Molecular Cloning and Bioinformatics characterization of ATP-Binding Cassette (ABC) Transporters from Toxoplasma gondii

• Molecular Cloning and Bioinformatics characterization of ATP-Binding Cassette (ABC) Transporters from Toxoplasma gondii

Thesis · November 2015

DOI: 10.13140/RG.2.1.3180.7440

1 author:

Umer Naveed Chaudhry
The University of Edinburgh
55 PUBLICATIONS **153** CITATIONS

Molecular Cloning & Bioinformatics characterization of ATP-Binding Cassette (ABC) Transporters from *Toxoplasma gondii*

Honours Thesis in Molecular Biology, 30 hp (MB701A)

(15/05/2008 -20/09/2008)

Report Version II

Umer Naveed Chaudhary
Molecular Biology
a08chaum@student.his.se

Supervisor

Prof. Dr. Richard Lucius
Dept. of Molecular Parasitology
Humboldt-University, Berlin, Germany
richard.lucius@hu-berlin.de

Nishith Gupta
Dept. of Molecular Parasitology
Humboldt University Berlin Germany
Gupta.Nishith@staff.hu-berlin.de

Examiner

Patric Nilsson
Dept. of Molecular Biology
University of Skövde, Sweden
patric.nilsson@his.se

University of Skövde Sweden

HÖGSKOLAN SKÖVDE

Humboldt-Universität Berlin Germany

Abstract

Title: Molecular Cloning & Expression Analysis of ATP-Binding Cassette (ABC) Transporters from *Toxoplasma gondii*

Dept: Molecular Biology, University of Skövde Sweden

Course: Honours Thesis in Molecular Biology, 30 hp (MB701A)

Author: Umer Naveed Chaudhary

Supervisor: Prof. Dr. Richard Lucius

Toxoplasma gondii is an obligate intracellular protozoan parasite of the subphylum Apicomplexa and causes toxoplasmosis in immunocompromised individuals and infected infants. *T.gondii* is uniquely adapted to the intracellular parasitism, being able to invade and survive in a wide range of cell types. Successful replication of the parasite within its parasitophorous vacuole necessitates substantial membrane biogenesis. The parasite meets its great demand of phospholipids by *de novo* phospholipids synthesis or by scavenging pre synthesized complex lipids and their precursors from the host cell.

T. gondii has 14 chromosomes and 8032 annotated genes currently available in Toxo DB. ATP-Binding Cassette (ABC) transporters represent the largest evolutionarily conserved super family of proteins involved in lipid and drug transport.

The aim of this project was to determine the classification and domain organization of TgABC transporters by using bioinformatics tools and molecular cloning of ABC transporters from *T. gondii* in to yeast vector for their expression in *S.Cerevisiae* (AD-18) mutant strain. For this purpose, six TgABC transporters genes were screen out and classification was done on the basis of their size, exon number, chromosomal location, domain organization and assigned families and subfamilies by amino acid sequence similarity to ABC proteins of other organisms.

Key words: *Toxoplasma gondii*, ATP binding cassette transporter, Resistance, Phospholipids synthesis

Table of contents

Abbreviations

ABC	ATP-Binding Cassette
Tg-ABC	*Toxoplasma gondii* ATP-Binding Cassette
ABC G2	ABC superfamily G member 2
AIDS	Acquired immunodeficiency syndrome
NBD	Nucleotide-binding domain
TMD	Transmembrane domain
PCR	Polymearse Chain Reaction
S.Cerevisiae	Saccharomyces cerevisiae
HIV	Human immunodeficiency virus
GAL	UDP- glucose-4- epimerase
LB	Laria Bartani
DMSO	Dimethlsulfoxid
E. Coli	Escherichia coli
YPD	Yeast peptone dextrose
SD	Synthetic dextrose

1 Introduction

The word Toxoplasma (*toxo* = arc, *plasma* = form) is derived from its crescent shape. *Toxoplasma gondii* is an intracellular protozoan parasite (Smith 1995) which infects humans as well as wide variety of mammals and birds (Hill *et al.,* 1998). The organism was first discovered by Nicolle and Manceaux (1908) as a tissue parasite of gondii (an African rodent) and Darling found it in Man (Subash 1990). The infection has been confirmed in 200 species of mammals including man and in Domestic / Wild felines, which serve as a definitive host (Pedro *et al.,* 2003).

Toxoplasmosis is found throughout the world (Dubey 1999). *T. gondii* can cause severe acquired infection in animals and human being, which may be localized or generalized. Lymphadenitis is the most frequently observed clinical sign (deep cervical nodes). Encephalitis is an important sign of Toxoplasma in later stages. During the 1980,s Toxoplasmic encephalitis in human emerged as a common complication associated with AIDS (Subash, 1990). Congenital infection in animals and pregnant women develop the most serious side effects leading to spontaneous abortion, still birth, birth defects, mummification, neonatal losses or fetal abnormalities. The mechanism of vertical transmission is not yet understood (Remington *et al.,* 1995).

The treatment and prophylaxis of *T. gondii* infection is one of the most serious problem facing all over the world. Approximately 500 million populations are estimated to have antibodies to *T. gondii* infection (Subash Ch.1990). The Studies show that 16% to 40% of the human populations in North America and Great Britain, 50% to 80% of the populations in Europe and Latin America have antibodies against *T. gondii,* indicating that they harbor infection at some time (Pedro *et al.,* 2003).

T. gondii exhibits a predator-prey type life cycle having two phases (Torada 2001). The sexual phase of the infection occurs only in felines when the cysts are ingested by a cat and the parasite multiplies asexually by mergony and then sexually by gametogony to produce oocyst (Pedro *et al.,* 2003). The asexual phase occurs in the intermediate host, when sporulated oocyst penetrates through intestinal wall and spread by haematogenous route. This stage is called tachyzoite that infect tissues and replicate intracellularly. As a result there is formation of tissue cysts containing bradyzoites. (Urquhart *et al.,* 1996).

The source of transmission is the ingestion of vegetables, fruits, water, soil, food contaminated by cat faces, raw or undercooked meat. Flies and Cockroaches may act as a mechanical carrier to transfer oocysts to different variety of foods. Other sources include transplcental transmission, from mother to the offspring through milk, transplantation of organs, transfusion of blood and venereal transmission (Pedro *et al.,* 2003).

Tachyzoites (2 X 7 um) contain a unique cytoskeleton (subpellicular microtubules, conoid), secretory organelles (rhoptries (6-12), micronemes (50-100), dense granules) along with apical complex, endosymbiotic organelles (mitochondria, apicoplast), eukaryotic organelles (nucleus, endoplasmic reticulum, golgi bodies, ribosomes) and specific structures (acidocalcisomes) (Louis 2007). Secretory organelles play an important role in the host cell invasion and maturation of the parasitophorous vacuole.

Dense granules release their contents after invasion has completed, rhoptries release their contents as invasion proceeds and micronemes release their contents during attachment process in parasitophorous vacuole (Black & Boothroyd 2001). The Conoid are thimble shape structure at the extreme apex of the parasite. It is connected with two apical and polar rings. Two microtubules extend from apical ring and terminate within the body of the cell. Polar ring function as a microtubule organization center from which 22 subpellicular microtubules move down toward the parasite body. Apicoplast are responsible for the synthesis of fatty acids, lipoic acid and isoprenoid precursors (Louis 2007).

More than hundred ABC transporters have been described, and each one is specific for a particular substrate including lipids, sugars, peptides, inorganic ions, cell surface components, iron, sulfur, heavy metals and drugs. Although most family members have been described in prokaryotes, the number is rapidly increasing in eukaryotes. The physiological roles of ABC transporters range from transporting nutrients (in prokaryotes) and the efflux of waste products to regulation. Usually they are low capacity, but high affinity transporters, able to transport substrates against the concentration gradient up to more than 10 000 fold. Characteristically ABC transporters are comparatively specific for a particular set of substrates (Alberts. *et al.,* 2002).

The ABC transporter proteins represent the largest super family found in most organisms including bacteria, yeast, plants and protozoa. The members of ABC transporter super family contain a highly conserved ATP-binding cassette, whose amino acid sequence display three major conserved motifs (Walker A & B, ABC signature). ABC transporter has two types of structural/domain organization. Full size ABC proteins consist of four structural domains having two cytoplasmic domains called ATP or nucleotide-binding domain (NBD) and two transmembrane domains (TMD). A number of ABC proteins are half size transporter consisting of single NBD and TMD (Virginie. *et al.,* 2006).

T. gondii has fourteen chromosomes and 8032 annotated genes currently available in Toxo data base (2008). The aim of this project was to determine the classification and domain organization of TgABC transporters by using bioinformatics tools and cloning of ABC Transporters from *T. gondii* in to yeast vector for their expression in *S.Cerevisiae* (AD-18) mutant strain. It was predicated that ABC transporters of *T. gondii* may function as a multi drug efflux pump and lipid transport. For this purpose, six TgABC transporters genes were screen out and classification was done on the basis of their size, exon number, chromosomal location, domain organization (toxo data base, 2008) and assigned families and subfamilies by amino acid sequence similarity to ABC proteins of other organisms using BLAST program (transport classification data base, 2008).

2 Materials and Methods

2.1. Annotation of *T.gondii* ABC Transporters:

Six genes encoding putative ABC transporters of *T.gondii* were identified and screen out from *T.gondii* database (2008). Identification of sorting signal was performed by using the program signalP (2008). In addition human and protozoan ABC sequences were used as queries for Blast searches (transport classification data base, 2008).

2.2. Gene purification from pCDNA 3.1 + Vector:

Six ABC transporters genes were retrieved from pcDNA 3.1 + Vector and simply (Appendix I). The plasmid was absorbed on filter paper, which was cut out to recover the DNA. Tris-HCl was added and filter was hydrate for 5 minutes. After a brief centrifugation supernatant was used to transform in XL-1b competent cells (Stratagene USA) according the laboratory compendium (Sazzad & Magnus, 2008). Purification was done from the culture of bacteria harbouring pCDNA 3.1 + plasmid along with TgABC clones by using Miniprep kit (Analytik Jean Germany).

2.3. Polymearse Chain Reaction (PCR) Amlification:

The pCDNA 3.1 + clones containing TgABC cDNA were used as a templete for gene specific PCR reactions. Master Mix was prepared for PCR having all the reagents and its composition is given in Phusion high fidelity PCR kit (Stratagene USA). The thermocycler was run according to the program given in Phusion high fidelity PCR kit. Optimization of annealing temperature (T_A) was varied between 50°C to 65°C for individual PCR. Gene specific primers (Invitrogen Germany) were designed contain *Not1* restriction site, maximum length of the primers varied between 31 to 32, melting temperature varied between 59°C to 69°C and guanine and cytosine content was 70% to 75% (Primer3: Pick primers from a DNA sequence, 2008) was given in Table 1.

Table1: Description of TgABC Genes Primers (Forward and Reverse).

Assigned Name	Forward Primer	Reverse Primer	TA	TM	Length	GC %
TgABC-A	CTCGCGGCCGC ATGGACGTCGG GCCTCGAGT	CTCGCGGCCGCCT ACCCACGCCCGCG AGC	61 ° C	69 ° C	FR 31 RP 29	FR 74 RP 82
TgABC-B	CTCGCGGCCGC ATGCCCTCGCA TCCTAATT	CTCGCGGCCGCTC AGGACTTCAGTTTA AGAGATTTCTT	57 ° C	60 ° C	FR 30 RP 38	FR 63 RP 50
TgABC-C	CTCGCGGCCGC ATGGACTTTCTG CGCAGACG	CTCGCGGCCGCTC ACGCTTCAAGGCC GAT	57 ° C	63 ° C	FR 31 RP 29	FR 67 RP 68

TgABC-D	CTCGCGGCCGC ATGGCAGACTC TGCCCTC	CTCGCGGCCGCTT AGTTTCGGCGATTC AAGAT	58° C	59 ° C	FR 29 RP 32	FR 72 RP 56
TgABC-E	CTCGCGGCCG CATGAATTCC CTCGCTGGC	CTCGCGGCCGCT CATTGGTTCTCCT TCGTTCG	57 ° C	61°	FR 29 RP 32	FR 68 RP 62
TgABC-F	CTCGCGGCCG CATGGACGCT CCGGACATGA	CTCGCGGCCGCT CAGCGACGAAGC CGGAC	62 ° C	66° C	FR 31 RP 29	FR 70 RP 75

Not1 is the Restriction site of all above primers.

The 1% agarose gel (Carl Roth Germany) was prepared according to the instruction given in the laboratory compendium (Karim and Magnus, 2008) Run the gel at 100-120 mA for 1-2 hours. The separation of DNA fragments was analyzed by UV irradiation of the gel. The PCR product was purified by pure-link PCR purification kit (Invitrogen Germany) and concentration was measured by the Nanodrop method.

2.4. Digestion of ABC transporters genes:

The digestion was performed by *Not1* restriction enzyme (Invitrogen Germany) at 37° C incubation for 2 hour. After digestion, DNA was cleaned with pure-link PCR purification kit (Invitrogen Germany) and the concentration was measured by the Nanodrop method.

2.5. Purification of pESC-Ura Plasmid:

The purification of plasmid was performed from overnight culture of bacteria harbouring plasmid pESC-Ura (Appendix IV). The protocol was given in the plasmid miniprep kit (Invitrogen Germany). Concentration of purified vector was measured by the Nanodrop method.

2.6. Digestion of pESC-Ura plasmid DNA and gel electrophoresis:

Plasmid DNA was digested with *Not1* restriction enzyme at 37° C incubation for 2 hours. Digested plasmid DNA was loaded in the 1% agarose gel along with ladder and visualized under ultraviolet irradiation. The fragment of digested product was cut out from gel and cleaned with gel extraction kit (Invitrogen Germany). The concentration was measured by the Nanodrop method. For symmetric (non directional) cloning, the linearized vector was deposphorylated and purified. Finally the concentration of dephosphorylated plasmid DNA was measured by Nanodrop method.

2.7. Ligation of digested vector (pESC-Ura) and insert:

Not1 digested pESC-Ura and insert DNA were used for ligation at molar ratio of 1:3 and 1:5 at 16^{o} C for 12 hrs. The use of purified water instead of the insert served as a negative control to determine the re-legation efficiency of empty vector.

2.8. Preparation and Transformation of E. Coli XL-1b competent cells:

L.B. media was inoculated with XL-1b cells and incubated overnight at 37^{o} C. The further procedure was done according to the protocol of Stratagene USA. XL-1b competent cells were mixed with ligation mixture. The cells were plated on L.B plate with antibioics ampicilllin at 37^{o}C for 16 hours. All the procedures were done according to the according the laboratory compendium (Sazzad & Magnus 2008).

2.9. Colony PCR for verification of vector with insert:

Colony PCR consists of three steps of denaturation, annealing, extension and total PCR cycles were 30 that was run according to the program given in Dream Taq PCR kit (Fermentas Germany). Individual colonies were examined to verify the presence of inserts containing pESC-URA plasmid by colony PCR. The colony was picked up from the plate and resuspended in HPLC-grade water (Carl Roth Germany). The suspension was used as a template for colony PCR using gene specific primers (Table 1) and flanking vector primers (Table 2). Master Mix was prepared containing Dream Taq PCR kit (Fermentas Germany).

Table 2: Description of pESC –Ura flank specific primer (Forward and Reverse).

pESC-Ura	Forward Primer Sequence (5-3)	Reverse Primer Sequence (5-3)
GAL10	GGTGGTAATCCATGTAATATG	GGCAAGGTAGACAAGCCGACAAC

The PCR products were analysed for the size of gene inserted in to the vector on an agrose gel. The forward and reverse PCR primers flank the multiple cloning sites. Positive bacterial clones with insert were grown overnight in LB media at 37 ºC. Purification was done from the culture of bacteria harbouring pESC-Ura plasmid along with TgABC clones by using Miniprep kit (Analytik Jean Germany).

2.10. Restriction digestion for verification of insert orientation:

pESC-Ura clones containing cDNA was digested at 37^{o} C for 2 hours. Restriction summaries of both genes (New England biolab international, 2008) and vector (Stratagene an agilent technologies division, 2008) are given in Appendix II & III. Digested product was loaded in the agarose gel along with DNA ladder. The product fragments were visualized by ultraviolet irradiation for verification.

2.11. Sequencing of cloned genes:

TgABC-B, TgABC-D and TgABC-F in pESC-Ura were sequenced by Agowa GmbH using GAL10-Fwd and GAL10 Rev Primers.

2.12. Preparation and Expression of Yeast competent cells (S. cerevisiae):

S. cerevisiae AD-18 lacking ABC transporters (Yeast genetic resource center, 2008) was taken from the frozen glycerol stock and streaked on YPD media. The further procedure was done according to the protocol of Stratagene USA. Yeast expression plasmid was added with salmon sperm DNA (Invitrogen Germany). All other procedures were done according to the laboratory compendium (Nishith 2008).

3 Results and Discussion:

Toxoplasmosis is one of the most common infections in men and animals cause by protozoan parasite *T. gondii,* which is responsible for significantly higher morbidity and mortality in both human and warm-blooded animals. Toxoplasmosis is world wide distribution, zoonotic in nature and depending upon the geographic location. 15-85% of the global population can be symptomatically infected (Subash 1990).Toxoplasmosis is also responsible for abortion and congenital defects in human and domestic livestock including sheep, goat, camel, cow and buffaloe (Pedro *et, al.,* 2003).

ABC transporters were first discovered and studied in bacteria in the fifties but in 1976, 170 kDa glycoprotein was identified in the plasma membrane of multidrug resistance cells (MDR) and termed P-glycoprotein. In protozoans, the first Pgp-like ABC transporter was described in the human malaria parasite *P. falciparum* (Pgh1 encoding by Pfmdr1). (Aline *at al.,* 2009).

ABC transporters represent the largest evolutionarily conserved super family of proteins. Virginie and his team had been identified 24 ABC transporters of *T. gondii* and their amino acid sequences exhibit all the distinctive biochemical features of the ABC family members. Fifteen of the ABC transporters of *T. gondii* cluster in to five out of seven Human ABC transporters families; six belongs to ABCB (lipid export, drug and peptide), two to ABCC (drug export and organic anion conjugate), one to each ABC-E (Rnase L inhibitor, translation regulation and antibiotics resistance) and ABCF (regulation of gene expression and drug resistance) and five to ABCG (drug export and resistance). The remaining nine ABC transporters of *T. gondii* included four from ABC-H (energy generating subunit), four from SMC (Structural maintenance of chromosomes) and one member of unclear (Virginie *et, al.,* 2006).

There are two basic aim of our study. Firstly we determined the bioinformatics classification and domain organization of five ABCG and one ABCH family of *T. gondii*

by using bioinformatics tools. Secondly molecular cloning and expression analysis of TgABC transporters genes in Yeast expression vector to find out their role in lipid and drug transport. For this purpose, *T. gondii* ABC transporters genes were cloned in pESC-Ura and expressed in *S. Cerevisia.* These findings are in concomitant with the results of Schmid *et al.,* 2009.

3.1. Classification of *T.gondii* ABC transporters:

A detailed inventory of five ABCG and one ABCH family of *T. gondii* including assigned name, chromosomal localization, sizes, overall structural organization and exons number (*T.gondii* genome data base, 2008) was presented in Table 3.

The exon encoded proteins of ABCG family of *T. gondii* are half transporters that exhibit a unique structural organization having single N-terminal ATP or nucleotide binding domain fused to a single C-terminal set of tranmembrane domain as predicted by intrpro program of *T. gondii* data base (2008) shown in Table 3. it was reported that ABCH family includes members containing only a single ABC domain (no TMD). These ABC domains are often the energy-generating domains of multi component membrane-bound transporters that are either importers (uptake system) or exporters (Virginie *et, al.,* 2006). These results shows contraversy of our results of ABCH family of *T. gondii*, that have single ATP binding domain fused to the single tranmembrane domain as predicted by intrpro program as shown in Table 3 (*T. gondii* data base, 2008). It was also predicted that there is no signal peptide at N terminus of five ABCG and one ABCH family of *T. gondii* (*T. gondii* data base, 2008)).

Table 3: TgABC transporters genes are classified on the basis of assigned families, chromosomal location, size, domain organization, exon number.

Assigned Name	Gene ID	Chromosomal Localization	Size (Nucleotide)	Size (Amino acid)	Domain structural arrangement	Exon
TgABC-A	80.m02179	IX	2649	882	NBD-TMD	7
TgABC-B	583.m05692	XI	2106	701	NBD-TMD*	6
TgABC-C	50.m03178	XII	2946	981	NBD-TMD*	11
TgABC-D	80.m00083	IX	2301	766	NBD-TMD*	10
TgABC-E	80.m02273	IX	2439	812	NBD-TMD*	15
TgABC-F	49.m05729	VI	2385	794	NBD-TMD*	11

* Nucleotide-Binding Domain (NBD), Trans Membrane Domain (TMD)

The homology search against the database of five ABCG and one ABCH family of *T. gondii* were performed using BLAST algorithm assigned ABCG family (represents the second largest family of *T. gondii* having only one subfamily termed as White subfamily) and ABCH family respectively (four members and one of them was included in our study).(transport classification data base, 2008). It was suggested that the amino acid sequencing of five ABCG and one ABCH family of *T. gondii* had closest similarity with the Human ABC superfamily G member 2 (Gene ID Q9UNQ0) genes can be seen in Table 4.

Human ABCG members ABCG1, ABCG5, ABCG8 and ABCG4 have been involved in the regulation of lipid-trafficking mechanisms in macrophages, hepatocytes, intestinal mucosa cells and brain tissue respectively. Human ABCG2 also functions as a multi drug efflux pump in some cancer cells. Its high level of expression in placenta trophoblast cells also suggests a physiological role in lipidic homeostasis and cellular detoxification. ABCG2 is a plasma membrane protein with minor fraction within the intracellular membranes (Virginie. *et al.,* 2006).

Amino acid sequences of all ABC transporters of *T.gondii* showed sequence identity between (25–41%), and similarity between (44–61%), which was consistent with findings obtained for Virginie. *et al.,* 2006. The alignment score of TgABC-D was high as 228 bits and E value of TgABC-C was low as 9e-40. E value indicate that the number of hits see when searching a database of a particular size and alignment score is the sum of the scores specific for each of the aligned pairs of letters and as the value alignment score is high when the similarity will be high.

Table 4: Family and subfamily assigned was based of BLAST results with the Transport classification database.

Assigned Name	Family	Human match (accession no)	Score (Bits)	E value	Identity	Similarity
TgABC-A	*ABCG(white)* *	ABC-G2 (Q9UNQ0)	174	6e-44	25%	44%
TgABC-B	*ABCH*	ABC-G2 (Q9UNQ0)	183	5e-47	41%	59%
TgABC-C	*ABCG(white)* *	ABC-G2 (Q9UNQ0)	160	9e-40	38%	57%
TgABC-D	*ABCG(white)* *	ABC-G2 (Q9UNQ0)	228	3e-60	30%	50%
TgABC-E	*ABCG(white)* *	ABC-G2 (Q9UNQ0)	184	3e-47	38%	61%
TgABC-F	*ABCG(white)* *	ABC-G2 (Q9UNQ0)	186	7e-48	37%	60%

*ABC superfamily G member 2 (ABC-G2)

3.2. Cloning and Expression analysis of TgABC transporters:

TgABC transporters genes were retrieved from pcDNA 3.1+ vector and further used for gene specific PCR reactions. The optimisation of various parameters was done in PCR reaction as described by Henrik and Elie (2007). Annealing temperature was optimised in between 50°C to 65°C (Fig. 1 a, b, c) because at these optimised values higher PCR product was obtained. Generic DNA (ABC transporters genes) was digested with *NotI* restriction enzyme and further used for ligation in pESC-Ura vector.

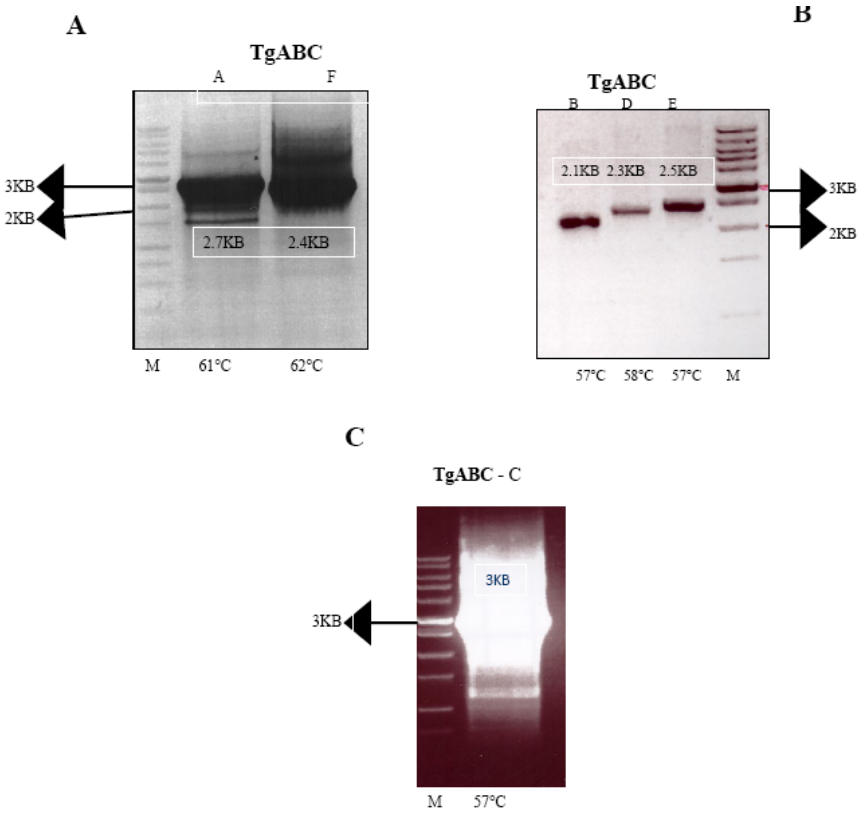


Figure 1: Amplifications with the different specific primers yielded PCR products of the expected size of TgABC. (A) Gel image show the amplification of TgABC-A & F gene product. The size of the product was 2.7 & 2.4 kb respectively and annealing temperature was optimised at 61°C. (B) Gel image show the amplification of TgABC-B, D & E gene product. The size of the product was 2.1, 2.3 & 2.5 kb respectively and annealing temperature was optimised at 57°C. (C) Gel image shows the amplification of TgABC-C gene product. The size of the product was 3kb and annealing temperature was optimised at 57°C.

pESC-Ura yeast expression was digested with *Not1* restriction enzyme along with undigested vector as a control and the degree of digestion was tested (Fig. 2). *Not1* digested pESC-Ura are cloned in to TgABC genes as described by Jung *et al.,* 2008. The use of purified water instead of insert served as a negative control to determine the relegation efficiency of empty vector. There were very few colonies seen in the LB^{AmpR} plate due to deposhorylated vector (pESC-Ura)

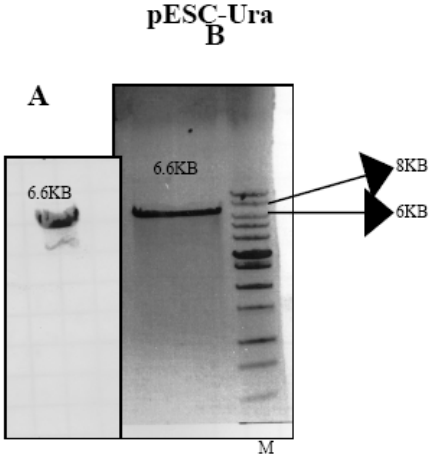
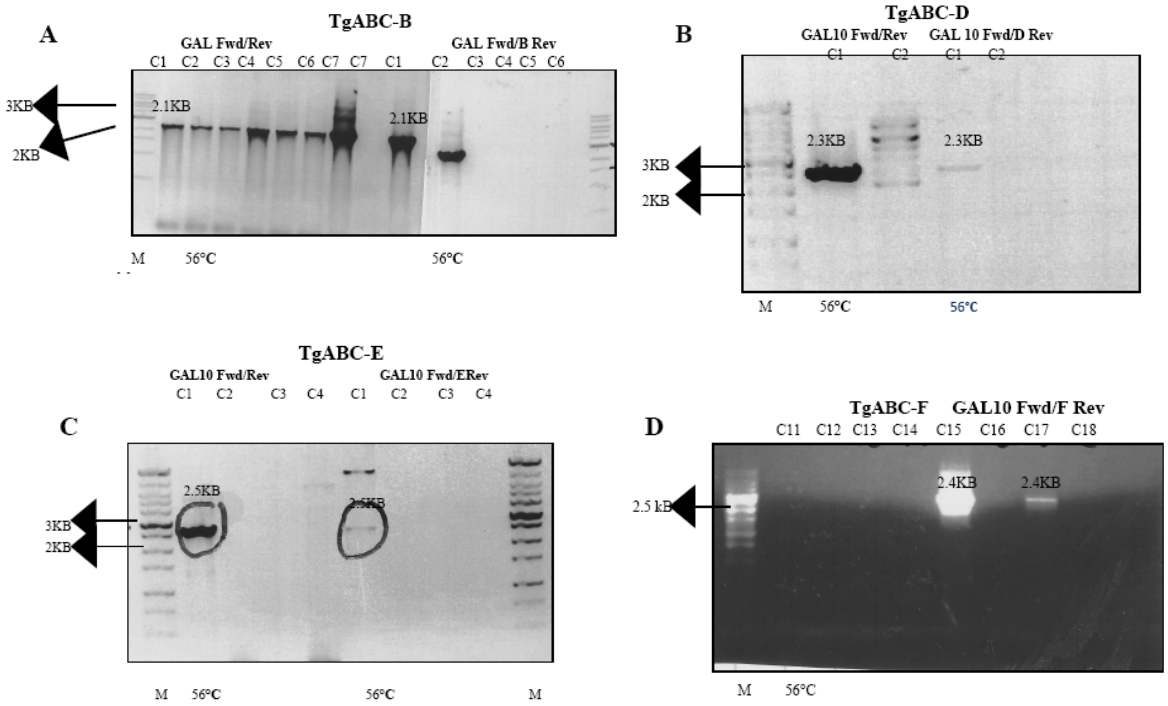


Figure 2: Fragments of pESC-Ura vector were observed. (A) Gel image show the undigested pESC-Ura vector product uses a control. The size of the product was 6.6 Kb. (B) Gel image show the digested pESC-Ura vector product. The size of the product was 6.6 Kb.

Individual colonies were examined to verify the presence of clone by PCR using gene specific primers (Table 1) or flanking vector primers (Table 2). PCR products were analysed for the size of TgABC transporters gene (Fig. 3). These findings are in concomitant with the results of Kong *et al.,* 2005. Seven bacterial colonies were found in TgABC-B LB^{AmpR} plate. TgABC-D showed only two bacterial colonies, TgABC-E revealed four colonies, TgABC-F had twenty bacterial colonies and TgABC-A had two colonies. Clone 1 & 2 of TgABC-B, clone 1 of TgABC-D, clone 1 of TgABC-E, clone 15 and 17 of TgABC-F and clone 1 of TgABC-A demonstrate the success of ligation and transformation of bacteria (Fig. 3).



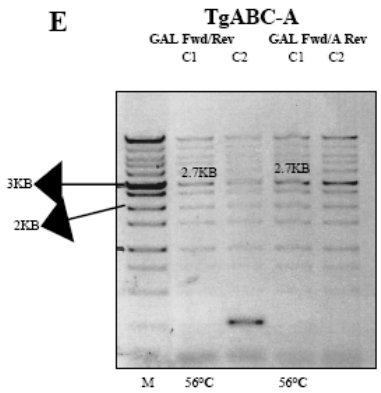


Figure 3: Amplification of different colony PCR products was shown the presence of insert in vector (pESC-URA). (A) The gel image shows the amplification of Clone 1 to 7 of TgABC-B using gene specific primers i.e. B reverse and vector specific primers i.e. GAL10 forward and reverse. Clone 1 & 2 give the positive results. (B) The gel image shows the amplification of Clone 1 & 2 of TgABC-D using gene specific primers i.e. D reverse and vector specific primers i.e. GAL10 forward and reverse. Clone 1 gives the positive results. (C) The gel image shows the amplification of Clone 1 to 4 of TgABC-E using gene specific primers i.e. E reverse and vector specific primers i.e. GAL10 forward and reverse. Clone 1 gives the positive results. (D) The gel image shows the amplification of Clone 1 to 18 of TgABC-F using gene specific primers i.e. F reverse and vector specific primers i.e. GAL10 forward and reverse. Clone 15 & 17 give the positive results. (E) The gel image shows the amplification of Clone 1 & 2 of TgABC-A using gene specific primers i.e. A reverse and vector specific primers i.e. GAL10 forward and reverse. Clone 1 gives the positive results.

Miniprep was performed for bacteria harbouring pESC-Ura vector along with insert and then digested with restriction enzymes for mapping (Fig. 4) as described by Sayada et al., 1995. To test the orientation of TgABC-A, TgABC-B, TgABC-D, TgABC-E & TgABC-F clones, diagnostic restriction mapping was performed in adding a set of restriction enzymes as shown in Table 5.

Table 5: The table shows that TgABC transpoters digest by different restriction enzymes to give the orientation (right or wrong).

Restriction enzymes	Clones	Right orientation	Wrong orientation
NcoI & Bgl II	TgABC-A(pESC-Ura)	7.2Kb & 2.1Kb	8.7Kb & 0.6Kb
Bgl II	TgABC-B(pESC-Ura)	0.6Kb & 8.1Kb	1.6Kb & 7.1Kb
Bgl II	TgABC-D (pESC-Ura)	1.6Kb & 8.3Kb	0.6Kb & 9.3Kb
SacI	TgABC-E (pESC-Ura)	8.15Kb & 0.85Kb	7.3Kb & 1.8Kb
EcoR1	TgABC-F (pESC-Ura)	8.4Kb & 0.6Kb	7.2Kb & 1.8Kb

TgABC-B, TgABC-D and TgABC-F were successfully cloned in pESC-Ura Yeast expression vector as shown in Fig. 4 and further confirmed by commercial sequencing. There was no mutation observed when sequenced DNA (TgABC-B) and amino acid were

aligned (Sequences Lalign, 2008). The alignment of nucleotide and protein sequence of TgABC-B was given in Appendix IV. TgABC-D did not yield any sequencing and were compared due to some unknown error.

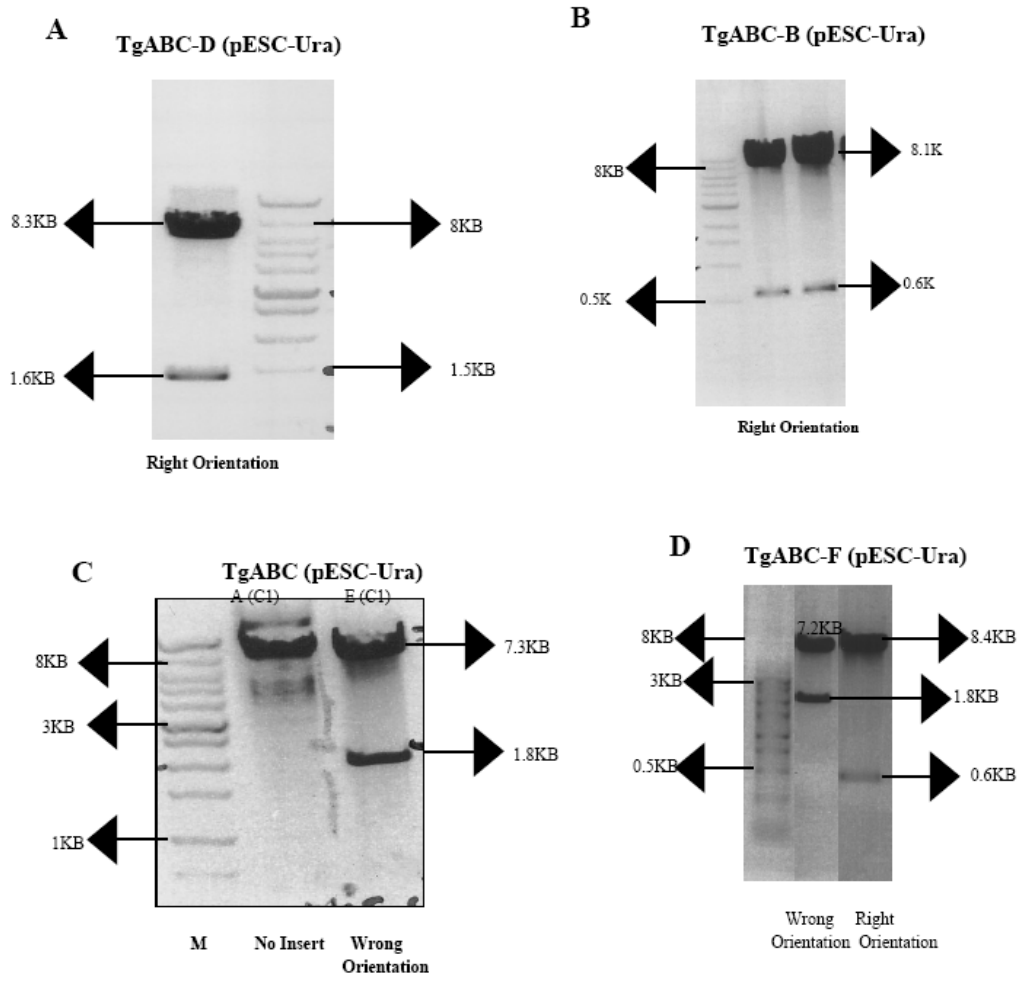


Figure 4: Fragments of different products were shown the orientation of insert in vector (pESC-URA). (A) Clone 1 of TgABC-D digest with *Bgl II* enzyme. The size of digested product was 1.6Kb & 8.3Kb that mean gene is in right orientation. (B) Clone 1 & 2 of TgABC-B digest with *Bgl II* enzyme. The size of digested product was 0.6Kb & 8.1Kb shows that gene is in right orientation. (C) Clone 1 of TgABC-A digest with *Nco I & Bgl II* enzyme. The product shows one band due to unknown error. The right orientation of the product was 7.2Kb & 2.1Kb as shown in Table 5. Clone 1 of TgABC-E digest with *SacI* enzyme. The size of digested product 7.3Kb & 1.8Kb mean gene is in wrong orientation. The right orientation of the product was 8.15Kb & 0.85Kb as shown in Table 5. (D) Clone 17 & 15 of TgABC-F digest with *EcoR1* enzyme. The size of clone 17 product was 7.2Kb & 1.8Kb means gene is in wrong orientation as well as the size of clone 15 product was 8.4Kb & 0.6Kb shows that gene is in right orientation.

S. cerevisiae AD-18 licking endogenous ABC transporters (Yeast genetic resource center, 2008) cells was used for transformation of pESC-Ura with inserts as performed by Ito *et al.,* 1983. *S. cerevisiae* strain was designed to test the function of ABC proteins. TgABC-B was cloned in to pESC-Ura and the construct was successfully expressed in *S.Cerevisiae* AD 18 strain for drug sensitivity. Due the unavailability of time, the work

could not be continued further. Future research will be needed to clarify the function of TgABC transporters for drug and lipid transport processes.

5 Conclusion

In summary, we identified six genes related to the ABCG and ABCH family in *T. gondii* genome as well as several proteins potentially interacting with them. This number falls in the range reported for other Apicomplexas; 15–20 members in *P. Falciparum* and *P. yoelii*, and 33 in *C. parvum* (Zapata *et al.,* 2002).

The presence of ABC transcripts in tachyzoite and bradyzoite infectious forms suggests an important role of these proteins in the basic biology of this parasite. Further research on membrane transport mechanisms and intracellular biochemical processes mediated by these ABC proteins could result in the identification of novel therapeutic targets. On the other hand, the implication of some of these proteins in the resistance phenomenon in different protozan parasites justifies further investigations on ABC's potential contribution to antiparasitic resistance in *T. gondii* (Klokouzas *et al.,* 2003).

6 References

Aline S, Virginie S, Sandie E-B, Dominique A, Christine Te, Roselyne G, Isabelle V, (2009) Molecular characterization and expression analysis of a P-glycoprotein homologue in Toxoplasma gondii, *Molecular & Biochemical Parasitology* 163 54–60.

Albert. B., Johnson. A., Lewis. J., Robert. K. and Walter, P. (2002) Molecular Biology of the Cell, *New York, Garland Science*, (4Th Edition) (P309, 500-502,532-533).

Black, M.W and Boothroyd, J.D., (2000) Lytic cycle of Toxoplasma gondii, *Microbiology and Molecular biology reviews.*

Dubey JP. (1999) Advances in the life cycle of Toxoplasma gondii. *Int J Parasitol*, 28(7):1019-24.

DE Hill and Dubey JP, (1998) Prevalance of viable T.gondii in beef, chicken, pork from retail meat stores in the United States: risk assessment to consumers.

G M Urquhart, J. Armur, JL. Duncan and Am.Dunn, FW. (1996) *Vet Parasitology*. Jennings Blackwell publication.

Henrik Ericson, Elie Järnmark (2007) Optimisation of Polymerase Chain Recation, Recombinant DNA Technology Laboratory Compendium.

Ito, H., Y. Fukuda, K. Murata, and A. Kimura. (1983) Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. 153: 163-168.

Jung, Joon-Y and Min-K.O (2008) Enhanced Production of 1,2-Propanediol by tpi1 Deletion in Saccharomyces Cerevisiae, J. Microbiol. Biotechnol. , 18(11), 1797–1802

Kong P., Richardson, P. A. and Hong, C. X (2005) Direct colony PCR-SSCP for detection of multiple pythiaceous oomycetes in environmental samples. J Microbiol Methods. Apr; 61(1):25-32

Klokouzas A, Shahi S, Hladky SB, Barrand MA, Van Veen HW (2003) ABC transporters and drug resistance in parasitic protozoa. *Int J Antimicrob* ; 22:301–17.

Louis M., Weiss, Kami K (2007) Toxoplasma gondii the model apicomplexan: perspectives and methods. *Academic press Elseviers UK*.

Nishith Gupta (2009) *Manual of Cloning work*, Department of Molecular Parasitology Humboldt University Berlin Germany.

New England biolab international (2008), available at: http://www.neb.com/nebecomm/default.asp.

Oligo Calc: Oligonucleotide Properties Calculator (2008), available at http://www.basic.northwestern.edu.

Pedro N. A and Boric S, (2003) *Zoonoses & communicable diseases common to man & animals*, (Third edition) PAHO HQ Library cataloguing- in – publication.

Primer3: Pick primers from a DNA sequence (2008) available at: http://frodo.wi.mit.edu/.

Remington JS, McLeod R and Desmonts G. (1995) *Toxoplasmosis Infectious diseases of the fetus and newborn infant*, (4th Edition) In: JS Remington and JO Klein (ed.) W.B. Saunders Co.,Philadelphia :140-266.

Simth J. E. (1995) Aubiquitous intracellular parasite, the cellular biology of T. gondii. *Int J parasitol*. 25:1301-1309.

Subash Ch. P (1990) *Review of parasitic Zoonoses*. AITBS Publication Delhi.

Sazzad K and Magnus F (2008) *Cloning of export signals by in vivo gene fusions using bla (β-lactamase) as a reporter gene,* Department of Natural sciences, University of Skovde, Sweden.

Schmid A, Sauvage V, Escotte-B S, Aubert D, Terryn C, Garnotel R and Villena I (2009) Molecular characterization and expression analysis of a P-glycoprotein homologue in Toxoplasma gondii. Mol Biochem Parasitol. Jan; 163(1):54-60. Epub 2008 Oct 17

Sayada C, Andersen AA, Storey C, Milon A, Eb F, Hashimoto N, Hirai K and Elion J, Denamur E. (1995) Usefulness of omp1 restriction mapping for avian Chlamydia psittaci isolate differentiation. Res Microbiol. Feb;146(2):155-65.

Sequences Lalign - find multiple matching subsegments in two sequences (2008), available at: http://www.expasy.org.

Stratagene an agilent technologies division (2008), available at: http://www.stratagene.com/products/displayProduct.aspx?pid=249.

Torada A. (2001) Toxoplasmosis: are cats really the source. *Aust fam phys*. 28: 743-747.

Virginie S, Jean-M.M, Dominique A and Vincent V, (2006) Identification and expression analysis of ABC protein-encoding genes in Toxoplasma gondii, Toxoplasma gondii ATP-binding cassette superfamily. *Molecular & Biochemical Parasitology* 147 177–192

Toxoplasma gondii genome data base (2008), available at: http://www.toxodb.org.

Transport classification data base (2008), available at: http://www.tcdb.org/progs/blast.php

Yeast genetic resource center, Department of biotechnology, Osaka University Japan (2008), available at: http://www.yeast.lab.nig.ac.jp/nig/english/

Zapata F, Perkins ME, Riojas YA, Wu TW, Le Blancq SM. (2002) The *Cryptosporidium parvum* ABC protein family. *Mol Biochem Parasitol*; 120:157–61.

6 Appendixes

Appendix I

Sequence of TgABC-A (2649 nucleotides)

ATGGACGTCGGGCCTCGAGTCGGCGAAGGCGGTTTCGGCGACGGCGACGATCCGAACGAGGACGGCGCGTGGGCGTCT
CTCTACCGTCGGCCCTTTGGGGCGTGCACGGTTGTGGTGAGAGACATTTGCTTCACTCTGCCTCGGAAACGGGAGTCGA
CGAGATGCAGCTTTTTGAGACGGTGTGCCTCGGCACTCTGCTCGCTCATCTGCTTCTCCAAGGACCTCCTAACAGTCCCC
CTGCAGCAGCTGGTGGTTGGCGAAAGCCAGCAGGAGAGACGCAGCTCACAGTACGCGTGGGGTCGACCGGCCAGTGC
CTGTAGACCTGCGAGCGAGATGGATGCGAGCAGCAGCGAGGCGACGCGCATGCGCAGCAGTAGGCTGGCGGCGGAAG
AGAGTGAAAAAGAAAGTGGAAGAAGTGGACGGGAAGCTGCGGAAGGCGACGCAGGCCGCGGGCATCCCGCAGAGGA
ACCGGAAGCCGACACACGCATGCAGACCTCGGAAAGCGGGCGCGACGAACGAGGGAGTTCCAACTCTGGGGGCGGCA
GAGGCGGCCATGGACCAGCTGCGCCGGAACAGTGGCGGGAGGATGAAGCAGCGAAGGACAAACAAGAGGCAGACGA
ACACAGACCGGAAGAACGACAGCTGGTGCTTCATCCCTTCTCCGCCGTGTTTGAAGCAGGCACCATGACGGCGGTGAT
GGGGCCGAGCGGGTGCGGCAAGACGACGCTGCTCAACGTGGTGGCGCACCGGACGCAAGCGCGGCAGACTGGCGGCC
AGGTGTTCGTCAACGGGAAGCCCCGAGGTCGCTCGTTCAAGCGCCTGGTGGCTTTTGTCCAGCAAGACGACATCTTTGA
CGGAAAAGAGACAGTGCGAGAATGCCTCGAGTTTTCACGAGACTTGCGCATGAACTTCTCGTCGATTCCAGACTCCACA
GAGAGACAGCGAGCCGCGGACGCCTACACCGACCAGGCCCTGAGGGTTCTAGGCCTGCAGGCTGTCGCCGACAGCCCC
ATAGGAAATGAAGCAGTCAGAGGCGTCTCTGGCGGCCAAAAACGAAGGGTCACTCTCGGCCTTGGTCTGATGAGCGAC
GCGCAGATTTTCCTTTGCGATGAACCGACGACAGGCCTCTCGGCTGCGGACGCGCTGGCGGTGGTGAGGACGTTGCGG
CGCCTCTGCGACGTGTACACCGTGACAGTGGTTGCCGTCATTCACCAGCCCTCCATGCAGATTCTGTCGCTCTTCTACAA
CTTGATTTTGCTCTCCTGCGAAGGCCGCTGCGCGTACAACGGGAAGGTCGCTGACTGCCAAGCATGGTTCGAGAACTGC
GGCTTCCCCTTCCCCCTCCACCAAAACCCCGCGGACTACCTCAGCGACCTTGTCTCCCCACACAAGGGACATCCTGCGC
AACTCGCCGCCTTCTACGAGGCACACCAATTCCCTGCAGTCCAAGCGCGCGTCGCCGAGGCGCTGAGGGGTGCCCAGA

CACCCAGCGGAGGTCGCCGGGACAGCGCGCGGGCCGACGCAGACGAAAACATGCCTCCAGATGGTGCACAGGCTTCC
AGTTTCGCGGCTCTCGATTTGCATTCGGAAGCGACGGGAGGTTCGCCTTCTTCTTCAGAGGCCGGGGGCGAAGACGCTC
CACACATCGCGAGTTTGTGGAAGCAGCTTTCGATGATAGGACGACGAAGCGCGCGTCTGTGGCTGCGCGACAGAGGGA
CGTTGGCGGCGATCTACTGCGACGCCGCCGTCGAGGGCGTCATTCTCGGCTTGGTCATGGCTCGCGTGCGTCAAACGCA
GCCGCCGTACTACCAGCTGTCCGCGTTGTTTCTTCTGGTCTACTGTGTTTGCGCTTCTGCGCTCTGGACGATCCCGCTCTT
CGTTCAACAGAAGGCGCAGCTGATTATGGAAGTCACGGGGGGGGTACTACTCGGCGTTGCCGCACTACTTGGCGACCAC
CAGCGTGTCTGCATGCGTCGTGGGGGGCTCCGACGTCGTTCTCTTCTCCATTCTCTGGTTCCTCGCCGGATTCGAGTGGA
CGGCGCTGCCCTTCTCGCTCTTCGTCTCACTTCTCGCCTTCCTCGTCGTCGACGGCGCCTTCTACCTCGCGTCCATCGCCA
GCAGCTCGTTCGCTCACGCCAACAGCGTCACCGCGGTGGCCTTCATGCTCTTCACCTTCGTCAACGGCTTCACCACGAA
CCCTCAGAGCATGCCGCTCTACGTCGGCTGGGTGAGCTACCTGTGCCCGTTCTTCCTCGCCTTCGAAGCCACAGCCGTA
CACGTCATGAAGGCGTACCCTTTCGCGGACCAACAGGCCTCAGGCCGCGGTCGCACTTTGCCCGCGGGAGAGCCGACC
CTGGCGTCGGCGGAAGAGCTGTTCAAGCAGTACGGACTAGCTGGAAGGGTGTATGGCGTAACGATGGATCCGGGCACC
TACGTGTGGGTTGTCGACGTCCTCATTCTCGTCCTTCTCGCCGTTGCTGTGAAAGGTTCCGCGGCCGTTTTCCAGAGCGT
CTGGGTCGCGCCAAACACCGAGAGCACCTGGAGGCGGAGTCGACTCCGAGGTGTGAACAAGCAGGCGAAAACGAACG
AGGAGGAAGCCGGGGAGATCGAGCCGCGAAAGTCGAAGCTGAGAGCTCGCGGGCGTGGGTAG

Amino acid sequence of TgABC-A (882 amino acids)

MDVGPRVGEGGFGDGDDPNEDGAWASLYRRPFGACTVVVRDICFTLPRKRESTRCSFLRRCASALCSLICFSKDLLTVPLQQ
LVVGESQQERRSSQYAWGRPASACRPASEMDASSSEATRMRSSRLAAEESEKESGRSGREAAEGDAGRGHPAEEPEADTRM
QTSESGRDERGSSNSGGGRGGHGPAAPEQWREDEAAKDKQEADEHRPEERQLVLHPFSAVFEAGTMTAVMGPSGCGKTTL
LNVVAHRTQARQTGGQVFVNGKPRGRSFKRLVAFVQQDDIFDGKETVRECLEFSRDLRMNFSSIPDSTERQRAADAYTDQA
LRVLGLQAVADSPIGNEAVRGVSGGQKRRVTLGLGLMSDAQIFLCDEPTTGLSAADALAVVRTLRRLCDVYTVTVVAVIHQ
PSMQILSLFYNLILLSCEGRCAYNGKVADCQAWFENCGFPFPLHQNPADYLSDLVSPHKGHPAQLAAFYEAHQFPAVQARV
AEALRGAQTPSGGRRDSARADADENMPPDGAQASSFAALDLHSEATGGSPSSSEAGGEDAPHIASLWKQLSMIGRRSARLW
LRDRGTLAAIYCDAAVEGVILGLVMARVRQTQPPYYQLSALFLLVYCVCASALWTIPLFVQQKAQLIMEVTGGYYSALPHY
LATTSVSACVVGGSDVVLFSILWFLAGFEWTALPFSLFVSLLAFLVVDGAFYLASIASSSFAHANSVTAVAFMLFTFVNGFTT
NPQSMPLYVGWVSYLCPFFLAFEATAVHVMKAYPFADQQASGRGRTLPAGEPTLASAEELFKQYGLAGRVYGVTMDPGTY
VWVVDVLILVLLAVAVKGSAAVFQSVWVAPNTESTWRRSRLRGVNKQAKTNEEEAGEIEPRKSKLRARGRG

Sequence of TgTABC-B (2106 nucleotides)

ATGCCCTCGCATCCTAATTTGGAGAGTACTACACCGCCTGTCCACGACCACAGTGGTGGGCAGCCAACGGAGTCCTTAA
CTTCAATTTGTCAGAAGAAGATTGCAGACTATGCGGTACCGGTGGAGAATTACTGGTCAACTGACCGTCTTTCGCCAGT
GGCTCTTGACAGAGAGACCACGACTGTTTCAGAGCCACGCCCCATCCCCTCTTGCTATGCACTGTCTACCATAGGTAGC
CGGAGTCCCGCGGCTGAGGTCAGAAACTCCTCAGCTTCCAGATGTGGAGGTGCAAGTCCACCTTCGCCTGAAAACGGT
GCTGCACGAAGTGGGGTTGTTCAGGTGCTCACACCGCCTCGCGTTGTCACTCTTCATAAGGAAGCCGATGATTCTGTAC
AGATGAAAGAGACCGAACAAGCGAAGCGTGCTGCAAGGCTACAAGTGTTCATACGCTTCACAAACCTCACATTCACTC
CTCCGCGAGGCCTTGCACGCCAATCCGTGAAAGAATGGGCTAAAACGGCTTCCTGTACACAGTCGTCACGACAGGCAA
TTGCTGGTTCCGATTCTGTTTCGGGACCGTGTTCAAAGGCGCCTCCTCCGTCCGCATATGCTCGCCCCCGACAAATTCTC
TTTGGGCTGAGCGGCTACTTTGCCCCTGGCGAAATTGTGGGAATCCTCGGTCCTTCTGGAGCAGGCAAAAGCACTTTCC
TGTCTGTACTGTGTGGAAGGCTCAAAAAAGGAGTCGGCGGACTCATTGATATCAACGGGGAACCTGCCCCAGCGCGAA
TGAAGAAGATTGTTGGTTACGTGATGCAGCAGGAGTATTTCTTCGGAAATTTGACAGTCGAAGAGACACTGATGTACAC
CGCCAGACTGCGTTTGGGGAAGAAAACCAGTTTCGCAGAGAAAAAAGCACGGGTCGAAGAAGTTATAAACTCAATGA
AGCTAGACAAGTGCAGAGGGACTCGCGTGGGGAGCGCTTTTTGCCGTGGCCTCTCCGGAGGAGAGCTGAAGCGTCTAA
ACATCGCGAACGAGCTGTTACTTCATCCGTCACTCTTCCTCTTGGATGAGCCCACAAGCGGACTGGACGCCTCGATCTC
GACCGCCCTTCTCGATATGTTGAAACAGCTAACACGAGCCAACAGGTGTACTACCGTCTGCACCATACACCAGCCTAAC
AGCAATGTGTTCATGAGGTTTGATCGGGTGATCTTTCTCAAAGACGGCCACATCGTGTACCAGGGGAAGCCCTCCGACG
TATGCGCATACTTTTCTTCTCTCAATTTCCACTGTCCAGAAGGGTGGAACATTGCAGATTACGCACTGCATGTCATTAGT
GCATCCGATGATGAGAACTTCAGGCGGCGCTCACTCACCAGCACGACTGCAGCCGGGGGAGATCCGATCACTGGCAGT
ACACATGGTCTCCTTTACTTCACTACTTCTTACTACAGCTTCGGACCTGCCTACATGGCATTCACATCTTTTCCAGCTGA
GCGCGCCGTTATCTCCAGAGAAAGAAGCAGTAAAGCATATAAAGTGTCATGTTATCTTCTCGCCAAAACCATGATAGA
CCTCGTCGTCCAGATCTTTACTCCGTCACTGTGGCTGGCAATCGTATATCCGCTTGTCGGATTGCCCTCAGACCTTGGCG
TGTTTATGGCTTTTTGGGCTCAGCTGGTACTGCTTGTCTGCATAGCTCAAGCCATAGGTCAACTGATTGCTGCGATCGTT
GTGGACGATGCGCGCCTAGGGGGTTTGCTGTTGTCAGTCATTCTGATTTCATCTTCCATCAGTAGTGGCTTCTATGTTCA
GCAACAGAGATTAGGCCCATGGATATCTTGGTTTCGATGGTTGTCTTTCCAGAACTACGCGGTGACAAACTTCGTAATA
GTTACAGTGGGATCGTCTTCGACACTCTCGTGCTCGGAGTTCTCTTCGTTCCCAACTTGCCCAGGGCAGCCGATCACAG
CAGAGATGATTACCCACAGGTTCACAACAGCACTTTCGCCTTTATCCAATATCACCTTAATGATTTGTACATGGCTAGCC
ATCAAACTAATGTGCTATGGAGTCCTCAAGAAATCTCTTAAACTGAAGTCCTGA

Amino acid sequence of TgABC-B (701 amino acids)

MPSHPNLESTTPPVHDHSGGQPTESLTSICQKKIADYAVPVENYWSTDRLSPVALDRETTTVSEPRPIPSCYALSTIGSRSPAAE
VRNSSASRCGGASPPSPENGAARSGVVQVLTPPRVVTLHKEADDSVQMKETEQAKRAARLQVFIRFTNLTFTPPRGLARQSV
KEWAKTASCTQSSRQAIAGSDSVSGPCSKAPPPSAYARPRQILFGLSGYFAPGEIVGILGPSGAGKSTFLSVLCGRLKKGVGGL
IDINGEPAPARMKKIVGYVMQQEYFFGNLTVEETLMYTARLRLGKKTSFAEKKARVEEVINSMKLDKCRGTRVGSAFCRGL
SGGELKRLNIANELLLHPSLFLLDEPTSGLDASISTALLDMLKQLTRANRCTTVCTIHQPNSNVFMRFDRVIFLKDGHIVYQGK
PSDVCAYFSSLNFHCPEGWNIADYALHVISASDDENFRRRSLTSTTAAGGDPITGSTHGLLYFTTSYYSFGPAYMAFTSFPAER
AVISRERSSKAYKVSCYLLAKTMIDLVVQIFTPSLWLAIVYPLVGLPSDLGVFMAFWAQLVLLVCIAQAIGQLIAAIVVDDAR
LGGLLLSVILISSSISSGFYVQQQRLGPWISWFRWLSFQNYAVTNFVIVTVGSSSTLSCSEFSSFPTCPGQPITAEMITHRFTTAL
SPLSNITLMICTWLAIKLMCYGVLKKSLKLKS

Sequence of TgABC-C (2946 nucleotides)

ATGGACTTTCTGCGCAGACGCGATGTCTTTCGCAGCAGCAAAGACGCTCGAGAGAAGCGAACTGGTTCTTCCTCGCCGA
GTCGCGAGGGCGGCTCTTCCTCGCTGTTCTCTCCGGCGGGGGGTTTGCATGTGACGCCTGCTCGTCCCATCACGCTTGTT

GCGCGCGGCTTTTCCTACGCCGTTCGGATGAAGCCGCAGATCCCCTGCACCCAGTGTGCCTCCCCGTGTCTCCCTCGAA
AGAAGCAGAGTAGCAAAGCAGAGGATGCGAACAAGCCGAGCGACGAAAACGAGCCGTTCGCGGCGCCGAGCGCTCTC
AGAAGACCCGAGAGCGCAAATTCGTGGCGAAGCAAAGGAGGCCGCGCAGACTCCCCGGCACACGGTGCCTTGGATAT
CGAGAAAGGTTCAGAAGACCAGCACATCGGAGACTGTGCCACTGAAAACCGAACGCCCGAAGAGAGGAGCGCCGACG
GCGAGACGGAAACGGAGGGGCCTTTGAAAATTATTCTCAAAGACATCAATCTCTGCGCGAGACCTGGAGAAATGCTCG
TCATCATGGGTCCCAGCGGCTCAGGAAAAACGACGCTCCTCAACGCATTCGCCGGACGGTCCCCTCCCTCGTCTCGCGT
CGTCCAAGACGGATCGCTTCTCTACCTCGGTCTCCAGCCTGGAGAGCCTCTGTCCTCAATTGCAACTTACGTCATGCAG
AAAGACATGGTCCCCGAACTCTTAACAGTTCAGGAGTACGTCACTTTTTTCTCTCGTCTCAAGATGCGGGACGCGACGG
AAGAAGAGAGAGCAGAGAGGGTGGAGGTCGTCCTGCGCGAGCTCGGCTTGTGGGGTAGCCGTTTCACCCGAGTCGGCG
GAAGCGCCAAGAAAGGTCTCAGTGGCGGAGAGATCAAACGCCTCGCTCTTGCCGTCGAGCTACTTCACAATCCGTCTCT
CATTTTCCTTGATGAACCAACGAGCGGCCTTGATGCAGCGCTCGCGTTCGAGACCATGAAGCTGCTTCTTCGCCTCGCG
CGCCATGGAGGCCGAACTATTCTCTGCACGATTCACCAGCCCCGTAGTCAGCTCTTCGCGATGTTTGATCGCTTGGTCCT
GATGTTCGAGGGGCGCATAGTGTTCCAGGGCCCCGCCAGGCACTGCGTATCCTACTTTGCCAAGCGCGGGTTCCACTGC
CCGCCTCAGTTCAATCCTGCTGACTTCATTCTGGATCTCCTGAATGTGACAACTCTTACTGCGAACACCCCTCTCGCCTG
TCCCCAGTCGGAGGGGTCTCTGAGCGCTCAATCCCTCCTCGAGGCGGAGTTGAAGAAGGTCCAGACGCAAGAGCTGCA
GCGGCTCGCACTCGCGACGCGGGAGGGCTCGCGACACGGAGGCCTGGGGGAGCCCCAGGAAGCTCCGAGGCAGGTCG
GCGAGCTTTTGGAGCCCAGTGCCTCGTCGGACGGCTTCACTCCGGGCGCGGAAAGGCCCCAGGCGACCGCAGCCGCAG
GGACAGATGCCGAAGAGACCCGCGAACGACCCGAGAGAGACGGCGGGGGACCCATGGGGGCGGCAGCGGGCAGGGC
GGGAGATCAGGAGGAAGAGACAGGTCCGCGTCACGAGGCTAGTCCGGCGCCCCTGGCGAGACATCGCTCCCACGGCC
AGGAGTCGGGCGCGAGCGGGCGACGGCGGCGCGGAAACGACGAAGCGGACACTCCTGGGGGAAGCGCTCAGGATGG
ATCTGTGGAGGGCCGAAAGGTCGTCTCTGTGGTCCCGATCGTTTCTGCGCCGTCGTCGGAGGTCGTGGAGGCGAACGGC
GAGGATGCGGGAGGGGCGCGAGCGGCGGTGAGTCAGGGAGACGCGAGGGACAAGCCGAGGCAGTCGCGGCTGCAGG
GCGAGGCCCGAATCGCGAGTGGCTTTCTGGACGCCGAGGAGACTCAGGGCGAACTCCACAGGGTTCTTGTCGACGAGA
ACGATGTGAAGCGGCTAGCAGACTCGTATGCGGCGAGCCCAGAGAGAGCGCATGTCGAGGAACTGATCGTTCAGTGTC
TCCAAGCGGCCCCGCCAGACCATGTGAAGCCGTCGGGACGAGCTGCCGCCGTCAAGCGACTTCTTCCGCAGCGCCGGT
GGTCCGACTGGGGCCGGGAAATTTGTGTCCTCATTCAAATGGGTTTTCTAAATCATGTCAGGAATCCAATGAGTTCTGT
TGTTCAGCTCGCGCTGAACATTCTCTTCGGTTTAATCTTTGGCGCGATTTTCTTCAACATTCCTGGGCAGGGGCAGACCA
TTGATTCGGCGAGGAACATGTTGGGGTGTCTCTTCTTCCTCGGGTCTCAACTTATCTTCGGCCCACTTGACTGCCTTGTT
CTGTTCTGCACGAATCGTGAACTCTTCAATAGAGACACCGCAAATGGCAACTACACTCCATTCTCGTACTTCGTCGCCA
AGTGTCTCTCGAATCTGCCTTTCGAGCACCTTCCGCTGACCTGCGTTACGCTGGTCGCCTACGGGATGTGTGGCTTGCAC
AGAGGAGCCGCGCACTTCTTCATCTATTTCGTCATCGGTCAGTTGTCGATCTTTGCCTCGACCTCTCTCCTCGGTTTGATT
TCTGCCGCATCGCCGCGGGTTGCAGTTGCCCAGGCCGTCGCTCCCATCATTCTTCTCATTTTTCTCCTGGTTACGGGCTA
CTACATTCGCGCTGATGACATTCCAGCGGTCATCCGGTGGCTGAAATACCTGTCGCCTATTCACTATTCGTATGTGGCTC
TAGCTCTCAATCAATTCCCGCCTGACGAGAAATGGGGTGATTCGTCCAACAAGGACTTTCTCGAGTCCTACGGAGGCAT
CACTAAGACAGATCTTGGGTTCTACATTGGCATGCTTGCTCTGCTCGGCTGCGTTTGCCGCGTCCTGTCTTTCTTCTGTCT
GAAGTACATGCACAGGCGCATCGGCCTTGAAGCGTGA

Amino acid sequence of TgABC-C (981 amino acids)

MDFLRRRDVFRSSKDAREKRTGSSSPSREGGSSSLFSPAGGLHVTPARPITLVARGFSYAVRMKPQIPCTQCASPCLPRKKQSS
KAEDANKPSDENEPFAAPSALRRPESANSWRSKGGRADSPAHGALDIEKGSEDQHIGDCATENRTPEERSADGETETEGPLKI
ILKDINLCARPGEMLVIMGPSGSGKTTLLNAFAGRSPPSSRVVQDGSLLYLGLQPGEPLSSIATYVMQKDMVPELLTVQEYVT
FFSRLKMRDATEEERAERVEVVLRELGLWGSRFTRVGGSAKKGLSGGEIKRLALAVELLHNPSLIFLDEPTSGLDAALAFET
MKLLLRLARHGGRTILCTIHQPRSQLFAMFDRLVLMFEGRIVFQGPARHCVSYFAKRGFHCPPQFNPADFILDLLNVTTLTAN
TPLACPQSEGSLSAQSLLEAELKKVQTQELQRLALATREGSRHGGLGEPQEAPRQVGELLEPSASSDGFTPGAERPQATAAA
GTDAEETRERPERDGGGPMGAAAGRAGDQEEETGPRHEASPAPLARHRSHGQESGASGRRRRGNDEADTPGGSAQDGSVE
GRKVVSVVPIVSAPSSEVVEANGEDAGGARAAVSQGDARDKPRQSRLQGEARIASGFLDAEETQGELHRVLVDENDVKRLA
DSYAASPERAHVEELIVQCLQAAPPDHVKPSGRAAAVKRLLPQRRWSDWGREICVLIQMGFLNHVRNPMSSVVQLALNILF
GLIFGAIFFNIPGQGQTIDSARNMLGCLFFLGSQLIFGPLDCLVLFCTNRELFNRDTANGNYTPFSYFVAKCLSNLPFEHLPLTC
VTLVAYGMCGLHRGAAHFFIYFVIGQLSIFASTSLLGLISAASPRVAVAQAVAPIILLIFLLVTGYYIRADDIPAVIRWLKYLSPI
HYSYVALALNQFPPDEKWGDSSNKDFLESYGGITKTDLGFYIGMLALLGCVCRVLSFFCLKYMHRRIGLEA

Sequence of TgABC- D (2301 nucleotides)

ATGGCAGACTCTGCCCTCGCACCTCAAGAGTCTGCCGGCACTGTTGCCGCTGACAATGCCGCTTCTTCCAGGGAGGGAG
GTCGAGGTACGTCCGGCGCGCATGGACAGGGCGCGCGCGCGACTGCAGAGCGGCGACTTTCTCTGGACGGCGCGACCG
CGGAGCCGAGCAGTGGAGGCGCTCTCCGGCCGCAGTCGGACGCGGACATGCTCGAAAGAGGCCTTCTCGGAGCTTCCG
CCTCTCCGGAGGCTGCCGCATGCAAGGACGAGGCAGTCGATAACTCGCCAAGGCCGCGCAAGGAAAAGGAAGGTGCC
CCCTGCTCCGAGCCTCTCCGTCGCTGTCGAGAGATTGGAGTCTCGTGGACGGATATCTCCTTTGAAGCGCGGGTGCGAA
GGCCGCGAACGGGCATTCTCGGTACGCCTCCCCTGTGCGACGTTCTCAACGCCCTCGACAGCCTCACGCAGAAAGGCCC
TCTGAAAAAGATTTTGTGCGGCGTCAGCGGAGAAGTCCTTCCCGGCGATATGGTGGCTCTGATGGGCGCAAGTGGAGC
AGGAAAGTCGACGCTGTTGAACATTTTGAGTCGATATTTGCGAGAGACCTCTGGAACGGTAAAGTACGGAGACGCGCA
GTTGGAGTTGAAGGAGGCGAAGAAGGTTTCGTGCTTCATTCAGCAAGAAGATCTCTTCAACGGATTCATCACTGTTCGA
GAGCATCTCCAGTGCATTGTTCGACTCCGAACTACGCTGCCGCCGAAAGAGCGCGAGGCTCTCGTCGACCGCCTCCTCG
TCGCCTTTGAACTTTCGAAAGCGGCAGACACATGCATAGGCAATCTTCAGATGGGAGCGCGAAGAGGAATCAGTGGAG

GCGAGAAGAAGCGCCTGAGTGTCGCGACGGAAATCGTGACAAATCCTTCCATCATTTTCGCGGACGAACCGACGACAG
GTCTGGACTCGTTCATGGCCGAGGCTGTGATGACTGTTCTCGAGCGGCTCGCCCAAAATGGGCGTTCGATCATTTGCAC
AATCCATCAGCCGAGCACGACGGTGTTTGAGAAATTCAACAAAGTGATTCTGCTCGCAGAAGGCCGCATGGTCTTCGCT
GGAGACCGACTCGCCCTGCGAGTGTATTTCGCGCGCGTCGGAAAGTCGATTCCGCCCTACACCAGTGTCGCTGACTTCG
TCATCGACGTCCTCTCCTCCAGCGAAGGCGCAGAAGCGACTCTCCAAATCGCGGAGAAGATGCATGCGGCGTGGATCA
ACACAGGCGTTCCCTTCATGCGTGACTGGCACGAATCCCTGCGAGAGCAATACATCCAGAGGCTGCAGTCTGAGGGAG
TTGCGGACAAACTCAGCTCGTTCAAGGGCGAGACGAAGCGTCTCGCTTTGCGCAAAGAAGCGCTGGAGGCTCCGACTG
CGACTGATGAGGAGCCAGGTCTCAAGAAGAAGGAAGCCGCAGTGGAGGGTCGCGTGAGTTGGTGGACGCAATTCCAG
GTGCTGCTCCACCGCCGCTCGCTGGCGAACAAACGAAATCCACAGATTCTGCAAGCTCGTGTGGGCCAGACTCTCGTCT
CTGCTCTTCTCCTCGGCTTTATTTTTCTTCGCCTTCGTAAAGGGGATGCCATTTCCAAGAACGGCGCTGCCAACTTCATC
AACCTGAATCAGGGCATGACAGGTCTGGTCACCGTTCTGCAGACGTTTACGACGGACAAAATCGTCGCGCTGCGCGAA
TATCGCTCGGGGACGTACAGCCTCGTTCCGTATTTCCTGGCAAAGACGGCTGCCGACGCGGCGTTCCAAATTTTTAACC
CAGTCGTCTTCTTCAGCATCGCATGGTACATGATGAACTTGAATCCGTCGGCGACCCGCTGGCTCTGGGGCCTTGGCTTC
ATCTTCCTGCAAACGAACGCGTCGATTTCGATGGGCTACTTGATCTCGTGCATGTGCCCCGACCTTGAAATTGCCCTCTC
GGTCATGCCGCTCCTGACGATGCCACTAATCCTTGTTGCGGGATTCATGATCATTCTCGACTCTCTGCCCAGATTCTGGA
TATGGGTGCCGTATCTCTCACCCTTCCGGTGGGCCTTCTCCGGCATCATGCACGCCGTCTGGGAAGACGTCGAACTCGA
TCCTTGCCCTGCGGGCATGAATCCGCCGTCCTGCTACAGCTCAGGCGCTGAGGTTCTCGAGTACTACTGCCTTGATGGC
GACAGCATGTGGCTGAACGCTCTCTACCTCGTGATCATGGTCGTCGGCTACCGCGTTGTCGGCCTCTTGGTTCTCCTCAT
CTTGAATCGCCGAAACTAA

Amino acid sequence of TgABC-D (766 amino acids)

MADSALAPQESAGTVAADNAASSREGGRGTSGAHGQGARATAERRLSLDGATAEPSSGGALRPQSDADMLERGLLGASAS
PEAAACKDEAVDNSPRPRKEKEGAPCSEPLRRCREIGVSWTDISFEARVRRPRTGILGTPPLCDVLNALDSLTQKGPLKKILCG
VSGEVLPGDMVALMGASGAGKSTLLNILSRYLRETSGTVKYGDAQLELKEAKKVSCFIQQEDLFNGFITVREHLQCIVRLRT
TLPPKEREALVDRLLVAFELSKAADTCIGNLQMGARRGISGGEKKRLSVATEIVTNPSIIFADEPTTGLDSFMAEAVMTVLER
LAQNGRSIICTIHQPSTTVFEKFNKVILLAEGRMVFAGDRLALRVYFARVGKSIPPYTSVADFVIDVLSSSEGAEATLQIAEKM
HAAWINTGVPFMRDWHESLREQYIQRLQSEGVADKLSSFKGETKRLALRKEALEAPTATDEEPGLKKKEAAVEGRVSWWT
QFQVLLHRRSLANKRNPQILQARVGQTLVSALLLGFIFLRLRKGDAISKNGAANFINLNQGMTGLVTVLQTFTTDKIVALREY
RSGTYSLVPYFLAKTAADAAFQIFNPVVFFSIAWYMMNLNPSATRWLWGLGFIFLQTNASISMGYLISCMCPDLEIALSVMPL
LTMPLILVAGFMIILDSLPRFWIWVPYLSPFRWAFSGIMHAVWEDVELDPCPAGMNPPSCYSSGAEVLEYYCLDGDSMWLN
ALYLVIMVVGYRVVGLLVLLILNRRN

Sequence of TgABC-E (2439 nucleotides)

ATGAATTCCCTCGCTGGCAGCGTCGTCGGGCAAGGCCAGCCTGTTGAAGAGACAGGAGTGGAGGCCCCGGCCTCTGAC
CCTCTTGCAGATGGAGAGAGACACCCCAGCCGAGAGGAACCTTCTGGTGCAATTTTCTCTTCTCAAGGCGTCTCACGCG
CGGTTGTGGAGAGTCGCACGGAAGAACAAGTTCACCTGCTCAACGACGGAGTGTCTAGACACCCCGGGGGACTTCCGG
ATGCAGAGGGCTGGGATGAGAATCCAGTCCGAGTCGACACAGGCGACGCAGTGGCTCATTACGACGGACTCCCCCTCG
TCACCCTCAACTTCGAGGACCTCACAATGGACGTCGTGACTCCGGGCGACGGCTGCGTGAAGAGAGCGTTGAAGGGTC
TCTTCAAACGGCCGATCAAGGAGCCGACGCGAAAGAGAATTCTCTCCCTCGAAGGCCTGAAAAGTTCCTTCCGGCCTG
GAGACTGTGTCGCGCTGATGGGAAGCAGTGGAGCAGGCAAAACGACTCTTCTGAATGTTCTGTCGGGGCGCGTGACGA
AGAATGTCGGAGGCCGCGTTCAGTACAACGGTCTGGAATTGCCTCCTGAGGCACTGAAGGCAATTTCTTGCTTCGTTCA
ACAAGAAGTGATTTTCTTCGGAACCTTGACCGTGCAGGAGCATCTCGAGTACCAGGCTGCTCTACGGCTGCCGCCTTCG
CTCTCTGCTCGTGACCGCGCAGCGACAGTGAACGCGATGATTGAAAAAGTCGGTCTCTCCAAAGTCGCGGATTCCCTTA
TTGGCAATGTCTCTCAGCAGCAGCTCGTCGGCATTTCCGGAGGCGAACAGCGACGCCTGTCTGTGGCGACGGAGCTTCT
GACAGAGCCTTGTGTCATCTTCGCCGACGAACCGACCAGCGGCCTCGACTCTTACATGGCGATGCAAGTCGTCAAGCTT
TTCAAGGGCCTCGCTCTCGACGGTCGAACTGTCGTCTGCACCATTCACCAGCCCAGCTCGAGTGTCTTTGCGCAGTTCA
ACAAGGTCGTTCTGATGTCGGAAGGGCACCTTCTGTACTGCGGAGACCGCGAGGCTTGCATCGGTTGGTTCGCGCATTT
AGGGCAAGTTTGCGAGGCGGACATGAATCCAGCTGAGTTTTTGATTAAAGTCACTGCAGTGACGGACGACAATCGCGA
GGCGGCTGTCCAGAGAACGGTGGAGTGGGCTGAACGTTGGCGGCAGGAAGGCGCTATGTTTCTCGAGCAGTGGGAGGC
TCTTGGCGGACGCGCTGCCGCTTCCTCCGATCAGATGCGCATTCAGCGCCTTTTTTCATCGATGGAAGAACCGGCTGAG
GCCTGCGCAAAGGAGGCTCCAGGTGAAAACGGAGCAGGTCCGGGGGGGTTCAGGAGCGCTAGTCGAGACCAAGCAGCC
TGGAGAAGTCGTATCACATGCGGTCCTCGTCGACCAGAGCGCCCAGACGTCGGAGGCCGATCGGCCGGAAACGCAGAA
CGCCGAGCCGACGCTCTTGCATCGAAAACTGAGTTCGCGCGCCCAGTACATGCTGAAGGGCGGTGCCATGTCGAAGGC
TTGTTTGGAGGAGATGAAAACAGACAGAATCGGAGTGCTCAGAGAAACCTGGCTGCAGATCCAACGCAGTACTCTGCT
CAGAGGACGCGACCCCTTCTCCACGTATGTCCGACTCGTCACAACCGTGATCAGCGCTCTCATCCCTGCCCTCATGTAC
TACCGCTTGACGTGGCAGAGCTCAGACGCGTGGAACAAAGTATCTTCTTCTTTTTACATCATCCTCAGCGAGTCGATGG
CTTGTCTCTTCGGCGCTTCCATGGCCTTCAACAAGGAACGCGCTGTCATCCAGCGGGAATACGAGTCCGGAGTGACACG
AATGCCTCTCTACTTCATCGGCCGCATCACCGCAGACTCCCTCCTCTGGATGTTTTTTCCGTTCATCTATCATTTAATCGT
CTACTGGATCTCCGATCTCGGGGGAGACTCTGTGAGCAAGTACTTTGCTTCCCTCGCGATCACGCTGCTCCTGATTCAGG
TTGTCCTCTCCTACACCTACGTCGTCGTCGCGCTCATCAAACATCCAGTTGCGTCGACAGTTGTTCTTCAGATCATGCAG
ATGATCCTCACTCTCTTCTCTGGATTCATGGTCAAGTTGGACGAACTGGGAAAGTTCTGGATCTGGATTGTGTACCTCTC
GCCATTCAAGTACGCGCTCCCTTGCTTCACCGTCACTATTTTCTGGAACACAGAAATCTCGTCGCCCTCAGGATCCACCG

TCTCCGGCGTTGACTTCCTGAACGATACCTTCGGATTTCAACATGACAAGTTCTGGCTGTACGTCGGCCTGCTATTCGTC
CTCGGAATCAGTGGGAGGCTGTTGGGAATGGTTGCGCTGTCGTGGAAGGCGAGTCGAACGAAGGAGAACCAATGA

Amino acid sequence of TgABC-E (812 amino acids)

MNSLAGSVVGQGQPVEETGVEAPASDPLADGERHPSREEPSGAIFSSQGVSRAVVESRTEEQVHLLNDGVSRHPGGLPDAEG
WDENPVRVDTGDAVAHYDGLPLVTLNFEDLTMDVVTPGDGCVKRALKGLFKRPIKEPTRKRILSLEGLKSSFRPGDCVALM
GSSGAGKTTLLNVLSGRVTKNVGGRVQYNGLELPPEALKAISCFVQQEVIFFGTLTVQEHLEYQAALRLPPSLSARDRAATV
NAMIEKVGLSKVADSLIGNVSQQQLVGISGGEQRRLSVATELLTEPCVIFADEPTSGLDSYMAMQVVKLFKGLALDGRTVVC
TIHQPSSSVFAQFNKVVLMSEGHLLYCGDREACIGWFAHLGQVCEADMNPAEFLIKVTAVTDDNREAAVQRTVEWAERWR
QEGAMFLEQWEALGGRAAASSDQMRIQRLFSSMEEPAEACAKEAPGENGAGPGGSGALVETKQPGEVVSHAVLVDQSAQT
SEADRPETQNAEPTLLHRKLSSRAQYMLKGGAMSKACLEEMKTDRIGVLRETWLQIQRSTLLRGRDPFSTYVRLVTTVISALI
PALMYYRLTWQSSDAWNKVSSSFYIILSESMACLFGASMAFNKERAVIQREYESGVTRMPLYFIGRITADSLLWMFFPFIYHL
IVYWISDLGGDSVSKYFASLAITLLLIQVVLSYTYVVVALIKHPVASTVVLQIMQMILTLFSGFMVKLDELGKFWIWIVYLSPF
KYALPCFTVTIFWNTEISSPSGSTVSGVDFLNDTFGFQHDKFWLYVGLLFVLGISGRLLGMVALSWKASRTKENQ

Sequence of TgABC-F (2385 nucleotides)

ATGGACGCTCCGGACATGAGGCGGGGAGCTCGTTGGTCGTCGGACTCTGTCGCTTCCCCCTCGGGTGGTGGCTCTTCCG
AAGCCGGCAAGAAGGAGGACGGGGAGGCTGGGGATCCTTCCTCTGCCTCGCTTGCAGAACGAAGAGACGGCCGCGCC
TTCAGCGGCGCAGAGACTTGCGGGAGGCGAGGGGACGCGCTTCTCTCGCGCTTCCGCGGCTGGCTGTGGAGACAGCCA
GCGAGAGAAGAGATCTCGAAGCGGCCGGGAGGGCAGAGGGAAGAGCGCGAACTCGGCCGGCTCTTGCCTGCATGCGA
GACGCCCAGGTCGAGTGAGTCGAAGGCGTGCAGTCCCCACAAAAAGGGCAACCTGCTCAAGAGAGATCGCCAGACCG
ATGCAGAGCGGTGGGGACGAAGTCGAGGCGTGTCTTTCTCCTTCAGAGACGTCCAATGCAGCGTCCGCTTCGCCGGCCC
CGACGGAAAAAAACAGAAGAAAACGATTCTGTGGCCGATCTCGGGCGACTTTCCTCCAGGAAGCGTCGTCGGAATTCT
CGGACCCTCTGGAAGTGGAAAGACGACGCTTCTTGACGTCTTGGCGAGTAAAAAAGTCCCCAAGGCCTACACCGGCGA
GATCTTTGTCGACGGCAAGAAACGAGAGGATAAACGCTTCAAAGCTCTCTCCATCTACGCTCCTCAGGAAAACATTTTC
AATGGAAACGAAGAGGTTCGGGAAGTGCTGGCTTTCGCTGCGGCTCTGAAGTCGGGTCGTTCGAAGGAGGATCAGAAG
AGGCTGGTGGACGCAACTCTTTCTTTTCTCGGTCTTGAGCGCGTTGCTACTCAACGCGTTGGAAACACCGTCGTTCGGG
GCATCTCCGGAGGCCAGAAGCGTCGTCTCGTCATCGGCCGCGCATTGGTCACCAACGCGAGTCTCGCCTTTTGCGACGA
ACCGACTTCAgGCCTGAGCTCGACGGACGCAGAGGCGCTGATGCTTGGCATCAGGGAAGTGGCGAAGGCCTGTCGCGT
CACCTTCCTCATCGTTATCCACCAGCCGCGTGTGGAAGTCTTTGAGCTTTTCGACGAGGTTATTCTCCTCGCACAAGGCC
GGTGTCTGTACAACGGCCCACGGTCCCAGATGGAAGCATACTTCTCTGCCCTAGGCTTTCCTCTTCCACCCTACGTGAAT
CCGGCAGAGTTTTACATGGAGCTCACCGCGCACGACGCGACTGTGGAGCAACTCGCGACCGCGTACGCCTCGAGAGCG
GAGACAGGTGTCTCCTCGCGCCCAGAGGCGGCGGCTTCGGGCTCCCTGCTGGAGGACGCGCAGAGACAACAGCGAGA
AGAGCTGGCGGAGGAGGTCGACGGTGCCGACGCGGCAGACGAAGAGGAAAGAGGCGAACACCTTGTGGAAGAAAAC
GAAGAAGACGAAACGCCGAACAGTCGAAGAGAATCTACAGCAGCGTCTCTCGCAGGCATGCACGCGCTCGTGACGGC
ACTTCACGACGATCCTCGATCGATGCATCCACTCCCGGGATTCTTCGGACGGTTCATGATTTTGACGCGTCGCAATGTG
ACGCTGGCGTGGCGAGATCGAAATGTTGTCTTCATGGTTGCATGCAACGCCCTGCTCGTTGCAATGCTCGCGTCCGTCC
TCTTCTTCAACGTCTACCAAGAAAAAGCCATCATGTATCACCTGTCGGTCTTCTTCTTGGTTTCGCTTTCTCTGGCAAGC
ATCCCGGCGGTGAACATGGCGATGTATCTCGAAAAGAAGGTGTCTTATCTCTACGAGGCCTCCGACGGCTTTTACACCG
CGGGGCCGTACCTTCTGTCTGAAATGACCACCGGCTTCCTGATGCTGCTGGGCGCGACGCTTCTGGTCTTCGTCGTGGTC
TTCCCCTGCTGCGCCTTTCCGTTTTCCAAATTCGGGATGGTCTTTCTGCTCTTCATTCTTTTTCTTCTCGTGGTGGACGCC
GTCATGCAGCTCGCTGCTGCCATGTGCGCCACTTTCATTATGGCGAGCACCTTCGCGGGCGGCTGGCTGGCGTTGACTT
CCGTTGTGAATGGGTACAACGCAAATCCCAAGTCGATGCCTGCATGGCTCACGTGGCTGGTGTACCTCTCGCCCTTCTA
CTACCTCATGGATGGTGTGGCGATTGTCCTCTGCTGGGAGAACCAAGATCTTTTCGGAACTCCAGAAAAAGCAGCGAA
ATATGGATACACCAGTTGTGAAGAACTTCTGAAGGCTTACGGGTTTGCTGGGACCCTCGGGGGAAGCTCCCTGACTCCG
CCACAGTGGCTCTGGTCTGTGGACGTCCCAGTTTTGTGCTTAATGACTGTCGGCATGAAGACGTTCGCCTGTTTTTACCA
GGCGTACTGTGTCCGGCTTCGTCGCTGA

Amino acid sequence of TgABC-F (794 amino acids)

MDAPDMRRGARWSSDSVASPSGGGSSEAGKKEDGEAGDPSSASLAERRDGRAFSGAETCGRRGDALLSRFRGWLWRQPAR
EEISKRPGGQREERELGRLLPACETPRSSESKACSPHKKGNLLKRDRQTDAERWGRSRGVSFSFRDVQCSVRFAGPDGKKQK
KTILWPISGDFPPGSVVGILGPSGSGKTTLLDVLASKKVPKAYTGEIFVDGKKREDKRFKALSIYAPQENIFNGNEEVREVLAF
AAALKSGRSKEDQKRLVDATLSFLGLERVATQRVGNTVVRGISGGQKRRLVIGRALVTNASLAFCDEPTSGLSSTDAEALML
GIREVAKACRVTFLIVIHQPRVEVFELFDEVILLAQGRCLYNGPRSQMEAYFSALGFPLPPYVNPAEFYMELTAHDATVEQLA
TAYASRAETGVSSRPEAAASGSLLEDAQRQQREELAEEVDGADAADEEERGEHLVEENEEDETPNSRRESTAASLAGMHAL
VTALHDDPRSMHPLPGFFGRFMILTRRNVTLAWRDRNVVFMVACNALLVAMLASVLFFNVYQEKAIMYHLSVFFLVSLSLA
SIPAVNMAMYLEKKVSYLYEASDGFYTAGPYLLSEMTTGFLMLLGATLLVFVVVFPCCAFPFSKFGMVFLLFILFLLVVDAV
MQLAAAMCATFIMASTFAGGWLALTSVVNGYNANPKSMPAWLTWLVYLSPFYYLMDGVAIVLCWENQDLFGTPEKAAK
YGYTSCEELLKAYGFAGTLGGSSLTPPQWLWSVDVPVLCLMTVGMKTFACFYQAYCVRLRR

Appendix II Restriction summary results

TgABC-A

#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AcuI	CTGAAG(N)$_{14}$NN	list	1598/1596
2	BamHI	G^GATC_C	list	2418/2422
3	BbsI	GAAGACNN^NNNN_	list	1637/1641
4	BfuAI	ACCTGCNNNN^NNNN_	list	*330/334
5	BsaAI	YAC_GTR	list	*2433
6	BseRI	GAGGAG(N)$_{8}$NN	list	2602/2600
7	BspHI	T^CATG_A	list	2279/2283
8	BspMI	ACCTGCNNNN^NNNN_	list	330/334
9	BsrFI	R^CCGG_Y	list	*303/307
10	BsrGI	T^GTAC_A	list	1181/1185
11	BssSI	C^ACGA_G	list	890/894
12	BstBI	TT^CG_AA	list	*2258/2260
13	BstYI	R^GATC_Y	list	2418/2422
14	EagI	C^GGCC_G	list	*2492/2496
15	MluI	A^CGCG_T	list	*290/294
16	NcoI	C^CATG_G	list	552/556
17	NruI	TCG_CGA	list	*1648
18	PflFI	GACN^N_NGTC	list	*1375/1376
19	PflMI	CCAN_NNN^NTGG	list	2009/2006
20	PpuMI	RG^GWC_CY	list	219/222
21	SacI	G_AGCT^C	list	2634/2630
22	SbfI	CC_TGCA^GG	list	992/988
23	Tsp509I	^AATT_	list	1432/1436
24	Tth111I	GACN^N_NGTC	list	1375/1376

TgABC-B

#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	Acc65I	G^GTAC_C	list	*114/118
2	AccI	GT^MK_AC	list	223/225
3	AfeI	AGC_GCT	list	*978
4	AgeI	A^CCGG_T	list	*117/121
5	AvrII	C^CTAG_G	list	1749/1753
6	BbsI	GAAGACNN^NNNN_	list	1901/1905
7	BdaI	_NN^(N)$_{10}$TGA(N)$_{4}$TCA(N)$_{10}$_NN^	list	#1197/1195+1231/1229
8	BglI	GCCN_NNN^NGGC	list	1471/1468
9	BglII	A^GATC_T	list	1586/1590
10	BmtI	G_CTAG^C	list	2050/2046
11	BsaAI	YAC_GTR	list	*807
12	BseYI	C^CAG_C	list	*777/781
13	BsmI	GAATG_CN^	list	1474/1472
14	BspEI	T^CCGG_A	list	*997/1001
15	BspHI	T^CATG_A	list	1190/1194
16	BsrBI	CCG_CTC	list	*641
17	BsrFI	R^CCGG_Y	list	*117/121
18	BstAPI	GCAN_NNN^NTGC	list	109/106
19	BtgZI	GCGATG(N)$_{10}$^NNNN_	list	*1008/1012
20	BtsI	GCAGTG_NN^	list	1321/1319
21	DrdI	GACNN_NN^NNGTC	list	1579/1577
22	EaeI	Y^GGCC_R	list	*1224/1228
23	EciI	GGCGGA(N)$_{8}$_NN^	list	*761/759
24	EcoNI	CCTNN^N_NNAGG	list	476/477
25	FalI	_(N)$_{5}$^(N)$_{3}$AAG(N)$_{5}$CTT(N)$_{3}$_(N)$_{5}$^	list	282/277+314/309
26	FauI	CCGCNNNN^NN_	list	*252/254
27	FspAI	RTGC_GCAY	list	*1263
28	FspI	TGC_GCA	list	*1263
29	KasI	G^GCGC_C	list	*588/592
30	KpnI	G_GTAC^C	list	118/114
31	MfeI	C^AATT_G	list	548/552
32	NarI	GG^CG_CC	list	*589/591
33	NciI	CC^S_GG	list	*1392/1393
34	NcoI	C^CATG_G	list	1831/1835
35	NdeI	CA^TA_TG	list	606/608
36	NheI	G^CTAG_C	list	2046/2050

TgABC-C

#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AccI	GT^MK_AC	list	*1946/1948
2	AcuI	CTGAAG(N)$_{14}$_NN^	list	2929/2927
3	AflIII	A^CRYG_T	list	2285/2289
4	AhdI	GACNN_N^NNGTC	list	22/21
5	AlwNI	CAG_NNN^CTG	list	#1142/1139
6	ApaI	G_GGCC^C	list	#1132/1128
7	BaeGI	G_KGCM^C	list	#1132/1128
8	BciVI	GTATCC(N)$_{6}$_N^	list	1158/1157
9	BglI	GCCN_NNN^NGGC	list	*1028/1025
10	BglII	A^GATC_T	list	2838/2842
11	BmtI	G_CTAG^C	list	1972/1968
12	BpuEI	CTTGAG(N)$_{14}$_NN^	list	748/746
13	BsaWI	W^CCGG_W	list	2702/2706
14	BseYI	C^CAG_C	list	*559/563
15	BsmI	GAATG_CN^	list	592/590
16	BsrFI	R^CCGG_Y	list	*2105/2109
17	BssSI	C^ACGA_G	list	1600/1604
18	EaeI	Y^GGCC_R	list	*#1641/1645
19	EcoNI	CCTNN^N_NNAGG	list	1297/1298
20	EcoRV	GAT_ATC	list	*390
21	FspI	TGC_GCA	list	*13
22	HincII	GTY_RAC	list	*1947
23	MfeI	C^AATT_G	list	683/687
24	NheI	G^CTAG_C	list	1968/1972
25	PciI	A^CATG_T	list	2285/2289
26	PflFI	GACN^N_NGTC	list	712/713
27	PspOMI	G^GGCC_C	list	#1128/1132
28	PspXI	VC^TCGA_GB	list	*1297/1301
29	PvuI	CG_AT^CG	list	*1760/1758
30	SacI	G_AGCT^C	list	827/823
31	SacII	CC_GC^GG	list	*2605/2603
32	SalI	G^TCGA_C	list	*1945/1949
33	SgrAI	CR^CCGG_YG	list	*2105/2109
34	SphI	G_CATG^C	list	2862/2858
35	StuI	AGG_CCT	list	#1379
36	Tth111I	GACN^N_NGTC	list	712/713

TgABC-D

#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AfeI	AGC_GCT	list	*1394
2	AflIII	A^CRYG_T	list	1901/1905
3	AleI	CACNN_NNGTG	list	#1039
4	AlfI	_NN^(N)$_{10}$GCA(N)$_{6}$TGC(N)$_{10}$_NN^	list	797/795+831/829
5	BbvCI	CC^TCA_GC	list	*2171/2174
6	BglII	A^GATC_T	list	675/679
7	BmrI	ACTGGGNNNNN_N^	list	1798/1797
8	BsaXI	_NNN^(N)$_{9}$AC(N)$_{5}$CTCC(N)$_{7}$_NNN^	list	1174/1171+1204/1201
9	BsiHKAI	G_WGCW^C	list	#1036/1032
10	BsmFI	GGGAC(N)$_{10}$^NNNN_	list	*1749/1753
11	BsmI	GAATG_CN^	list	404/402
12	BspEI	T^CCGG_A	list	*240/244
13	BspHI	T^CATG_A	list	#2009/2013
14	BspQI	GCTCTTCN^NNN_	list	1577/1580
15	BsrFI	R^CCGG_Y	list	*34/38
16	BstAPI	GCAN_NNN^NTGC	list	10/7
17	BstBI	TT^CG_AA	list	*799/801
18	BstEII	G^GTNAC_C	list	1674/1679
19	BstYI	R^GATC_Y	list	675/679
20	BtgI	C^CRYG_G	list	155/159
21	BtgZI	GCGATG(N)$_{10}$^NNNN_	list	*1806/1810
22	EagI	C^GGCC_G	list	*184/188
23	EcoNI	CCTNN^N_NNAGG	list	1654/1655
24	EcoRV	GAT_ATC	list	366
25	FspI	TGC_GCA	list	*1384
26	MluI	A^CGCG_T	list	*1901/1905
27	MseI	T^TA_A	list	1799/1801
28	MslI	CAYNN_NNRTG	list	#1039
29	NaeI	GCC_GGC	list	*36
30	NciI	CC^S_GG	list	*512/513
31	NgoMIV	G^CCGG_C	list	*34/38
32	NruI	TCG_CGA	list	*886
33	PflFI	GACN^N_NGTC	list	*1174/1175
34	PflMI	CCAN_NNN^NTGG	list	2039/2036
35	PshAI	GACNN_NNGTC	list	1561
36	PsrI	_(N)$_{6}$^(N)$_{7}$GAAC(N)$_{4}$TAC(N)$_{7}$_(N)$_{6}$^	list	*2209/2204+2241/2236

TgABC-E

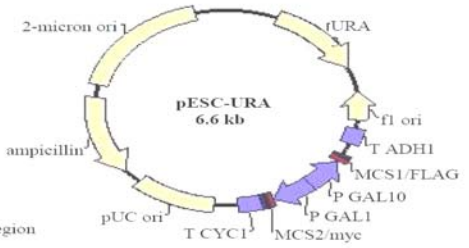
#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AarI	CACCTGCNNNN^NNNN▲	list	200/204
2	AclI	AA^CG▲TT	list	*1212/1214
3	AflIII	A^CRYG▲T	list	1751/1755
4	AhdI	GACNN▲N^NNGTC	list	207/206
5	AjuI	▲(N)5^(N)9GAA(N)7TTGG(N)14▲(N)5^	list	1555/1550+1587/1582
6	AloI	▲(N)5^(N)9GAAC(N)6TCC(N)14▲(N)5^	list	*548/543+580/575
7	BaeGI	G▲KGCM^C	list	1049/1045
8	BamHI	G^GATC▲C	list	2275/2279
9	BanII	G▲RGCY^C	list	1747/1743
10	BclI	T^GATC▲A	list	#1694/1698
11	BdaI	▲NN^(N)10TGA(N)4TCA(N)10▲NN^	list	1131/1129+1165/1163
12	BmgBI	CAC▲GTC	list	*1736
13	BpuEI	CTTGAG(N)14▲NN^	list	125/123
14	BsaAI	YAC▲GTR	list	*1670
15	BsaBI	GATNN▲NNATC	list	254
16	BsmFI	GGGAC(N)10^NNNN▲	list	239/243
17	BspDI	AT^CG▲AT	list	*1315/1317
18	BsrDI	GCAATG▲NN^	list	795/793
19	BsrFI	R^CCGG▲Y	list	*1326/1330
20	BssHII	G^CGCG▲C	list	*1527/1531
21	BssSI	C^ACGA▲G	list	713/717
22	BtgI	C^CRYG▲G	list	1823/1827
23	ClaI	AT^CG▲AT	list	*1315/1317
24	EcoNI	CCTNN^N▲NNAGG	list	2038/2039
25	FauI	CCCGCNNNN^NN▲	list	*1850/1852
26	FspAI	RTGC▲GCAY	list	*1294
27	HindIII	A^AGCT▲T	list	937/941
28	MluI	A^CGCG▲T	list	*1751/1755
29	NcoI	C^CATG▲G	list	1823/1827
30	PflFI	GACN^N▲NGTC	list	*475/476
31	PpiI	▲(N)5^(N)9GAAC(N)5CTC(N)14▲(N)5^	list	*548/543+580/575
32	PpuMI	RG^GWC▲CY	list	330/333
33	PshAI	GACNN▲NNGTC	list	*840
34	PspXI	VC^TCGA▲GB	list	*999/1003
35	PvuI	CG▲AT^CG	list	*1473/1471
36	PvuII	CAG▲CTG	list	1132

TgABC-F

#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AvrII	C^CTAG▲G	list	1146/1150
2	BaeI	▲(N)5^(N)10ACNNNNGTAYC(N)7▲(N)5^	list	2196/2191+2229/2224
3	BamHI	G^GATC▲C	list	111/115
4	BanI	G^GYRC▲C	list	*1354/1358
5	BciVI	GTATCC(N)5▲N^	list	2197/2196
6	BfaI	C^TA▲G	list	1147/1149
7	BfuAI	ACCTGCNNNN^NNNN▲	list	371/375
8	BmrI	ACTGGGNNNNN▲N^	list	2299/2298
9	Bpu10I	CC^TNA▲GC	list	952/955
10	BsaXI	▲NNN^(N)9AC(N)5CTCC(N)7▲NNN^	list	1765/1762+1795/1792
11	BsgI	GTGCAG(N)16▲NN^	list	360/358
12	BsiWI	C^GTAC▲G	list	*1238/1242
13	BspDI	AT^CG▲AT	list	*#1504/1506
14	BspHI	T^CATG▲A	list	1538/1542
15	BspMI	ACCTGCNNNN^NNNN▲	list	371/375
16	BsrGI	T^GTAC▲A	list	1103/1107
17	BstBI	TT^CG▲AA	list	*763/765
18	BstEII	G^GTNAC▲C	list	906/911
19	Bsu36I	CC^TNA▲GG	list	686/689
20	BtgZI	GCGATG(N)10^NNNN▲	list	*1756/1760
21	ClaI	AT^CG▲AT	list	*#1504/1506
22	DraIII	CAC▲NNN^GTG	list	1395/1392
23	DrdI	GACNN▲NN^NNGTC	list	*318/316
24	EcoNI	CCTNN^N▲NNAGG	list	1303/1304
25	EcoO109I	RG^GNC▲CY	list	2249/2252
26	EcoRI	G^AATT▲C	list	539/543
27	HaeII	R▲GCGC^Y	list	*976/972
28	MseI	T^TA▲A	list	2317/2319
29	MslI	CAYNN▲NNRTG	list	1741
30	NmeAIII	GCCGAG(N)16▲NN^	list	266/264
31	NruI	TCG▲CGA	list	*1231
32	NsiI	A▲TGCA^T	list	1511/1507
33	PaeR7I	C^TCGA▲G	list	*1245/1249
34	PmlI	CAC▲GTG	list	*2093
35	PpuMI	RG^GWC▲CY	list	2249/2252
36	PshAI	GACNN▲NNGTC	list	1261

Appendix III

Diagram of pESC-Ura vector



pESC-URA Multiple Cloning Site 1 Region

(Sequence shown 2180-2263, bottom strand)

GAA TTC AAC CCT CAC TAA AGG GCG GCC GCA CTA GTA TCG ATG GAT TAC AAG GAT GAC GAC GAT AAG ATC TGA GCTCTTAATTAA

M D Y K D D D D K I stop

FLAG epitope

pESC-URA Multiple Cloning Site 2 Region

(Sequence shown 2936-3033, top strand)

G GAT CCG TAA TAC GAC TCA CTA TAG GGC CCG GGC GTC GAC..

stop

...ATG GAA CAG AAG TTG ATT TCC GAA GAA GAC CTC GAG TAA GCTTGGTACCGCGGCTAGC

M E Q K L I S E E D L E stop

myc epitope

Feature	Nucleotide Position
Yeast *URA3* selection marker ORF	417-1217
f1 origin of ss-DNA replication	1483-1789
Yeast *ADH1* terminator	1885-2049
Multiple cloning site 1	2180-2263
FLAG tag	2198-2224
Yeast *GAL1/GAL10* divergent promoters	2268-2934
Multiple cloning site 2	2936-3033
c-myc tag	2976-3011
Yeast *CYC1* terminator	3038-3227
pUC origin of replication	3414-4081
Ampicillin resistance (*bla*) ORF	4232-5089
2μ yeast origin of replication	5223-6378

Above Figure is pESC-URA vector. The complete sequence and list of restriction sites are available at *www.stratagene.com*. Note that the sequence shown for MCS1 corresponds to the bottom strand sequence (the reverse-complement of the sequence available for this vector at *www.stratagene.com*).

Appendix IV

TgABC-B- pESC-Ura sequencing with GAL10-Fwd (Initiation Methionine is highlighted yellow)

AAGAATTTTTGAAAATTCGAATTCAACCCTCACTAAAGGGCGGCCGCATGCCCTCGCATCCTAATTTGGAGAGTACTAC
ACCGCCTGTCCACGACCACAGTGGTGGGCAGCCAACGGAGTCCTTAACTTCAATTTGTCAGAAGAAGATTGCAGACTAT
GCGGTACCGGTGGAGAATTACTGGTCAACTGACCGTCTTTCGCCAGTGGCTCTTGACAGAGAGACCACGACTGTTTCAG
AGCCACGCCCCATCCCCTCTTGCTATGCACTGTCTACCATAGGTAGCCGGAGTCCCGCGGCTGAGGTCAGAAACTCCTC
AGCTTCCAGATGTGGAGGTGCAAGTCCACCTTCGCCTGAAAACGGTGCTGCACGAAGTGGGGTTGTTCAGGTGCTCACA
CCGCCTCGCGTTGTCACTCTTCATAAGGAAGCCGATGATTCTGTACAGATGAAAGAGACCGAACAAGCGAAGCGTGCT
GCAAGGCTACAAGTGTTCATACGCTTCACAAACCTCACATTCACTCCTCCGCGAGGCCTTGCACGCCAATCCGTGAAAG
AATGGGCTAAAACGGCTTCCTGTACACAGTCGTCACGACAGGCAATTGCTGGTTCCGATTCTGTTTCGGGACCGTGTTC
AAAGGCGCCTCCTCCGTCCGCATATGCTCGCCCCCGACAAATTCTCTTTGGGCTGAGCGGCTACTTTGCCCCTGGCGAA
ATTGTGGGAATCCTCGGTCCTTCTGGAGCAGGCAAAAGCA

MPSHPNLESTPPVHHSGGQPTESLSICQKKIADYAVPVENYWSTDRLSPVALDRETTTVSEPRIPSCYALSTIGSRSPAAEVRN
SSASRCGGASPPSPENGAARSGVVQVLTPRVVTLHKEADSVQMKETEQAKRAARLQVFIRFTNLTFTPPRGLARQSVKEWA
KTASCTQSSRQAIAGSDSVSGPCSKAPPPSAYARPRQILFGLSGYFAPGEIVGILGPSGAGKS

TgABC-B-pESC-Ura sequencing with GAL10-Rev (Reverse complement) (Termination codon is highlighted yellow)

ACTTCACTACTTCTTACTACAGCTTCGGACCTGCCTACATGGCATTCACATCTTTTCCAGCTGAGCGCGCCGTTATCTCC
AGAGAAAGAAGCAGTAAGGCATATAAAGTGTCATGTTATCTTCTCGCCAAAACCATGATAGACCTCGTCGTCCAGATCT
TTACTCCGTCACTGTGGCTGGCAATCGTATATCCGCTTGTCGGATTGCCCTCAGACCTTGGCGTGTTTATGGCTTTTTGG
GCTCAGCTGGTACTGCTTGTCTGCATAGCTCAAGCCATAGGTCAACTGATTGCTGCGATCGTTGTGGACGATGCGCGCC
TAGGGGGTTTGCTGTTGTCAGTCATTCTGATTTCATCTTCCATCAGTAGTGGCTTCTATGTTCAGCAACAGAGATTAGGC
CCATGGATATCTTGGTTTCGATGGTTGTCTTTCCAGAACTACGCGGTGACAAACTTCGTAATAGTTACAGTGGGATCGTC
TTCGACACTCTCGTGCTCGGAGTTCTCTTCGTTCCCAACTTGCCCAGGGCAGCCGATCACAGCAGAGATGATTACCCGC
AGGTTCACAACAGCACTTTCGCCTTTATCCAATATCACCTTAATGATTTGTACATGGCTAGCCATCAAACTAATGTGCTA
TGGAGTCCTCAAGAAATCTCTTAAACTGAAGTCCTGAGCGGCCGCACTAGTATCGATGGATTACAAGGATGACGACGA
TAAGATCTGAGCTCTTAATTAACAATCTCGCCAGA

FTTSYYSFGPAYMAFTSFPAERAVISRERSSKAYKVSCYLLAKTMIDLVVQIFTPLWLAIVYPLVGLPSDLGVFMAFWAQLVL
LVCIAQAIGQLIAAIVVDDARLGGLLLSVILISSSISSGFYVQQQRLGPWISWFRWLSFQNYAVTNFVIVTVGSSSTLSCSEFSF
PTCPGQPITAEMITRRFTTALSPLSNITLMICTWLAIKLMCYGVLKKSLKLKS

Appendix V

Buffer and Media:

E. coli competent cell buffers

TFB-I Buffer	Ingredients	TFB-II Buffer	Ingredients
$CaCl_2$	10mM	$CaCl_2$	75mM
Glycerol	15%	Glycerol	15%
KOAc,PH 5.8	30mM	MOPS,PH 5.8	10mM
$MnCl_2 4H_2O$	50mM	RbCl Autoclave	100mM
RbCl Autoclave	100mM	Store 4 ^{o} C	
Store 4 ^{o} C			

E. coli transformation media

SOB Media	Ingredients
NaCl	500mg
KCl	186mg
Tryptone	20g
Yeast Extract	5g
ddH_2O	Ad 1 L
Autoclave	

Add sterile $MgCl_2$ solution to final concentration of 10mM.

E. coli culture

LB Media	Ingredients	LB Plates	Ingredients
NaCl	10g	NaCl	10g
Tryptone	10g	Tryptone	10g
Yeast Extract	5g	Yeast Extract	5g
ddH_2O	Ad 1 L	Agar	15g
Autoclave		ddH_2O	Ad 1 L
		Autoclave	

*LB:Laria Bartani.100ug/ml Ampicillin) was added in a media cooled to 55 °C.

S. cerevisiae competent cell buffers

10x TFB-I Buffer	Ingredients	10x LiAc Buffer	Ingredients
Tris-HCl	10mM	Lithium acetate	1 M
EDTA	100mM		
Autoclave			

50%polyethylene glycol (PEG3350) solution: Filter sterilize

LiAc-TE buffer: 1 part 10x LiAc, 1 part 10x TE buffer, 8 part ddH2O, Autoclave.

(PEG3350)- LiAc-TE buffer: 1 part 10x LiAc, 1 part 10x TE buffer, 8 part 50%PEG3350, Prepare fresh.

S. cerevisiae transformation media: Dimethlsulfoxid (DMSO)

S. cerevisiae culture

YPD Media	Ingredients
Dextrose	20g
Peptone	20g
Yeast Extract	10g
ddH_2O	Ad 1 L
Autoclave	

YPD: Yeast peptone dextrose

SD Media	Ingredients	SD Plates	Ingredients
Yeast nitrogen base	0.85g	Yeast nitrogen base	0.85g
Ammonium sulphate	2.50g	Ammonium sulphate	2.50g
ddH_2O		Agar	10g
Autoclave	Ad 425 ml	ddH_2O	Ad 425 ml
		Autoclave	

*SD: Synthetic dextrose minimal media (SD drop out media)
Add 1x amino acid and 2% appropriate sugar.

10x amino acid mix

Amino Acid	Concentration (mg/liter)	Amino Acid	Concentration (mg/liter)
Adenine sulfate	40	L-phenylalanine	50
L-arginine (HCl)	20	L-serine	375
L-aspartic acid	100	L-threonine	200
L-glutamic acid	100	L-tryptophan *	40
L-histidine *	20	L-tyrosine	30
L-leucine *	60	L-valine	150
L-lysine	30	Uracil *	20
L-methionine	20		

*The omission of histidine, leucine, tryptophan, and uracil allows selection of plasmids harboring the cDNA of interest.

40%suger stock (glucose) in ddH_2O: Filter sterilizes and Store at 4^{o} C

Agarose gel electrophoresis

50x TAE buffer Ph 8.0	Ingredients
Acetic acid	57.1ml
EDTA	0.5M
Tris-HCl	242g
ddH_2O	Ad 1L
Autoclave	