Parasitol Res (2003) 91: 283–289
DOI 10.1007/s00436-003-0962-y

ORIGINAL PAPER

Ji-Sook Lee · Jongweon Lee · Soon-Jung Park
Tai-Soon Yong

Analysis of the genes expressed in *Clonorchis sinensis* adults using the expressed sequence tag approach

Received: 21 February 2003 / Accepted: 10 April 2003 / Published online: 22 August 2003

Abstract *Clonorchis sinensis* is a biliary tract parasite, which infects over 30 million people in China, Korea and Southeast Asia through the ingestion of undercooked freshwater fish that harbour the infective metacercariae. The genes expressed in *C. sinensis* adults were identified in order to develop novel drugs, better diagnostics and vaccines for the parasite. The *C. sinensis* cDNA library was constructed and DNA sequencing was performed with 450 randomly selected clones. Four hundred and fifteen clones contained the amino-acid-encoding sequences. The functions of these genes could be assigned by DNA sequence homology. Basic Local Alignment Search Tool X analysis showed that 277 of the 415 clones were strongly matched ($P < 10^{-9}$) to previously identified proteins, while the remaining 138 fell into the "no database match" category. Among the clones matching previously identified proteins, 220 putatively identified genes were sorted into seven functional categories. These included the genes associated with energy metabolism (38), gene expression/RNA metabolism (21), regulatory/signalling components (14), protein metabolism/sorting (98), the structure/cytoskeleton (29), membrane transporters (ten) and antigenic proteins (ten). The remaining 57 clones were not included in these categories. The dataset included the genes encoding the proteases, a lipid binding protein, the antigen proteins and the other genes of interest from a diagnostics, drug or vaccine development viewpoint. The present expressed sequence tag analysis proved to be an effective tool for examining gene expression and identified several important genes which increase and complement our knowledge of the biology of *C. sinensis*.

J.-S. Lee · J. Lee · S-J. Park · T.-S. Yong (✉)
Department of Parasitology,
Institute of Tropical Medicine and
BK 21 Project for Medical Science,
Yonsei University College of Medicine,
Seodaemungu Shinchondong 134, Seoul 120–752, South Korea
E-mail: tsyong212@yumc.yonsei.ac.kr
Tel.: +82-2-3615297
Fax: +82-2-3638676

Introduction

Clonorchis sinensis is an important human parasite in eastern Asia, including China, Taiwan, northern Vietnam and Korea. In Korea, it is well known that human infection of *C. sinensis* is widely distributed along the rivers and streams (Rim 1998; Crompton et al. 1999). When the fluke infects humans, the bile duct is severely dilated and the ductal wall is thickened due to mucosal hyperplasia and fibrosis. As the infection becomes chronic and heavier in intensity, complications such as obstructive jaundice, dull epigastric pain, biliary stones, ascites and cholangiocarcinoma can develop (Chapman et al. 1999; Kim et al. 1993).

Over the past decade, the world has witnessed the emergence and progress of several genome projects. The last release of the db expressed sequence tags (ESTs) in 1999 contained a prodigious number of entries for parasite genomes including >20,000 ESTs for *Brugia malayi*, ~12,500 for *Schistosoma mansoni*, and over 12,000 for *Trypanosoma cruzi*. In addition, an increasing number of ESTs for other parasites including *Paragonimus westermani*, *Strongyloides stercoralis*, *Toxoplasma gondii* and *Aedes aegypti* etc., have been identified (Paul 2000). Analysis of ESTs has proven to be a rapid and efficient method for characterizing the subset of genes which are expressed in a life-stage specific manner in a wide variety of tissues and organisms (Adams et al. 1995). Gaining knowledge on the genome of parasites is increasingly important in order to understand a parasite's biology, the drug resistance mechanism and antigenic variations that determine the escape from a host's immune system (Franco et al. 2000; Tawe et al. 2001). Although a *C. sinensis* infection is diagnosed by identifying its eggs through a microscopic examination of stools, immunological methods have recently been adopted for its diagnosis (Kang et al. 1969). The

assembly of EST data specific to a protein or a DNA sequence can be used to rapidly produce a gene/protein model to understand the physical and biological properties of *C. sinensis*, and is useful for immunodiagnosis, determining novel target drugs and identifying potential vaccine candidates.

Materials and methods

Collection of parasite

The *C. sinensis* metacercariae were collected from naturally infected fish (*Pseudorasbora parva*) caught in the Nakdong River, Korea. These metacercariae were orally administered to the experimental rabbits. Eight weeks later, adult *C. sinensis* worms were obtained from the rabbit bile ducts.

Construction of the cDNA library

C. sinensis mRNA was purified by using a messenger RNA isolation kit (Stratagene, La Jolla, Calif.). Briefly, live *C. sinensis* worms recovered from the infected rabbits were homogenized in a denaturing solution of 4 M guanidium isothiocyanate and 0.14 M β-mercaptoethanol, and centrifuged at 12,000 *g* for 10 min. The supernatant was transferred to a tube containing oligo (dT) cellulose resin. The resin was washed with a high-salt buffer, and the poly (A) + mRNA was eluted with a low-salt buffer. A cDNA library of *C. sinensis* was constructed using a ZAP Express cDNA Gigapack II gold cloning kit (Stratagene) according to the manufacturer's instructions. Briefly, the first strand cDNA was synthesized on 5 μg of *C. sinensis* mRNA and the second strand was synthesized by a nick translation. The *Eco*RI adaptors were ligated to the blunt ends. *Xho*I restriction enzyme digestion resulted in directional cDNA. The cDNA was then inserted into the predigested ZAP express vector arms. Packaging was carried out in vitro with Gigapack II packaging extract. The library was plaqued on *Escherichia coli* XL1-Blue MRF′.

Sequencing

The phage library was converted to the plasmid form by a mass excision according to the reported protocol (Stratagene). The obtained phagemid of the library was used to infect the *E. coli* strain XLOLR. The bacteria were grown for 45 min and then plated at a low density on a medium containing a Luria-Bertani broth including tetracycline (10 mg/l). The bacteria were cultured at 37°C overnight, and individual colonies were selected randomly for plasmid DNA purification and sequencing. All the sequencing reactions contained the T3 sequencing primer, and were read into the 5′ end of each DNA. The reactions were run and analysed on capillary automated sequencing machines (ABI 377 XL90; Applied Biosystem, USA). The machines generated two computer files, a chromatogram file and a plain text file.

The EST nucleotide sequences and completed cDNA sequence data reported in this paper were submitted to the GenBank dbEST database under the accession nos. AF480464, AF480465, AF527454–AF527457, AT006727–AT006731, AT006733–AT006735, AT006737–AT006740, AT006742–AT006744, AT006746, AT006749, AT006753–AT006755, AT006757, AT006760–AT006762, AT006764, AT006765, AT006768, AT006769, AT006772–AT006775, AT006777, AT006778, AT006780, AT006781, AT00678, AT006784–AT006786, AT006790, AT006797, AT006798, AT006800, AT006801, AT006804, AT006806, AT006809, AT006811, AT006817, AT006818, AT006820, AT006823, AT006833, AT006834, AT006838, AT006840, AT006844, AT006846, AT006847, AT006851, AT006853–AT006855, AT006859, AT006862, AT006864, AT006869, AT006870, AT0068 72, AT006873, AT006879 AT006884, AT006888, AT006891, AT006894, AT006898, AT006902–AT006904, AT006906, AT006907, AT006909, AT006911, AT006912, AT006914–AT006929, AT006931–AT006974.

Homology comparisons

Each edited EST was translated in all six reading frames and compared with the non-redundant database at the National Centre for Biotechnology Information using the Basic Local Alignment Search Tool X (BLASTX) program, which compares the translated nucleotide sequences with the protein sequences. The homologies to the negative reading frames, with the exception of clones with the insert in the reverse orientation, were disregarded. Putative identification of the ESTs was based on the BLAST searches and in some cases on the information contained in the MEDLINE database.

Results

Sequencing of ESTs

Of the 450 sequencing reactions attempted, 415 produced readable amino acid-encoding sequences from the *C. sinensis* adult cDNA library (Table 1). The leading and tailing vectors as well as the poor-quality sequences were trimmed from each text file. The 3′ vector and linker sequences were removed if the poly (A^+) tails could have been included in the sequencing results. Three classes of anomalous sequences were also excluded; sequences without the insert, sequences with the reverse inserts and the incorrect adaptor. The insert size of the clones ranged from 400 to 3,000 bp, and the mean size was 627 bp.

Putative identification of EST sequences

BLASTX analysis showed that 277 of the 415 clones were strongly matched ($P < 10^{-9}$) to other proteins. On the basis of database searches, the 277 different ESTs were classified two groups. The remaining 138 clones fell into a "no database match" category (Table 1). The first class comprised 53% (220 clones) of the sequences with an average read of 562 ± 95 bp, and were matched to the genes with the predicted or known functions of *C. sinensis* and other organisms. Forty-one clones of these cDNAs were matched with *C. sinensis* and are related to cysteine protease, glutathione-S transferase 28 kDa and the antigen proteins. The other organisms included helminths (*Paragonimus westermani*, *Fasciola hepatica*, *Schistosoma japonicum*, *Schistosoma mansoni*, *Caenorhabditis elegans*), mammals (*Homo sapiens*, *Rattus norvegicus*, *Mus musculus*), insects (*Drosophila melanogaster*) and others. Secondly, 14% (57 clones) of the sequences with an average sequence read of 548 ± 93 bp were matched with *C. elegans*, *H. sapiens*, *D. melanogaster*, *M. musculus*, and *P. westermani* with no assigned function. The remaining 33% of the sequence read of 417 ± 10 bp had no significant match in the database. The first class was sorted into seven functional categories, which included

Table 1 Expressed sequence tags (*EST*s) of the *Clonorchis sinensis* adult cDNA library[a]

EST match category	No. of cDNA clones (%)
Total	415 (100)
Database match[a]	277 (67)
Putative function/domain[b]	220 (53)
No function	57 (14)
No database match[c]	138 (33)

[a]Significant match to other organisms ($P < 10^{-9}$). Other organisms includes helminths (*Clonorchis sinensis, Paragonimus westermani, Fasciola hepatica, Schistosoma japonicum, Schistosoma mansoni, Caenorhabditis elegans*), mammals (*Homo sapiens, Rattus norvegicus, Mus musculus*), insects (*Drosophila melanogaster*) and others
[b]Matched to predicted or known function or protein domains
[c]No significant database matches

the genes associated with energy metabolism (38), gene expression/RNA metabolism (21), regulatory/signalling components (14), protein metabolism/sorting (98), structure/cytoskeleton (29), membrane transporter (ten), and antigen protein (ten) (Table 2). These were matched to either the ribosomal protein, cysteine protease or the heat shock protein. In addition, the genes related to energy metabolism were matched to those for the enzymes taking part in the glycolytic pathway, TCA cycle and oxidative phosphorylation. Among the other functional categories identified, genes encoding the proteins involved with transcription and translation and those associated with regulatory and signalling functions were identified.

Discussion

Even though the identification of novel parasite molecules using conventional methods are tedious, identifying some of these molecules might serve as starting points for various studies. However, the search for *C. sinensis* gene expression using the EST analysis performed in this study, showed many important genes which increase and complement our knowledge of the biology of *C. sinensis*.

Energy production in this fluke depends largely on glycolysis (Kang et al. 1969; Hong et al. 2000). Through glycolysis, it takes up approximately 1.13 mg glucose/h per gram wet weight, produces lactate, and forms several types of amino acids as the end product from exogenous glucose (Hong et al. 2000; Han et al. 1961). Therefore, the enzymes involved in glycolysis are essential for energy metabolism. In this work, the ESTs found represented a variety of proteins related to metabolism. It has been suggested that they produce ATPs through the glycolytic pathway or aerobic metabolism (Table 2; AT006919, AT006737, AT006806, AT006873, AF480465, AT006919, 006731, AT006921). However, an initial search of the genomic basis of the fundamental biochemical pathways of *C. sinensis* revealed that its biosynthetic networks are fairly consistent with those of humans.

The adenylate kinase 1 (AK 1, AF480464) homologous gene was identified from the ESTs. AK 1 is indispensable for *Escherichia coli* (Cronan et al. 1972) and *Schizosaccharomyces pombe* growth (Konrad et al. 1993), indicating that it is an essential enzyme for life in a single cell. Under normal conditions, AK 1-knockout mice showed no phenotypic changes. However, under metabolic stress, compromised energetics were detected in the heart and skeletal muscle, suggesting the physiological significance of AK-catalysed phosphoryl transfer between the intracellular compartments in cellular energetic homoeostasis (Qualtieri et al. 1997; Janssen et al. 2001). The nucleoside analogues of AK have been used clinically for treating certain viral infections and malignant diseases (Pucar et al. 2000; Bourdais et al. 1996; Schneider et al. 1998). Therefore, they can be also targeted for drug research with recombinant antigens.

Several genes encoding antioxidant and detoxification enzymes, such as Cu/Zn-superoxide dismutase, and glutathione-*S*-transferase were identified. These proteins are believed to play a crucial role in protecting the parasite from the host immune effector mechanisms, and are being pursued as drug targets in other parasitic infections (Selkirk et al. 1998; Smooker et al. 1999). Of the proteases identified from EST analysis, the high frequency of cysteine protease expression (30 out of 415 randomly selected clones) suggests it has an important role in the metabolism and/or pathogenesis of clonorchiasis. It has also been proposed as a target for a structure-based approach for drug design (Park et al. 2001; Song and Rege 1991). In addition, one protease inhibitor was also identified. Specific protease inhibitors have also been isolated from filarial nematodes, and are suggested to play a role in inhibiting the enzymes secreted from host immune cells, the blocking of antigen processing and the control of the endogenous proteases involved in parasite development (Zang et al. 1999; Yenbutr et al. 1995). The biological role of the protease inhibitor in *C. sinensis* requires further investigation. The ESTs identified in this study showed a significant number of homologous genes previously reported from closely related parasitic trematodes. A homologue of the *S. japonicum* fatty acid binding protein (FABP) was identified in *C. sinensis*, and it was demonstrated that the FABP with a cross-protective efficacy could be used as a vaccine against trematode infections, such as fasiolasis (Hillyer et al. 1987, 1988; Estuningsih et al. 1997) and schistosomiasis (Tendler et al. 1996). It was also found that the ESTs contained several antigenic protein genes for *C. sinensis* or other parasites. The recombinant glycine-rich *C. sinensis* protein and proline-rich antigen were reported previously to be useful for immunodiagnosis, with a high specificity (Yang et al. 2000; Kim et al. 2001).

AT006742 and AT006744 were found to be homologous to the *Schistosoma* tegumental antigens. Sm 20.8 was reported to be a member of a family of soluble tegument antigens that contained the EF-hand motifs, and were recognized as antigenic targets with protective

Table 2 Significant matches of *C. sinensis* adult ESTs with sequences present in DNA and protein databases[a]

EST	Accession no.	Putative identification	Species	No. of clones	Length of EST[b] (bp)	e-value
Structure/cytoskeleton (29)						
CS478	AT006944	Actin	*Crassostrea gigas*	1	624	e-105
CS128	AT006733	Alpha tubulin	*Schistosoma mansoni*	5	653	e-109
CS8	AT006749	Alpha tubulin chain	*Lytechinus pictus*	2	564	2e-47
CS467	AT006945	Alpha-1 tubulin	*Hirudo medicinalis*	1	654	3e-88
CS72	AT006728	Alpha-3 tubulin chain	*Drosophila melanogaster*	1	668	e-110
CS28	AT006762	Alpha-tubulin	*Xenopus laevis*	1	630	3e-80
CS98	AT006730	Beta-1 tubulin chain	*Physarum polycephalum*	1	635	3e-83
CS214	AT006740	Beta-tubulin	*Halocynthia roretzi*	5	549	7e-83
CS200	AT006851	Dynein light chain	*Schistosoma mansoni*	1	395	3e-21
CS232	AT006864	Fibrillar collagen	*Arenicola marina*	1	603	7e-15
CS334	AT006946	Fragile x-related protein	*Danio rerio*	1	632	2e-06
CS38	AT006947	Inner dynein arm right chain	*Strongylocentrotus purpuratus*	1	439	1e-39
CS469	AT006948	Intermediate chain 1	*Anthocidaris crassispina*	1	623	4e-33
CS220	AT006859	Paramyosin-related protein	*Echinococcus granulosus*	1	553	3e-07
CS185	AT006840	Prosaposin precursor	*Danio rerio*	1	607	6e-06
CS331	AT006949	Testicular microtubules-related protein	*Homo sapiens*	1	393	1e-22
CS433	AT006950	Thrombospondin precursor	Chicken	1	602	2e-13
CS475	AT006951	Vitelline protein B1	*Fasciola hepatica*	1	560	7e-30
CS177	AT006834	Vitelline protein B2		1	608	1e-21
CS471	AT006937	Female genital complex protein precursor		1	547	1e-13
Membrane transporter (10)						
CS328	AT006912	Hexosaminidase b	*Mus musculus*	1	626	2e-28
CS118	AT006800	Integral membrane protein 23	*Schistosoma mansoni*	1	653	2e-22
CS290	AT006894	Na^+/K^+-exchanging ATPase beta chain	Sheep	1	658	1e-26
CS367	AT006952	Nucleoporin 153	*Takifugu rubripes*	1	524	1e-07
CS154	AT006818	Oxalate/formate antiporter (oxlt-2)	*Archaeoglobusfulgidus*	1	572	3e-06
CS248	AT006870	Propionyl coenzyme a carboxylase	*Mus musculus*	1	643	3e-49
CS159	AT006820	Na-dependent amino acid transporter	*Chlamydia trachomatis*	1	626	2e-15
CS401	AT006953	Transmembrane protein	*Caenorhabditis elegans*	1	646	7e-20
CS22	AF527455	V-ATPase G subunit		1	560	1e-09
CS89	AT006786	Endoplasmic reticulum lumen protein-retaining receptor	*Drosophila melanogaster*	1	733	4e-91
Gene expression/ RNA metabolism (21)						
CS181	AT006838	14–3-3 Protein epsilon	*Xenopus laevis*	1	605	3e-38
CS141	AT006811	Asparaginyl-tRNA synthetase	*Homo sapiens*	1	659	5e-47
CS436	AT006925	Aspartyl-tRNA synthetase	*Rattus norvegicus*	1	489	2e-27
CS438	AT006926	Casein kinase II alpha subunit	*Spodoptera frugiperda*	1	543	1e-81
CS209	AT006738	DNA supercoiling factor	Silkworm	1	659	3e-24
CS202	AT006853	DnaJ-like protein 1	*Mus musculus*	1	655	1e-04
CS244	AT006743	GTP-binding protein	*Discopyge ommata*	1	650	2e-73
CS473	AT006927	Human translation initiation factor 6	*Homo sapiens*	1	554	4e-66
CS400	AT006928	Kruppel homologue 2	*Drosophila melanogaster*	1	462	4e-06
CS91	AT006729	Microrchidia	*Mus musculus*	1	636	3e-42
CS497	AT006929	Mitochondrial dicarboxylate carrier		1	559	2e-31
CS278	AT006888	Phenylalanine-tRNA synthetase-like protein	*Homo sapiens*	1	663	5e-32
CS115	AT006798	SET translocation		1	659	6e-69
CS458	AT006931	Small glutamine-rich tetratricopeptide		1	556	2e-29
CS207	AT006855	Small nuclear ribonucleoprotein d1 polypeptide		1	548	2e-33
CS5	AT006746	Small Zn finger-like protein	Rat	1	316	3e-07
CS59	AT006777	Splicing factor 3b, subunit 2, 145kd	*Homo sapiens*	1	632	1e-05
CS24	AT006760	TB2		1	586	6e-39
CS461	AT006932	Transcription regulatory protein	*Oryza sativa*	1	617	1e-12
CS543	AT006933	Y-box binding protein homologue	*Schistosoma mansoni*	2	517	1e-31
Signalling components (14)						
CS63	AF480464	Adenylate kinase	*Schistosoma mansoni*	1	648	5e-73
CS73	AT006780	Ca^{2+}/calmodulin-dependent protein kinase I	*Homo sapiens*	1	646	8e-37
CS296	AF527456	Ca-binding protein	*Fasciola hepatica*	1	362	8e-15
CS297	AT006954	caltractin	*Homo sapiens*	1	651	2e-15
CS349	AT006955	cAMP-dependent protein kinase (PKA C)	*Caenorhabditis elegans*	1	619	8e-70
CS341	AT006956	*Chlamydomonas* radial spoke protein 3	*Homo sapiens*	1	634	2e-29
CS313	AT006907	Dopamine-responsive protein	*Rattus norvegicus*	1	560	2e-47
CS38	AT006957	GTP:AMP phosphotransferase	*Bos taurus*	1	468	4e-26

Table 2 (Continued)

EST	Accession no.	Putative identification	Species	No. of clones	Length of EST[b] (bp)	e-value
CS191	AT006844	GTP-binding protein ara2	*Arabidopsis thaliana*	1	602	4e-67
CS114	AT006797	Pka-C1 gene product	*Drosophila melanogaster*	1	653	8e-71
CS413	AT006958	Programmed cell death 6	*Homo sapiens*	1	628	2e-36
CS176	AT006833	Senescence-associated protein	*Pisum sativum*	1	628	4e-42
CS323	AT006911	SH3-domain kinase binding protein 1	*Homo sapiens*	1	588	1e-20
CS298	AT006898	Vesicle -associated membrane protein 2	*Xenopus laevis*	1	642	2e-16
Protein metabolism/sorting (98)						
CS129	AT006804	Brain specific protein	*Homo sapiens*	1	630	2e-16
CS268	AT006884	Calpain	*Drosophila melanogaster*	1	647	6e-72
CS4	AF093242	Cysteine protease	*Clonorchis sinensis*	30	678	e-119
CS227	AT006862	Cysteine and histidine-rich protein	*Mus musculus*	1	603	1e-05
CS96	AT006934	Cysteine-rich protease inhibitor		1	619	3e-14
CS47	AT006772	Cytosol aminopeptidase	*Schistosoma mansoni*	1	501	8e-15
CS554	AT006936	Elongation factor Tu homologue precursor	*Caenorhabditis elegans*	1	590	7e-47
CS370	AT006937	Elongation factor-1a-related protein	*Anthocidaris crassispina*	2	434	1e-70
CS308	AT006904	Eukaryotic translation initiation factor 4	*Homo sapiens*	1	662	1e-06
CS33	AT006765	Ferritin-2 heavy chain	*Schistosoma mansoni*	1	496	7e-45
CS261	AT006879	Glycoprotein 96–92	*Leishmania major*	1	645	6e-08
CS282	AT006891	Granulin	*Homo sapiens*	1	448	9e-11
CS119	AT006801	Heat shock protein 110	*Strongylocentrotus franciscanus*	1	657	5e-55
CS16	AT006755	Heat shock protein 60	*Drosophila melanogaster*	2	604	4e-77
CS496	AT006938	Heat shock protein 67B2	*Mus musculus*	1	489	3e-12
CS546	AT006939	Heat-responsive protein		1	498	9e-38
CS521	AT006940	Heme binding protein 2		1	553	9e-22
CS465	AT006941	Hemoglobinase	*Schistosoma japonicum*	1	544	4e-42
CS510	AT006942	Hsp89-alpha-delta-N	*Homo sapiens*	2	563	5e-94
CS474	AT006943	Mago-nashi (Drosophila) homologue		1	346	9e-36
CS459	AT006959	Metaxin 2 (mitocondial membrane)	*Mus musculus*	1	583	6e-04
CS26	AT006761	Mitochondrial ribosomal protein L2	*Drosophila melanogaster*	1	624	3e-21
CS507	AT006960	Preprocathepsin c	*Schistosoma japonicum*	1	610	4e-34
CS56	AT006775	Ribosomal protein l11 60s	*Caenorhabditis elegans*	1	596	2e-72
CS195	AT006847	Prolyl endopeptidase	*Sus scrofa*	1	604	7e-55
CS133	AT006734	Polyubiquitin	*Suberites domuncula*	2	528	3e-72
CS212	AT006739	Ribosomal protein l3 60s	*Toxocara canis*	1	598	1e-69
CS466	AT006961	Ribosomal protein l34 60s	*Aedes albopictus*	1	443	2e-37
CS315	AT006909	Ribosomal protein l36 60s	*Homo sapiens*	1	315	3e-20
CS442	AT006962	Ribosomal protein s1	*Xenopus laevis*	1	523	4e-69
CS479	AT006963	Ribosomal protein s10 40s	*Ictalurus punctatus*	1	551	1e-24
CS36	AT006768	Ribosomal protein s15-b 40s	*Schizosaccharomyces pombe*	1	466	1e-36
CS105	AF527457	Ribosomal protein s24	*Mus musculus*	2	469	2e-33
CS483	AT006964	Ribosomal protein s4	*Gallus gallus*	1	593	3e-78
CS18	AT006757	Ribosomal protein s4 40s	Chicken	1	602	2e-85
CS184	AT006735	Ribosomal protein s8 40s	*Homo sapiens*	2	603	2e-58
CS206	AT006854	Sec23 gene product [alt 1]	*Drosophila melanogaster*	1	648	2e-80
CS62	AT006778	Sec24D protein-like protein		1	641	3e-36
CS455	AT006965	Senescence-associated protein	*Pisum sativum*	1	561	1e-46
CS443	AT006966	Thrombospondin	*Gallus gallus*	2	602	2e-13
CS374	AT006967	Transaldolase 1	*Homo sapiens*	1	439	2e-44
CS80	AT006785	Translation initiation factor 2, subunit 2 (beta)		1	642	1e-12
CS31	AT006764	Transpanin-like protein	*Caenorhabditis elegans*	1	602	2e-08
CS517	AT006968	Ubiquitin		1	513	5e-74
CS404	AT006969	Ubiquitin-conjugating enzyme E2 variant 2	*Homo sapiens*	1	573	3e-43
CS391	AT006970	Acute morphine-dependence related protein 2		1	458	1e-54
CS450	AF051318	Glutathione-S transferase 28 kDa	*Clonorchis sinensis*	7	574	1e-37
CS430	AT006971	Mitochondrial thioredoxin	*Bos taurus*	1	550	3e-30
CS360	AT006972	Proteasome (prosome, macropain)	*Homo sapiens*	1	540	6e-44
CS369	AT006973	Ubiquitin-conjugating enzyme		1	433	1e-59
CS78	AT006784	Myoglobin	*Paramphistomum Epiclitum*	5	568	4e-28
CS243	AT006869	B7	*Homo sapiens*	1	649	8e-26
Metabolism (38)						
CS501	AT006914	4-Hydroxybutyrate CoA-transferase	*Clostridium aminobutyricum*	1	594	3e-40
CS482	AT006915	Aldehyde dehydrogenase (NAD+)	*Enchytraeus buchholzi*	1	572	1e-52
CS250	AT006872	Alpha-mannosidase c1orf22	*Homo sapiens*	1	645	2e-07
CS50	AT006774	Alpha-propionyl-coa carboxylase	*Rattus norvegicus*	1	506	2e-27
CS139	AT006809	Carbonic anhydrase II	Bovine	2	656	1e-39
CS519	AT006916	Citrate synthase (CoA, Citrate) Complex	Pig	1	515	8e-64

Table 2 (Continued)

EST	Accession no.	Putative identification	Species	No. of clones	Length of EST[b] (bp)	e-value
CS163	AT006823	Coproporphyrinogen III oxidase	*Drosophila melanogaster*	1	628	7e-82
CS539	AT006917	Cu/Zn-superoxide dismutase	*Fasciola hepatica*	1	551	8e-59
CS417	AT006918	Cytochrome b5	*Oryctolagus cuniculus*	1	582	2e-26
CS307	AT006903	Cytochrome c	Desert locust	1	565	8e-47
CS368	AT006974	Cytochrome c oxidase subunit 2	*Fasciola hepatica*	2	527	1e-28
CS312	AT006906	Cytochrome c oxidase subunit i	*Oryctolagus cuniculus*	2	533	2e-42
CS194	AT006846	Dihydrolipoamide s-acetyltransferase	*Homo sapiens*	1	605	3e-28
CS306	AT006902	Dihydrolipoamide: NAD+ oxidoreductase	*Canis familiaris*	1	635	3e-50
CS213	AF480465	Enolase	*Ricinus communis*	2	656	1e-68
CS92	AF527454	Fatty acid-binding protein	*Schistosoma japonicum*	1	578	1e-30
CS77	AT006783	Formiminoglutamate hydrolase	*Streptococcus pyogenes*	1	647	7e-09
CS500	AT006919	Fructose-1,6-bisphosphate aldolase	*Echinococcus multilocularis*	1	454	1e-41
CS167	AT006737	Glyceraldehyde 3-phosphate dehydrogenase	*Chondrus crispus*	1	625	6e-79
CS131	AT006806	Glyceraldehyde-3-phosphate dehydrogenase	*Schistosoma japonicum*	1	644	2e-23
CS415	AT006920	Glycerol kinase	*Mesorhizobium loti*	1	618	5e-14
CS74	AT006781	Liver glycogen phosphorylase	*Rattus norvegicus*	1	660	3e-50
CS252	AT006873	Malate dehydrogenase	*Drosophila melanogaster*	2	646	7e-72
CS376	AT006921	NADH dehydrogenase subunit 4	*Paragonimus westermani*	1	240	6e-21
CS15	AT006754	NADH dehydrogenase subunit 6		1	100	1e-08
CS271	AT006731	Phosphoglycerate mutase	*Schistosoma japonicum*	1	567	3e-80
CS49	AT006773	Phosphoglycerate mutase (gpma)	Lyme disease *spirochete*	1	453	2e-32
CS502	AT006922	Dehydrogenase kinase 4	*Mus musculus*	1	654	5e-29
CS14	AT006753	Rieske Fe-S protein precursor	*Rattus norvegicus*	1	643	1e-18
CS527	AT006923	Succinyl-coa synthetase beta-subunit	*Sus scrofa*	1	537	7e-35
CS101	AT006790	Tyrosinase	*Pelodiscus sinensis japonicus*	1	640	3e-21
CS339	AT006924	Uridine phosphorylase	*Homo sapiens*	2	627	5e-44
Antigen protein/ Others (10)						
CS276	AF136608	Antigen cs44	*Clonorchis sinensis*	3	515	4e-27
CS42	AT006769	Autoantigen	*Rhipicephalus appendiculatus*	1	462	3e-07
CS153	AT006817	Ly6-C antigen	*Rattus norvegicus*	2	594	2e-06
CS239	AF343876	Proline-rich antigen	*Clonorchis sinensis*	1	444	9e-34
CS231	AT006742	Sj-Ts4 protein	*Schistosoma japonicum*	1	580	1e-18
CS39	AT006727	Sperm-associated antigen 6	*Mus musculus*	1	467	2e-54
CS9	AT006744	Tegumental antigen Sm20.8	*Schistosoma mansoni*	1	681	2e-07

[a]Matches were sort into functional categories and only one representative match is given. Length ESTs are in nucleotides of reading sequences

antisera (Mohamed et al. 1998). The most significant feature of the flatworm tegument is that the ultimate boundary between the parasite and host is a living plasma membrane and its associated polyanionic coating or glycocalyx. This knowledge has revolutionized our understanding of symbiotic relationships, in that it raises a new conceptual level of the significance of the host-parasite relationship. In vivo, the alimentary tracts in schistosomes and the liver fluke play a role primarily in macromolecular digestion and the subsequent absorption of soluble digestive products. However, it is likely that these are augmented by the host-derived sugars and amino acids absorbed by the tegument (Halton et al. 1997). This can also be a target for developing diagnostics with recombinant antigens.

The EST analysis using the various stages of the parasite life cycle, such as the metacercariae, which is the infective stage to the final host, will be useful for investigating gene expression and characterizing the purified proteins of interest in *C. sinensis*. A particular gene or a class of genes identified within the EST dataset can be used as an appropriate target for diagnostics, vaccines or drug research and for further study of the development of clonorchiasis.

Acknowledgement This study was supported by the Brain Korea 21 Project for Medical Science, Yonsei University.

References

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O, et al. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377:3–174

Bourdais J, Biondi R, Sarfati S, Guerreiro C, Lascu I, Janin J, Veron M (1996) Cellular phosphorylation of anti-HIV nucleosides. Role of nucleoside diphosphate kinase. J Biol Chem 271:7887–7890

Chapman RW (1999) Risk factors for biliary tract carcinogenesis. Ann Oncol [Suppl] 10:308–311

Crompton DWT (1999) How much human helminthiasis is there in the world? J Parasitol 85:397–403

Cronan JE Jr, Godson GN (1972) Mutants of *Escherichia coli* with temperature-sensitive lesions in membrane phospholipid synthesis: genetic analysis of glycerol-3-phosphate acyltransferase mutants. Mol Gen Genet 116:199–210

Estuningsih SE, Smooker PM, Wiedosari E, Widjajanti S, Vaiano S, Partoutomo S, Spithill TW (1997) Evalution of antigens of *Fasciola gigantica* as vaccines against tropical fasciolosis in cattle. Int J Parasitol 27:1419–1428

Franco GR, Valadado AF, Azevedo V, Rabelo EM (2000) The *Schistosoma* gene discovery program: state of the art. Int J Parasitol 30:453–463

Halton DW (1997) Nutritional adaptations to parasitism within the platyhelminthes. Int J Parasitol 27:693–704

Han SS, Hahn HJ, Seo BS (1961) The uptake of ^{14}C-glucose by *Clonorchis sinensis*. Korean J Int Med 4:281–285

Hillyer GV, Rosa MI, Alicea H, Hemandez A (1987) Acquired resistance to *Fasciola hepatica* in cattle using a purified adult worm antigen. Am J Trop Med Hyg 37:363–369

Hillyer GV, De Galanes M, Fosa MI, Montealegre F (1988) Acquired immunity in schistosomiasis with purified *Fasciola hepatica* cross-reactive antigens. Vet Parasitol 29:265–280

Hong SJ, Seong KY, Sohn WM, Song KY (2000) Molecular cloning and immunological characterization of phosphoglycerate kinase from *Clonorchis sinensis*. Mol Biochem Parasitol 108:207–216

Janssen E, Dzeja PP, Oerlemans F, Simonetti AW, Heerschap A, de Haan A, Rush PS, Terjung RR, Wieringa B, Terzic A (2001) Adenylate kinase 1 gene deletion disrupts muscle energetic economy despite metabolic rearrangement. EMBO J 19:6371–6381

Kang IK, Lee SH, Seo BS (1969) Study on the ^{14}C-glucose metabolism by *Clonorchis sinensis*. Paper chromatographic analyses in combination with autoradiography. Korean J Parasitol 7:43–52

Kim TY, Kang SY, Ahn IY, Cho SY, Hong SJ (2001) Molecular cloning and characterization of an antigenic protein with a repeating region from *Clonorchis sinensis*. Korean J Parasitol 39:57–66

Kim YI, Rim HJ, Park UB (1993) Effect of *Clonorchis sinensis* infection and dimethylnitrosamine administration on the induction of cholangiocarcinoma in Syrian golden hamster. Korean J Pathol 31:21–30

Konrad M (1993) Molecular analysis of the essential gene for adenylate kinase from the fission yeast *Schizosaccharomyces pombe*. J Biol Chem 268:11326–11334

Mohamed MM, Shalaby KA, LoVerde PT, Karim AM (1998) Characterization of Sm20.8, a member of a family of schistosome tegumental antigens. Mol Biochem Parasitol 96:15–25

Park SY, Lee KH, Hwang YB, Kim KY, Park SK, Hwang HA, Sakanari JA, Hong KM, Kim SI, Park H (2001) Characterization and large-scale expression of the recombinant cysteine proteinase from adult *Clonorchis sinensis*. J Parasitol 87:1454–1458

Paul J (2000) Brindley. Parasite genomes. Int J Parasitol 30:327

Pucar D, Janssen E, Dzeja PP, Juranic N, Macura S, Wieringa B, Terzic A (2000) Compromised energetics in the adenylate kinase AK1 gene knockout heart under metabolic stress. J Biol Chem 275:41424–41429

Qualtieri A, Pedace V, Bisconte MG, Bria M, Gulino B, Andreoli V, Brancati C (1997) Severe erythrocyte adenylate kinase deficiency due to homozygous A $\rightarrow$ G substitution at codon 164 of human AK1 gene associated with chronic haemolytic anaemia. Br J Haematol 99:770–776

Rim HJ (1998) Field investigations on epidemiology and control of fish-borne parasites in Korea. Int J Food Sci Technol 33:157–168

Schneider B, Xu YW, Sellam O, Sarfati R, Janin J, Veron M, Deville-Bonne D (1998) Pre-steady state of reaction of nucleoside diphosphate kinase with anti-HIV nucleotides. J Biol Chem 273:11491–11497

Selkirk ME, Smith VP, Thomas GR, Gounaris K (1998) Resistance of filarial nematode parasites to oxidative stress. Int J Parasitol 28:1315–1332

Smooker PM, Steeper KR, Drew DR, Strugnell RA, Spithill TW (1999) Humoral responses in mice following vaccination with DNA encoding glutathione S-transferase of *Fasciola hepatica*: effects of mode of vaccination and the cellular compartment of antigen expression. Parasite Immunol 21:357–364

Song CY, Rege AA (1991) Cysteine proteinase activity in various developmental stages of *Clonorchis sinensis*: a comparative analysis. Comp Biochem Physiol B 99:137–144

Tawe W, Hashmi S, Lustigman S (2001) Identification of filarial vaccine and drug target candidates by EST analysis. Trends Parasitol 17:204–206

Tendler M, Brito CA, Vilar MM, Serra-Preire N, Diogo CM, Almeida MS, Delbem AC, Da Silva JP, Savino W, Garratt RC, Katz N, Simpson AS (1996) *Schistosoma mansoni* fatty acid-binding protein, Sm 14, is the potential basis of a dual-purpose anti-helminth vaccine. Proc Natl Acad Sci USAA93:269–273

Yang HJ, Park SJ, Im KI, Yong TS (2000) Identification of a *Clonorchis sinensis* gene encoding a vitellaria antigenic protein containing repetitive sequences. Mol Biochem Parasitol 111:213–216

Yenbutr P, Scott AL (1995) Molecular cloning of a serine proteinase inhibitor from *Brugia malayi*. Infect Immun 63:1745–1753

Zang X, Yazdanbakhsh M, Jiang H, Kanost MR, Maizels RM (1999) A novel serpin expressed by blood-borne microfilariae of the parasitic nematode *Brugia malayi* inhibits human neutrophil serine proteinases. Blood 94:1418–1428