* I, then fractionated by 8% PAGE[6].

The 0.5~3.0kb length of fragments were ligated into Uni-ZAP XR vector arms[7] and tested the ratio of recombination on X-gal indicator plates. Packaging of cDNA library were used with the Gigapack II packging extracts(Stratagene, USA) according to manufacturer's instruction and amplification *in vitro*[8].

Immunoscreening : The cDNA library were plated on a top agar LB plate with a host cell, *E coli* strain MRF' (Stratagene, USA), and grown at 37℃ for 9 hours. The formated plaques were transferred to nitrocellulose filter paper saturated with 10mM IPTG(Promega Co., USA), and reacted with blocking solution(5% skim milk, 0.01% Tween 20, 0.02 % sodium azide in PBS). Then, antisera against *T sergenti* diluted in TBST were added to react with plaques for 90 minute and washed three times with TBST. Anti-bovine IgG AP(Sigma, USA) were applied according to manufacture's instruction. BCIP/NBT(Promega Co., USA) were used as the substrates for AP. Positive clones identified from first screening were rescreened to obtain pure phage clones.

Infection of cDNA clones : Pure phagemid were rescued

from the recombinant *in vivo* excision by the infection of *E coli* strain SOLR(Stratagene, USA) with f1 helper phage (Stratagene, USA). Plasmid DNA were isolated by alkali lysis method[9] and confirmed containing of inserts with *EcoR* I and *Xho* I enzyme digestion. Inserts of plasmid DNA were PCR amplified with T3 and T7 primer in pBluscripts vector (Table 1).

Table 1. Sequences of T3 and T7 primers for PCR amplification and sequencing of recombinant *T sergenti* cDNA library phage

Primer	Length (mer)	Nucleotide sequence 5' to 3'
T3	20	5'-AATTAACCTTCACTAAAGGG-3'
T7	22	5'-GTATTACGACTCACTATAGGGC-3'

DNA sequencing and analysis : DNA sequencing was performed with T3 and T7 primer on Automatic sequencer (ABI 377, USA) that apply multi-color fluorescence detection system. The obtained partial sequence were examined for similarities to the GenBank nucleic acid database using the BLASTN and BLASTX program[10] through the National Center for Biotechnology Information(NCBI, USA). Also, the deduced amino acid sequences were predicted from those sequences.

Results

cDNA library of *T sergenti* : Six micrograms of poly(A)$^+$ RNA was eluted from *T sergenti* isolates on oligo-dT cellulose column and used to construct the cDNA library. The synthesized cDNA having a *EcoR* I /*Xho* I sites were fractionated with a length of 0.5~3.0kb by PAGE, then ligated into Uni-ZAP XR vector arms. The recombinant cDNA library was packaged at 5×10^6 primary plaques, and amplified to a titer of 6×10^9 pfu/ml *in vitro*. In the test of recombination ratio, approximately 70% of recombination phages were resulted as plaques on X-gal indicator plates.

Immunoscreening : For the immunoscreening, 2.5×10^3 recombinant plaques were screened using antiserum against *T sergenti*. Among the phages, eight clones strongly reacted with antiserum were obtained and rescreened.

Identification of antigen gene : Pure phagemid was rescued from the recombinants *in vivo* excision with f1 helper phage to identify antigen genes. Isolated plasmid DNA were digested with *EcoR* I /*Xho* I enzyme(Fig 1) and PCR amplified(Fig 2). As a results, the recombinant cDNA were proved having 0.5~3.0kb of inserts.

Fig 1. Restriction endonuclease analysis of selected recombinant phagemids from *T sergenti* cDNA library with *EcoR* I and *Xho* I . M : 1kb DNA size marker, 1~5 : recombinant phage-mids.

Fig 2. PCR amplification of selected recombinant phagemids from *T sergenti* cDNA library.
M : 100bp DNA size marker 1~9 : amplified recombinant phagemids *T sergenti* cDNA clones using T3 and T7 primer.

Classification of partial cDNA sequence : The eight of partial cDNA clones' sequences were obtained and examined for their homology using the BLASTN and BLASTX. Among the eight sequenced clones, four clones have the

strong homology to already known sequences, two others have moderate to low homology in many different genes which make it difficult to assert as a known gene, and the rest two of them did not have any similarity to the known genes. On the basis of database searches, the partial cDNA sequences were classified into three groups ; putatively identified clones, unidentified clones, and no matched clones (data are not shown). A total 3,045bp of partial cDNA sequence were determined from six putatively identified and unidentified clones. The cDNA clones that show a homology with other gene were named as KTs-cDNA1, KTs-cDNA2, KTs-cDNA3, KTs-cDNA4, KTs-cDNA6, and KTs-cDNA8 respectively. All of the putatively identified partial cDNA sequences and deduced amino acid sequences were showed in following Figures 13~18.

KTs-cDNA1, 703bp of partial cDNA sequence(Fig 3),

```
GCATCGACTT CGACAAGTCT CGCAACATAT CTTGCAACCC AAARCRATCA CCAGCTTTCC
  I D F   D K S   R N I   S C N   P X X S   P A F   60
ACGTCCAGCC ACCTGCAGAA GCTCTCTTCC AGCATCGTTG CGCTCAGTGC CACACTGTCG
 H V Q   P P A E   A L F   Q H R   C A Q C   H T V  120
AGGCTGGTGG TCCCCACAAG GTCGGCCCCA ACTTGCACGG TCTCTTCGGT CGTAAGACCG
 E A G   G P H K   V G P   N L H   G L F G   R K T  180
GTCAGTCCGA GGGCTACGCC TACACCGACG CCAACAAGCA GGCTGGTGTC ACCTGGGATG
 G Q S   E G Y A   Y T D   A N K   Q A G V   T W D  240
AGAACACTCT GTTCTCCTAC CTCGAGAACC CCAAGAAGTT CATCCCCGGT ACCAAGATGG
 E N T   L F S Y   L E N   P K K   F I P G   T K M  300
CCTTCGGTGG TCTCAAGAAG GGCAAGGAGA GGAACGACCT CATCACTTAC CTCAAGGAGA
 A F G   G L K K   G K E   R N D   L I T Y   L K E  360
GCACCGCTTA AATCACCTAC CGTTATGAAC TCTTCTAGAA CCCGCCTGCC GAACATTGCC
 S T A   * I T Y   R Y E   L F *   N P P A   E H C  420
CACTTATTCC GTGCTGGAAA CAGGCGTTGA AGTCGGCAAC CCGACAGTCA GGGGACTGTA
 P L I   P C W K   Q A L   K S A   T R Q S   G D C  480
CATTACCAGC CCTTATTTAC CCGCATACCT TTGGGTTGCT GCGTTGGACA TCAGATCCAA
 X L P   A L I Y   P H T   F G L   L R W T   S D P  540
TAAGTCCTGT CTCCCTCCGC GACCGCTGTC TACATATTCC CTTCTGTTCC CTTTTTTTAC
 I S P   V S L R   D R C   L H I   P F C S   L F L  600
CCCCCATTTC CTAGGTCCTC TAAATGGTTT CTTCCAGCTC TTGTGGTATC AGGTTTTAAA
 P P I   S * V L   * M V   S S S   S C G I   R F *  660
CAATTGGTTG CTTAAGCCAT AATGAATTTG AAAGTCTCCT CTCAAAAAAA AAAAAAAAAA
 T I G   C L S H   N E F   E S L   L S K K   K K K  703
```

Fig 3. The nucleotide and deduced amino acid sequences of selected recombinant phagemid from KTs-cDNA1 library.

Table 2. Analysis of KTs-cDNA1 ESTg on the homology

Genbank No.	Identities	Gaps	Query	Subject	Length	Definition
L19358	169/192(88%)	2/192(1%)	156-346	1057-1247	1465	Neurospora crassa cytochrome c(cyc-1) gene. complete cds including cyc-1-12 mutation.
X05506	169/192(88%)	2/192(1%)	156-346	217-407	665	N. crassa mRNA for cytochrome c
M83141	83/89(93%)		258-346	932-1020	1312	Emericella nidulans cytochrome c(CYTc) gene, complete cds
X58150	48/55(87%)		267-321	229-283	333	K. lactis cytcl gene for cytochrome C
MB5253	82/89(92%)		258-346	561-649	1264	Arabidoposis thaliana AtCc-1 DNA, complete CDs.
X59459	82/89(92%)		258-346	561-649	1263	A. thaliana AtCc-1 gene for cytochrome c
J01318	53/59(89%)		246-704	646-704	846	Yeast(S. pombe) cytochrome c gene and flanks
X60834	48/55(87%)		267-321	600-654	913	K. lactis CYCK gene for cytochrome C
AL049644.1	53/59(89%)		246-304	10743-10801	25000	S. pombe chromosome I cosmid c191
X01760	175/222(78%)		99-320	822-1043	1559	Drosophila melanogaster cytochrome c gene DC4
AC006214	175/222(78%)		99-320	127103-127324	129779	Drosophila melanogaster cytomosome 2L, region 36A7-36A13, P1 clones DSO4680 and DS00592, complete sequence
AC005120	175/222(78%)		99-320	83927-84148	86603	Drosophila melanogaster, chromosome 2L, region 36A7-36A10, P1 clone DS00592, complete sequence
M11381	175/222(78%)		99-320	398-619	1015	D. melanogaster cytochrome C gene
X58249	48/53(90%)		267-319	208-260	315	C. glabrata DNA for cytochrome C
Z21499	94/115(81%)		244-358	240-354	477	S. longipes mRNA for mitochondiral cytochrome c
AF030426	48/54(88%)		267-320	837-890	1200	Pichia stipitis cytochrome c(CYC1) gene. complete CDs

Table 3. Analysis of KTs-cDNA2 ESTg on the homology

Genbank No.	Identities	Gaps	Query	Subject	Length	Definition
U81165	322/378(85%)	11/378(2%)	209-582	892-1262	3480	Podospora anserina suppressor of vegetative incompatibility MOD-E(mod-E) gene, complete cds
M55629	113/129(87%)	1/129(0%)	212-339	545-673	2801	Histoplasma capsulatum beat shock protein 82(hsp 82) gene, complete cds.
X70101	124/146(84%)		220-365	49-194	2633	G. gallus hsp 90 beta mRNA for heat shock protein 90 beta
U57463	111/129(86%)		219-347	299-427	1385	Drosophila melanogaster heat shock protein 83 (Hsp83) gene, partial cds
U57461	111/129(86%)		219-347	299-427	1385	Drosophila melanogaster heat shock protein 83 (Hsp83) gene, partial cds
X03811	111/129(86%)		219-339	1974-2102	3072	Drosophila simulans gene fragment for heat shock protein hsp 82
U57466	111/129(86%)		219-347	299-427	1385	Drosophila melanogaster heat shock protein 83 (Hsp83) gene, partial cds
U57465	111/129(86%)		219-347	299-427	1385	Drosophila melanogaster heat shock protein 83 (Hsp83) gene, partial cds
X03810	111/129(86%)		219-347	2181-2309	5024	Drosophila melanogaster gene for heat shock protein hsp 82
U57464	111/129(86%)		219-347	299-427	1385	Drosophila melanogaster heat shock protein 83 (Hsp83) gene, partial cds
U75687	134/160(83%)		220-379	1069-1228	3313	Drosophila auraria heat shock protein 83(hsp83) gene, complete cds
X00065	110/129(85%)		219-347	2194-2322	3292	Drosophila melanogaster 5' end of heat shock gene hsp83
K01685	110/129(85%)		219-347	2194-2322	3292	D. melanogaster heat shock gene hsp83 (locus 63B) 5' end and flank.
L47285	131/160(81%)		220-379	5293-5452	7587	Anopheles albimanus heat shock protein gene (hsp 82) genes, complete cds
U55859	86/101(85%)		220-320	137-237	2397	Triticum aestivum heat shock protein 80 mRNA, complete cds
M18186	104/126(82%)		238-363	161-286	2509	Mouse 84kD heat shock protein mRNA, complete cds.

Table 4. Analysis of KTs-cDNA3 ESTg on the homology

Genbank No.	Identities	Gaps	Query	Subject	Length	Definition
Z75201	38/42(90%)		173-214	1112-1153	1757	S. cerevisiae chromosome XV reading frame ORF YOR293w
Z54367	33/36(91%)		56-91	55-90	14189	H. sapiens gene for plectin
AD000019	20/20(100%)		437-456	20999-21018	42061	Mycobacterium tuberculosis sequence from clone y223
U43537	20/20(100%)		445-464	3773-3792	3938	Streptomyces argillaceus mithramycin resistance determinant, ATP-binding protein (mtrA) and membrane protein (mtrB) genes, complete cds
AE000607.1	20/20(100%)		282-301	2608-2627	10707	Helicobacter pylori 26695 section 85 of 134 of the complete genome
U80908	19/19(100%)		396-414	1538-1556	3957	Pseudorabies virus DNA-binding protein (DBP) gene, complete cds
M27444	19/19(100%)		30-48	352-370	1691	Bos taurus (clone pTKD7) dopamine and cyclic AMP-regulated neuronal phosphoprotein (DARPP-32) mRNA, complete cds.

```
GCACTCTTAT TCTCTATCAC TCTTCCCTCT CTCTCTACCT CCATCTCCCT TTTCTTTCTG
 A  L  L   F  S  I  T   L  P  S   L  S  T   S  I  S  L   F  F  L   60
TCAACTCGTG GGCAGTTCTT AATCTCCTTT CCTCCTATCT ATCTTGTGTG ATCCGTGACT
 S  T  R   G  Q  F  L   I  S  F   P  P  I   Y  L  V  *   S  V  T  120
TTGCTCCATT TGCCCTCCCA TCAAACCTCC ACACCTTTCA ATCCGTTTTA GAAAACCAAG
 L  L  H   L  P  S  H   Q  T  S   T  P  F   N  P  F  *   K  T  K  180
TCTCTCCTCC TACCGCCACC AGTCTGCCGA GACTTTCGAT TCCAGGCTGA GATCTCTCAG
 S  L  L   L  P  P  P   V  C  R   D  F  R   F  Q  A  E   I  S  Q  240
CTCCTCTCCC TCATCATCAA CACTGTCTAC TCCAACAAGG AGATTTTCCT GCGTGAGCTG
 L  L  S   L  I  I  N   T  V  Y   S  N  K   E  I  F  L   R  E  L  300
ACTCCCAACG CCTCCGATGC TCTGGACAAG ATTCGCTATG AGTCCCTGTC CGACCCCTCT
 T  P  N   A  S  D  A   L  D  K   I  R  Y   E  S  L  S   D  P  S  360
AAGCTCGACT CGGGCAAGGG ATCTCCGCAT CGACCTGATC CCCAACGCCG AGGCCAAGAC
 K  L  D   S  G  K  G   S  P  H   R  P  D   P  Q  R  R   G  Q  D  420
CCTCACCATC CGTGATACCG GTATTCGGTA TGACCAAGGC TGACCTCATC AACAACCTTC
 P  H  H   P  *  Y  R   Y  S  V   *  P  R   L  T  S  S   T  T  F  480
GGTACCATTG CTCGCTCTGG GTACTAAAGC AGTTCATGGG AGGCTCTCTC GCTGGTGCTG
 G  T  I   A  R  S  G   Y  *  S   S  S  W   E  A  L  S   L  V  L  540
ATATCTCATG ATCGGGCAGT TCGGGGTTGG GTTCTACTCT GCTACCCTTG GGCGACCCGG
 I  S  H   D  R  A  V   R  G  W   V  L  L   C  Y  P  W   A  T  R  600
TGTAACCGGT CATTTCAAGC ACAATGG
 C  N  R   S  F  Q  A   Q  W                                       627
```

Fig 4. The nucleotide and deduced amino acid sequences of selected recombinant phagemid from KTs-cDNA2 library.

were putatively identified as a cytochrome c gene, and the identities were 78~93% with other cytochrome c gene from *Arabidoposis thaliana, Drosophila melanogaster* etc(Table 2). KTs-cDNA2, 627bp of partial cDNA sequence(Fig 4), were putatively identified as a heat shock protein gene, and the identities were 81~87% with other heat shock protein gene from *D melanogaster, Histoplasma capsulatum ect* (Table 3). KTs-cDNA3, 648bp of partial cDNA sequence(Fig 5), belong to unidentified clones showed a moderate homology to ATP-binding protein(mtrA) gene of *Streptomyces argillaceus* and DNA-binding protein(DBP) gene of *Pseudorabies virus*. The homology of unidentified clones was restricted to only a short stretch of sequence but it could reached to more than 90%(Table 4). KTs-cDNA4, 116bp of partial cDNA sequence(Fig 6), belongs to unidentified clones showed a

Table 5. Analysis of KTs-cDNA4 ESTg on the homology

Genbank No.	Identities	Gaps	Query	Subject	Length	Definition
M99575	23/23(10%)		1-23	8-30	2026	Babesia bovis 85 kDa merozoite protein gene. complete CDs
V15794	23/23(100%)		1-23	2285-2307	2888	Theileria annulata mRNA for spm1 protein
AF026216	25/25(100%)		1-25	60-84	2313	Mus musculus mitogen-activated protein kinase kinase (MKK7) mRNA, complete cds
AF006627	22/22(100%)		1-22	266-287	4455	Toxoplasma gondii myosin-B mRNA, complete cds
AB002728	24/24(100%)		1-24	2-25	499	Entamoeba histolytica mRNA for elongation factor 1 alpha. partial cds
AB002741	24/24(100%)		1-24	1-24	262	Entamoeba histolytica mRNA for amoebapore C homologue, partial cds
AB002730	24/24(100%)		1-24	2-25	518	Entamoeba histolytica mRNA for ribosomal protein SA(P40) /laminin receptor, partial cds
AB002794	24/24(100%)		1-24	1-25	386	Entamoeba histolytica small subunit rRNA
AF037332	24/24(100%)		1-24	1743-1766	7261	Homo sapiens Eph-like receptor tyrosine kinase hEphB1b (EphB1) mRNA, complete cds
〃	22/22(100%)		1-22	1408-1429	〃	〃
U19482	24/24(100%)		1-24	3-26	1143	Mus musculus C10-like chemokine mRNA, complete cds
AB002739	24/24(100%)		1-24	1-24	428	Entamoeba histolytica small subunit rRNA, partial sequence
U46118	24/24(100%)		1-24	171-194	2101	Rattus norvegicus cytochrome P450 3AO mRNA, complete cds
UO8214	22/22(100%)		1-22	880-901	2370	Rattus sp. DNA binding protein(URE-B1) mRNA, complete cds.
AF090094	24/24(100%)		1-24	837-860	1608	Homo sapiens clone IMAGE 172979
AF098796	24/24(100%)		1-24	984-1007	2368	Mus musculus SLM-1 (Slm1) mRNA, complete cds
AF009411	24/24(100%)		1-24	10-33	821	Oryza sativa glycine-rich RNA-binding protein mRNA, complete cds

```
AGGAACTCGC GACATGCTTA TCCCCAAGGA AGACCGCAAG AAGATCCACG AGTACCTCTT
  G T R     D M L I     P K E D    R K K I    H E Y L    F          60
-------------------------------------------------------------------
CCGCGAGGGT GTGCTCGTGG CCAAGAAGGA CTTCAACCTT CCCAAGCATG GCGACATTGA
 R E G     V L V A    K K D     F N L      P K H G    D I D       120
-------------------------------------------------------------------
CACCAAGAAC CTCTACGTGA TCAAGGCCTG CCAGTCCCTG ACCTCCCGCG GCTACGTCAA
 T K N     L Y V I    K A C     Q S L      T S R G    Y V K       180
-------------------------------------------------------------------
GACCCAGTTC TCCTGGCAGT ACTACTACTA CACCCTCACC CCCGAGGGTC TTGACTACCT
 T Q F     S W Q Y    Y Y Y     T L T      P E G L    D Y L       240
-------------------------------------------------------------------
CCGTGAGTGG CTCCACCTCC CCGCTGAGGT TGTCCCCGCC ACCCACATCA AGCAGCAGCG
 R E W     L H L P    A E V     V P A      T H I K    Q Q R       300
-------------------------------------------------------------------
TTCCCACGCT CCCCCTCGTG GCATGATGGG CGGTGAGGAG CGTGAGCGTC GTCCCCGTGC
 S H A     P P R G    M M G     G E E      R E R R    P R A       360
-------------------------------------------------------------------
TCCTCGTGAG GGTGGCTACC GGCGCGCGAG CAGGACAAGG AGGGCGCCGG CCCCGGCGAG
 P R E     G G Y R    R A S     R T R      R A P A    P A S       420
-------------------------------------------------------------------
TTCGCTCCAG CTTTCGTGGT GGATTCGGCG TGGCCGTGGT GCTCCCTTCT CTAAGGGGTG
 S L Q     L S W W    I R R     G R G      A P F S    K G C       480
-------------------------------------------------------------------
TCGCTGGCCG GTCTTCAAGG GCTTATTGGG AATGGTGCCG GTGTGGCTTG AACGACTTTT
 R W P     V F K G    L L G     M V P      V W L E    R L L       540
-------------------------------------------------------------------
GAAAGAAGTC GGGAAGCTAT GCTGGGCACT TTCTAAAAAA GAAGTTATGC CAAGCTACTG
 K E V     G K L C    W A L     S K K      E V M P    S Y C       600
-------------------------------------------------------------------
TTTCACAGTT ATGACAACGT TAGGATTAAT GGCATGATTT CATGACTCAA AAAAAAAAAA
 F T V     M T T L    G L M     A * F      H D S K    K K K       648
```

Fig 5. The nucleotide and deduced amino acid sequences of selected recombinant phagemid from KTs-cDNA3 library.

```
CTCGTGCCGA ATTCGGCACG AGGGAATGGT TTCGGGATTA TGGAATTAAC GGGGGATTTT
   R A E    F G T    R E W    F R D Y    G I N    G G F      60
-------------------------------------------------------------------
GCTGGTTTAG GTTTCACGGA ATAGCATCTT GAATTGAGAA GCAATCTGGT TATTGCAAAA
C W F R    F H G    I A S    * I E K    Q S G    Y C K      120
-------------------------------------------------------------------
AAAAAAAAAA AAAA
K K K K    K
```

Fig 6. The nucleotide and deduced amino acid sequences of selected recombinant phagemid from KTs-cDNA4 library.

homology to 85kDa merozoite protein gene of *B bovis*, mRNA spm1 protein gene of *T annulata* and glycine-rich RNA-binding protein mRNA of *Oryza sativa* (Table 5). KTs-cDNA6, 339bp of partial cDNA sequence(Fig 7), were putatively identified as a mRNA for cyclophilin gene, and the identities was 86~97% with other mRNA for cyclophilin gene from *Emericella nidulans*, *Trichophyton mentagrophytes*

Table 6. Analysis of KTs-cDNA6 ESTg on the homology

Genbank No.	Identities	Gaps	Query	Subject	Length	Definition
AF107254	36/37(97%)		17-53	788-824	1139	Emericella nidulans cyclophilin B(cypB) gene, complete cds
AJ011956	33/35(94%)		18-52	370-404	573	Malassezia furfur mRNA for potential allergen similar to cyclophilins, strain ATCC no. 42132
X68678	30/31(96%)		16-46	1441-1471	2598	Z. mays gene for cyclophilin
M55021	30/31(96%)		16-46	404-434	792	Zea mays cyclophilin (CyP) rRNA, complete cds.
Z32674	44/51(86%)		2-52	1242-1292	1488	T. niveum(ATCC34921) cptA gene for cyclophilin
AB019518	32/35(91%)		18-52	493-527	1020	Trichophyton mentagrophytes mRNA for cyclophilin, complete cds
X53223	44/51(86%)		2-52	419-469	812	Schizosaccharomyces pombe gene for cyclophilin (peptidyl-prolylcis-trans isomerase)
X74403	27/29(93%)		18-46	600-628	969	P. vulgaris gene for cyclophilin
L29471	27/29(93%)		18-46	434-462	590	Oryza sativa cyclophilin 1(Cyp1) mRNA, complete cds.
Y08320	33/37(89%)		18-54	428-464	765	D. lanta mRNA for cyclophilin
Z15137	36/41(87%)		18-58	2791-2831	3538	Streptomyces chrysomallus orfA, estA, cypA genes
〃	21/22(95%)		85-106	2858-2879	〃	〃
S82440	29/32(90%)		12-43	405-436	1755	cyclophilin (Hemicentrotus pulcherrimus, early pluteus larva, mRNA, 1755 nt)
AB011651	26/28(92%)		74-101	185-212	860	Bipolaris papendorfii gene for Brni, partial CDs
M62398	23/24(95%)		29-52	407-430	707	Drosophila melanogaster CYP-1 protein mRNA, complete CDs.
J03963	31/35(88%)		18-52	663-697	928	N. crassa cytosolic cyclosporin A-binding protein mRNA, complete CDs.
AF126551	22/23(95%)		18-40	410-432	781	Solanum tuberosum subsp. tuberosum cyclophilin mRNA, complete cds
Z14081	27/30(90%)		18-47	79-108	168	N. tabacum mRNA for cyclophilin(partial)
X17692	31/35(88%)		18-52	2227-2261	2719	Neurospora crassa cyclophilin gene

```
GCACCGTTGT CACCTCTTGG CTGGATGGCA AGCACGTCGT CTTCGGTGAG GTTGCCGATG
   T  V  V  T  S  W  L  D  G  K  H  V  V  F  G  E  V  A  D     60
AGCAGTCCAG TCCGGTCGTG AAGGAGATCG AGGCTCTCGG CTCCTCCTCC GGCTCCGTCC
E  Q  S  S  P  V  V  K  E  I  E  A  L  G  S  S  S  G  S  V    120
GCTCTGCCGT CAAGCCCACC ATCGTCGACT GCGGTGAGTT GTAAGCCGTT TTCAAGGATA
R  S  A  V  K  P  T  I  V  D  C  G  E  L  *  A  V  F  K  D    180
ACAGATGGCA AAATTGGGAT ATCGTCCCAC GGGCTGGTGT AAACCAGCAT CGTGAGGGTT
N  R  W  Q  N  W  D  I  V  P  R  A  G  V  N  Q  H  R  E  G    240
AATGGTGGAC AATTCTCTTT ATCCCTCTAC CCGGTGTAAT CTAGAGTGTC CATACCAGTT
*  W  W  T  I  L  F  I  P  L  P  G  V  I  *  S  V  H  T  S    300
GTCAATAAGT TTCCTCCAAA AAAAAAAAAA AAAAAAAAA
C  Q  *  V  S  S  K  K  K  K  K  K                            339
```

Fig 7. The nucleotide and deduced amino acid sequences of selected recombinant phagemid from KTs-cDNA6 library.

```
GCGCCATCTT CACGCTCCGT CAGTCCTTCG CTGTCGTCAA GCCCAACTCC AAGATGGATT
   A  I  F  T  L  R  Q  S  F  A  V  V  K  P  N  S  K  M  D     60
CCGCCAAGCA GCCCGCTAAG CTCGTCAAGG TCACCCGTGT CCTCGGCCGC ACCGGATCCC
S  A  K  Q  P  A  K  L  V  K  V  T  R  V  L  G  R  T  G  S    120
GTGGTGGTGT CACCCAGCTC CGCGTCGAGT TCATGCACGA TACCTCCCGC AGCATCATCC
R  G  G  V  T  Q  L  R  V  E  F  M  H  D  T  S  R  S  I  I    180
GTAACGTCAA GGGCCCAGTC AAGGTCGACG ACATCCTCTG CCTGCTCGAG TCCGAACGCG
R  N  V  K  G  P  V  K  V  D  D  I  L  C  L  L  E  S  E  R    240
AGGCCCGCCG TCTCCGTTAA TTTCCTTTAC CAACTTCGAG TTTTCGTCAA CATCTTTTTC
E  A  R  R  L  R  *  F  P  L  P  T  S  S  F  R  Q  H  L  F    300
CCTCTCGAGG CTATGTACCC TTGCTTGCGT TGGAAATGAT TCTGGGAGGT CGGGAAGGCA
P  S  R  G  Y  V  P  L  L  A  L  E  M  I  L  G  G  R  E  G    360
AAAAAGAACT ATCTTATCTT ATGGGCTGAT GCGGCATCCG GGTTGGTGTG ACTTCGTTGT
K  K  E  L  S  Y  L  M  G  *  C  G  I  R  V  G  V  T  S  L    420
CCTTTTGTGG TTTTCTTATC ATTCCTGTCT GTGTGTTCAG AATCATCTTG CACTACTGGA
S  F  C  G  F  L  I  I  P  V  C  V  F  R  I  I  L  H  Y  W    480
CTTAATATAG AGAGAAAGAA GAAACCAGAG AAAACGAGGC TTTAACCTGA AATTCAAAAA
T  *  Y  R  E  K  E  E  T  R  K  N  E  A  L  T  *  N  S  K    540
AAAAAAAAAA AAAAAACTTG AGGTCCACGG TTCATAACTT GGAATTGAAT TTGGGACGCT
K  K  K  K  K  N  L  R  S  T  V  H  N  L  E  L  N  L  G  R    600
TATTCCAAAG TT
L  F  Q  S  X                                                 612
```

Fig 8. The nucleotide and deduced amino acid sequences of selected recombinant phagemid from KTs-cDNA8 library.

etc(Table 6).

KTs-cDNA8, 612bp of partial cDNA sequence(Fig 8), was putatively identified as a mRNA for ribosomal protein gene, and the identities was 90~100% with other mRNA ribosomal protein gene from *Kluyveromyces lactis* etc(Table 7).

Discussion

Complementary DNA library allowed us to identify the protein genes of *T sergenti*. Information of functions and structures of membrane protein genes such as nucleotide monophosphate kinase and L35 ribosomal protein gene[11], dextran sulphate protein gene[12], casein kinase Ⅱ alpha subunit gene[13] etc, were identified in *T annulata*, *T buffeli*, and *B bovis*, but few in *T sergenti*. Thus, cDNA library of *T sergenti* was constructed, and EST analysis which render us to have antigenic protein genes were performed for the establishment a effective control of theileriosis. Uni-ZAP XR vector was used to construct of cDNA library of *T sergenti* in this study. The selected eight recombinant phagemids confirmed having an insert by enzyme digestion and PCR. Then the sequence of plasmid DNA and its deduced amino acid sequences were analysed.

Among the eight selected clones, four clones that KTs-cDNA1, KTs-cDNA2, KTs-cDNA6, and KTs-cDNA8 were putatively identified as a cytochrome c gene, a heat shock protein gene, a cyclophilin gene, and a ribosomal protein gene respectively. KTs-cDNA3 and KTs-cDNA4 belong to unidentified clones, have a homology to ATP-binding protein (mtrA) gene of *S argillaceus*, DNA-binding protein(DBP) gene of *Pseudorabies virus*, 85kDa merozoite protein gene of *B bovis*, mRNA spml protein gene of *T annulata* and glycine-rich RNA-binding protein mRNA of *O sativa* etc. In protozoon like as *Theileria* spp. or *Babesia spp.*, only a so few of gene were identified[11~16]. This means that molecular research on disease of hemoparasite such as theileriosis or

Table 7. Analysis of KTs-cDNA8 ESTg on the homology

Genbank No.	Identities	Gaps	Query	Subject	Length	Definition
X69582	29/32(90%)		105-136	160-191	360	K. lactis mRNA for ribosomal protein S33
X69583	29/32(90%)		105-136	802-833	1308	K. marxianus gene for ribosomal protein S33
AC001231	25/27(92%)		487-513	83527-83553	135793	Genomic sequence from Human 17, complete sequence 〔Homo sapiens〕
AB008264	19/19(100%)		454-472	9289-9307	79976	Arabidopsis tahliana genomic DNA, chromosome 5, P1 clone : MB02, complete sequence〔Arabidopsis thaliana〕

babesiosis is just began recently.

Among the selected clones, three were seemed to be a membrane protein gene or structural gene, but their functional significance was not clearly defined yet. The number of screened clones in this study is extremely a small portion of whole profile of expressed genes. Thus more clones needed to be analyzed. The research on the structural analysis or functional assay about unknown antigenic proteins must be continually progressed.

Reference

1. Adams MD, Kelley JM, Gocayne JD, *et al*. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252(5013):1651-1656, 1991.
2. Le Provost F, Lepingle A, Martin P. A survey of the goat genome transcribed in the lactating mammary gland. *Mamm Genome*, 7(9):657-666, 1996.
3. Djikeng A, Agufa C, Donelson JE, *et al*. Generation of expressed sequence tags as physical landmarks in the genome of *Trypanosoma brucei*. *Gene*, 221(1): 93-106, 1998.
4. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem*, 162(1): 156-159, 1987.
5. Okayama H, Kawaichi M, Brownstein M, *et al*. High efficiency cloning of full-length cDNA; construction and screening of cDNA expression libraries for mammalian cells. *Methods Enzymol*, 154:3-28, 1987.
6. Sambrook J, Fritsch EF, Maniatis T. Polyacrylamide Gel Electrophoresis. *Molecular Cloning a Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, New York. pp 6.36-6.38.
7. Short JM, Fernandez JM, Sorge JA, *et al*. Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. *Nucleic Acids Res*, 16(15): 7583-7600, 1988.
8. Kretz PL, Reid CH, Greener A, *et al*. Effect of lambda packaging extract mcr restriction activity on DNA cloning. *Nucleic Acids Res*, 17(13): 5409, 1989.
9. Stuber D, Matile H, Garotta G. System for high-level production in *E coli* and rapid purification of Recombinant Proteins. *Immun Method*, 4:121-152, 1990.
10. Altschul SF, Madden TL, Schaffer AA, *et al*. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". *Nucleic Acids Res*, 25: 3389-3402, 1997.
11. Silins GU, Blakeley RL, Riddles PW. Characterisation of genes encoding a nucleoside monophosphate kinase and a L35 ribosomal protein from *Babesia bovis*. *Mol Biochem Parasitol*, 76(1-2): 231-244, 1996.
12. Dalrymple BP, Peters JM, Goodger BV, *et al*. Cloning and characterisation of cDNA clones encoding two *Babesia bovis* proteins with homologous amino and carboxy-terminal domains. *Mol Biochem Parasitol*, 59(2): 181-189, 1993.
13. ole-MoiYoi OK, Sugimoto C, Conrad PA, *et al*. Cloning and characterization of the casein kinase II alpha subunit gene from the lymphocyte-transforming intracellular protozoan parasite *Theileria parva*. *Biochemistry*, 31(27): 6193-6202, 1992.
14. Kairo A, Fairlamb AH, Gobright E, *et al*. A 7.1kb linear DNA molecule of *Theileria parva* has scrambled rDNA sequences and open reading frames for mitochondrially encoded proteins. *EMBO J*, 13(4):898-905, 1994.
15. Mason PJ, Shiels BR, Tait A, *et al*. Sequence and expression of a gene from *Theileria annulata* coding for a 70-kilodalton heat-shock protein. *Mol Biochem Parasitol*, 37(1): 27-35, 1989.
16. Swan DG, McDonald F, Shiels BR. A stage regulated cyclophilin gene from *Theileria annulata*. *Mol Biochem Parasitol*, 81(1):111-114, 1996.