Korean J Parasitol. Vol. 47, No. 2: 103-107, June 2009 DOI: 10.3347/kjp.2009.47.2.103

Construction of EST Database for Comparative Gene Studies of *Acanthamoeba*

Eun-Kyung Moon[1], Joung-Ok Kim[1], Ying-Hua Xuan[1], Young-Sun Yun[1], Se Won Kang[2], Yong Seok Lee[2], Tae-In Ahn[3], Yeon-Chul Hong[1], Dong-Il Chung[1] and Hyun-Hee Kong[1,*]

[1]Department of Parasitology, Kyungpook National University School of Medicine, Daegu 700-422, Korea; [2]Department of Parasitology and Malariology, PICR, College of Medicine and Frontier Inje Research for Science and Technology, Inje University, Busan 614-735, Korea; [3]Department of Biological Science, Seoul National University, Seoul 151-742, Korea

Abstract: The genus *Acanthamoeba* can cause severe infections such as granulomatous amebic encephalitis and amebic keratitis in humans. However, little genomic information of *Acanthamoeba* has been reported. Here, we constructed *Acanthamoeba* expressed sequence tags (EST) database (*Acanthamoeba* EST DB) derived from our 4 kinds of *Acanthamoeba* cDNA library. The *Acanthamoeba* EST DB contains 3,897 EST generated from amebae under various conditions of long term in vitro culture, mouse brain passage, or encystation, and downloaded data of *Acanthamoeba* from National Center for Biotechnology Information (NCBI) and Taxonomically Broad EST Database (TBestDB). The almost reported cDNA/genomic sequences of *Acanthamoeba* provide stand alone BLAST system with nucleotide (BLAST NT) and amino acid (BLAST AA) sequence database. In BLAST results, each gene links for the significant information including sequence data, gene orthology annotations, relevant references, and a BlastX result. This is the first attempt for construction of *Acanthamoeba* database with genes expressed in diverse conditions. These data were integrated into a database (http://www.amoeba.or.kr).

Key words: *Acanthamoeba*, ESTs, database, brain passage, encystation

INTRODUCTION

Free-living amebae belonging to the genus *Acanthamoeba* are the causative agents of granulomatous amebic encephalitis (GAE), a fatal disease of the central nervous system (CNS), and amebic keratitis (AK) [1]. The recent increased incidence in *Acanthamoeba* infections is due in part to infection in patients with acquired immune deficiency syndrome, while that for keratitis is due to the increased use of contact lenses [2]. In addition to these medical importances, *Acanthamoeba* is also well known as a good model system to study eukaryotic cell biology due to its relatively large size, rapid growth in culture, active motility, and well developed cytoskeleton [3,4]. Over the years, *Acanthamoeba* has gained increasing attention from the scientific community with these diverse roles [4].

Over the past decade, as the development of tools for genome study, knowledge on genome of protozoan parasites has grown exponentially. Based on these results of genome studies, constructions of various databases have been applied including parasitic protozoa such as *Plasmodium* species [5,6], *Entamoeba histolytica* [7], *Trypanosoma cruzi* [8,9], and free living protozoa such as *Dictyostelium discoideum* [10,11]. Although *Acanthamoeba* has been considered to be an important organism in medicine and biological researches, little genomic information of *Acanthamoeba* has been reported. The genome size of the ameba has been speculated as ~1 × 10^8 bp [3]. The complete primary sequence of *A. castellanii* mitochondrial genome was determined as 41,591 bp [12], and the small-sized expressed sequence tag (EST) analysis of *Acanthamoeba healyi* was reported [13]. Recently, gene discovery in *A. castellanii* was performed [14] and a taxonomically broad database (TBestDB) from 49 organisms including 13,814 ESTs of *A. castellanii* was constructed [15]. TBestDB database (http://tbestdb.bcm.umontreal.ca) containing ~370,000 clustered EST sequences of 49 organisms provided information of 5,262 clustered EST sequences in *A. castellanii* trophozoites [15]. However, these reported genes seem to be expressed in normal conditions or some genes silenced. The virulence of *Acanthamoeba* can be attenuated by a long-term in vitro cultivation and the cyst form of *Acanthamoeba* is resistant to immune responses and antibiotics. With these databases, it is difficult to get the information about enhanced virulence genes or encystation mediat-

• Received 21 January 2009, revised 20 March 2009, accepted 31 March 2009.
* Corresponding author (hhkong@mail.knu.ac.kr)

ing genes.

In this study, we constructed the specific database with our previously reported EST sequences generated with *Acanthamoeba* in a highly virulent condition by mouse brain passage or in encystation. This new database of *Acanthamoeba* could give more information of various genes concerned with pathogenesis or encystation of the cyst forming protozoa.

MATERIALS AND METHODS

Our previously reported EST sequences, randomly selected from 4 kinds of cDNA library [16, processing], were used to construct database to study various types of genes containing pathogenicity, differentiation, or stress-condition related genes of *Acanthamoeba*.

The BLAST server for the *Acanthamoeba* EST database (*Acanthamoeba* EST DB) was constructed on the basis of the dual Xeon CPU system. After installing the Cent operating system, NCBI www blast package was installed after web server configuration for cgi (common gate interface) (http://www.amoeba.or.kr). To build up the stand alone blast server, it was conducted as follows: first, own EST sequence data for *Acanthamoeba* and downloaded nucleotide and amino acid sequences related with *Acanthamoeba* available at NCBI and TBestDB were used [15]. Second, it was translated into the multifasta format that was stored as database by using the formatdb program provided by NCBI. Third, blast results of own EST sequences which were transformed into a table include QueryID (clone name), SubjectID (gi number of NCBI), KOG (Clusters of Orthologous Groups of proteins), QLen (query sequence length), CovQ (coverage of query sequence against subject sequence), SLen (subject sequence length), CovS (coverage of subject sequence against query sequence), Pid (percent identity in the HSP), Psi (percent similarity in the HSP), Frame, E-value, a kind of Database, Annotation results, Source (species), and Link service for original sequence and blast results.

Table 1. Statistics of ESTs of *Acanthamoeba* species

EST category	No. of clones				
	A. castellanii		*A. healyi*		Total
	Trophozoites	Cysts	Old[a]	MBP[b]	
Total clones sequenced	1,000	1,115	1,000	1,050	4,165
ESTs submitted for BLAST search	905	1,021	938	1,033	3,897
ESTs identified by homology	632	677	767	722	2,798
Unique ESTs identified	348	648	718	833	2,547
Cluster	179	129	101	94	503
Singlet	169	519	617	739	2,044
ESTs with homology to *Acanthamoeba* genes	11	15	26	17	69

[a]long-term in vitro cultivated *A. healyi*; [b]3 times mouse-brain passaged *A. healyi*.

RESULTS

Composition of *Acanthamoeba* EST DB

Based on our previous report [16], specific *Acanthamoeba* EST database (*Acanthamoeba* EST DB) was constructed (http://www.amoeba.or.kr: inaccessible). The sequence data of *Acanthamoeba* EST DB consisted of 3,897 ESTs data of *Acanthamoeba* from our previous studies (Table 1), 33,648 sequence data related with *Acanthamoeba* from NCBI, and 5,260 nucleotide data of *Acanthamoeba* from TBestDB. Total 42,805 sequence data were used for construction of the database (Tables 1, 2).

Information of *Acanthamoeba* EST DB

The contents of *Acanthamoeba* EST DB consisted of 3 search

Table 2. *Acanthamoeba* sequences used for EST database server

Database type	Database name	Type	Data No.
Generated	*Acanthamoeba castellanii* trophozoites	Nucloetide	905
Generated	*Acanthamoeba castellanii* cysts	Nucleotide	1,021
Generated	*Acanthamoeba healyi* Old	Nucleotide	938
Generated	*Acanthamoeba healyi* MBP	Nucleotide	1,033
Downloaded	NCBI Acanthamoebidae	Nucleotide	33,362
Downloaded	NCBI Acanthamoebidae mitochondria	Nucleotide	1
Downloaded	NCBI Acanthamoebidae	Amino acid	285
Downloaded	TBestDB *Acanthamoeba castellanii* trophozoites	Nucleotide	5,260
Total			42,805

EST, expressed sequence tags; MBP, mouse brain passed; NCBI, National Center for Biotechnology Information; TBestDB, taxonomically broad database.

Table 3. Organization of the database

Menu	Contents
Home	Go to start page
BLAST NT	Blastn, tblastn, tblastx
BLAST AA	Blastx, blastp
BLAST results	Interface for analysed data of EST
2-Sequence	Blast 2 sequences
Statistics	Statistic analysed data

Table 4. Comparison of redundancy between TBestDB and *Acanthamoeba* EST database

Sequence category	No. of cDNA clones	
	TBestDB	*Acanthamoeba* EST DB
Total sequences	13,770	3,897
Unique ESTs identified	5,260 (38.2%)	2,327 (59.7%)
Annotated	4,888 (92.9%)	1,623 (69.7%)
Not-annotated	372 (7.1%)	704 (30.3%)

tools and 2 depots of search results data. BLAST system with nucleotide (BLAST NT), BLAST system with amino acid (BLAST AA), and 2-Sequence were developed as the search tools, and BLAST results and statistics were the depots, respectively (Table 3). BLAST NT or BLAST AA contained nucleotide database or protein sequence database, respectively, which could provide predictive information for the functions of *Acanthamoeba* genes or proteins in any experiments through comparative analysis. The search of BLAST NT worked with blastn, tblastn, or tblastx program, while blastp or blastx program was used for the search of BLAST AA. 2-Sequence was an alignment tool to compare the homology and similarity between 2 genes using the blastn, tblastn, tblastx, blastp, or blastx program. Each searched sequences linked to information of annotated genes and showed the similarity with queried sequences. BLAST results could not only store the results of analysis but also could provide significant information, including the sequence data, blastX results, orthology annotations, KOG analysis, and relevant references for each gene. In statistics, the results of *Acanthamoeba* ESTs analysis were summarized. Each program or database in the search tool was optionally selected and comparative analysis of *Acanthamoeba* genes was also applicable for various investigations.

Specificity of *Acanthamoeba* EST DB

To show the specificity of our database, we compared the redundancy rates between TBestDB and our *Acanthamoeba* EST DB (Table 4). TBestDB database (http://tbestdb.bcm.umontreal.ca) provided 5,262 clustered EST sequences in *A. castellanii* trophozoites. Although total EST sequences of *Acanthamoeba* EST DB (3,897 ESTs) was smaller than that of TBestDB (13,770 ESTs), redundancy was relatively lower than that of TBestDB. Unique cluster EST of *Acanthamoeba* EST database (2,327 clones, 59.7%) was higher than that of TBestDB (5,260 clones, 38.2%). Among unique ESTs clusters, the not-annotated cluster ESTs including unknown genes, hypothetical or novel proteins of *Acanthamoeba* EST DB (704 clones, 30.3%) were also higher than TBestDB (372 clones, 7.1%) (Table 4).

Table 5. Statistics on searched proteins form *Acanthamoeba* including PA (protease-associated) domain (E-value $\leq e^{-05}$)

Menu	No. of clones
TBestDB	11
NCBI	21
Acanthamoeba EST database	49

Our *Acanthamoeba* EST DB included various genes concerned with enhanced virulence or different developmental stages of *Acanthamoeba*. To confirm the specificity of our database, we examined the blast results of *Acanthamoeba* EST DB. With the amino acid sequences of the protease-associated (PA) domain from *Acanthamoeba lugdunensis* (ABY6399), PA domain containing proteins were identified using the tblastn program in BLAST NT search tool (Table 5). Our database provided more various informations for the PA domain containing proteins than TBestDB or NCBI blast search results.

DISCUSSION

As the strategies and techniques for molecular biology are developed and advanced rapidly, the database of nucleotide sequences and genome become a very powerful tool to identify new genes and proteins and to suspect the function of novel genes. Over the past decade, together with genome studies, construction of database has been applied to many organisms including parasitic protozoa [5,7-9]. *Entamoeba histolytica* genome analysis was carried out on a 12.5-fold coverage of the total genome [7], but that of *A. castellanii* was carried out on a 0.5-fold coverage of the total genome [14].

Several reasons would explain the poor progress in *Acanthamoeba* genomic study. First, the gene structure of *Acanthamoeba* may be more complex than expected. In a previous genome study of *Acanthamoeba*, average 3.0 exons per gene were calculated and this was higher than those of *E. histolytica* which has

1.3 exons per gene [14]. Ploidy and chromosome numbers of the genus *Acanthamoeba* are still undiscovered. Second, the transfection system very useful to study functions or localization of a putative gene has not been completely established in *Acanthamoeba* yet. Kong and Pollard [17] recently developed the system which is for the transient transfection in *Acanthamoeba*. Peng [18] reported the system for the stable transfection of *Acanthamoeba castellanii*. However, these systems have to overcome the low transfection efficiency to be used commonly [17,18]. Third, little data on *Acanthamoeba* genes and proteins in public database makes more difficulty to identification and speculation of functions of new genes or proteins. When we search for a new gene or a protein in the NCBI blast, the result usually shows the matched genes or proteins of vertebrates. Thus, genes of *Acanthamoeba* may be shown at a lower part of the list or may not be shown because of a low HSP (high scoring segment pair). This reveals the requirement of more information in genomes of *Acanthamoeba*. For the proteomic researches, more genomic information of *Acanthamoeba* is also needed for a comparative genetic study.

In the present study, the specific database of *Acanthamoeba* named *Acanthamoeba* EST database (*Acanthamoeba* EST DB) was constructed. To promote the *Acanthamoeba* gene study, *Acanthamoeba* EST DB could provide the specific sequences concerned with specific conditions such as mouse brain passage or encystation. TBestDB showed the information of 13,814 ESTs from *Acanthamoeba* generated with trophozoites; however, in our database, 3,897 ESTs were generated with diverse conditions. Although the size of *Acanthamoeba* EST database was smaller than that of TBestDB, the redundancy of information was lower than TBestDB, and the number of non-annotated clusters, unknown, hypothetical, or novel protein was much higher than TBestDB. It means that *Acanthamoeba* EST DB may contain more diverse genes related with *Acanthamoeba* life- or infection cycle. Investigation of those unknown or novel proteins, which are expressed specifically in encystation or mouse infection, will provide the clues to understand the pathogenesis and encystation of *Acanthamoeba*.

This is the first attempt of specific database for comparative studies of *Acanthamoeba*. In fact, the entire genome of this organism has not been fully sequenced yet. Therefore, the number of ESTs should be increased to improve the usefulness of database for comparative genome studies. This database will be upgraded with new sequences which are related with cyst mediating genes. *Acanthamoeba* EST DB would make easy the gene study of *Acanthamoeba*, providing sequence data for proteomics and providing many new opportunities for the scientific community. *Acanthamoeba* EST DB can be freely accessible via http://www.amoeba.or.kr.

ACKNOWLEDGEMENTS

This work was supported by No. R01-2006-000-10757-0 from the Basic Research Program of the Korea Science & Engineering Foundation (KOSEF) and the Brain Korea 21 Project in 2008. We thank to a KOSEF program (System development for application of genomic sequence information) No. M107520000001-07N5200-00110 funded by the Korea Government (MEST).

REFERENCES

1. Marciano-Cabral F, Cabral G. *Acanthamoeba* spp. as agents of disease in humans. Clin Microbiol Rev 2003; 16: 273-307.
2. Marciano-Cabral F, Puffenbarger R, Cabral GA. The increasing importance of *Acanthamoeba* infections. J Eukaryot Microbiol 2000; 47: 29-36.
3. Byers TJ, Hugo ER, Stewart VJ. Genes of *Acanthamoeba*: DNA, RNA and protein sequences (a review). J Protozool 1990; 37: 17S-25S.
4. Khan NA. *Acanthamoeba*: biology and increasing importance in human health. FEMS Microbiol Rev 2006; 30: 564-595.
5. Watanabe J, Suzuki Y, Sasaki M, Sugano S. Full-malaria 2004: an enlarged database for comparative studies of full-length cDNAs of malaria parasites, *Plasmodium* species. Nucleic Acids Res 2004; 32: D334-338.
6. Watanabe J, Wakaguri H, Sasaki M, Suzuki Y, Sugano S. Comparasite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs. Nucleic Acids Res 2007; 35: D431-438.
7. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, Nozaki T, Suh B, Pop M, Duchene M, Ackers J, Tannich E, Leippe M, Hofer M, Bruchhaus I, Willhoeft U, Bhattacharya A, Chillingworth T, Churcher C, Hance Z, Harris B, Harris D, Jagels K, Moule S, Mungall K, Ormond D, Squares R, Whitehead S, Quail MA, Rabbinowitsch E, Norbertczak H, Price C, Wang Z, Guillen N, Gilchrist C, Stroup SE, Bhattacharya S, Lohia A, Foster PG, Sicheritz-Ponten T, Weber C, Singh U, Mukherjee C, El-Sayed NM, Petri WA Jr, Clark CG, Embley TM, Barrell B, Fraser CM, Hall N. The genome of the protist parasite *Entamoeba histolytica*. Nature 2005; 433: 865-868.
8. Aguero F, Zheng W, Weatherly DB, Mendes P, Kissinger JC. Tcruzi DB: an integrated, post-genomics community resource for *Trypanosoma cruzi*. Nucleic Acids Res 2006; 34: 428-431.
9. Luchtan M, Warade C, Weatherly DB, Degrave WM, Tarleton RL, Kissinger JC. TcruziDB: an integrated *Trypanosoma cruzi* genome resource. Nucleic Acids Res 2004; 32: 344-346.

10. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A. The genome of the social amoeba *Dictyostelium discoideum*. Nature 2005; 435: 43-57.
11. El-Sayed NM, Alarcon CM, Beck JC, Sheffield VC, Donelson JE. cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. Mol Biochem Parasitol 1995; 73: 75-90.
12. Burger G, Plante I, Lonergan KM, Gray MW. The mitochondrial DNA of the amoeboid protozoon, *Acanthamoeba castellanii*: complete sequence, gene content and genome organization. J Mol Biol 1995; 245: 522-537.
13. Kong HH, Hwang MY, Kim HK, Chung DI. Expressed sequence tags (ESTs) analysis of *Acanthamoeba healyi*. Korean J Parasitol 2001; 39: 151-160.
14. Anderson IJ, Watkins RF, Samuelson J, Spencer DF, Majoros WH, Gray MW, Loftus BJ. Gene discovery in the *Acanthamoeba castellanii* genome. Protist 2005; 156: 203-214.
15. O'Brien EA, Koski LB, Zhag Y, Yang L, Wang E, Gray MW, Burger G, Lang BF. TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). Nucleic Acids Res 2007; 35: 445-451.
16. Moon EK, Chung DI, Hong YC, Ahn TI, Kong HH. *Acanthamoeba castellanii*: gene profile of encystation by ESTs analysis and KOG assignment. Exp Parasitol 2008; 119:111-116.
17. Kong HH, Pollard TD. Intracellular localization and dynamics of myosin-II and myosin-IC in live *Acanthamoeba* by transient transfection of EGFP fusion proteins. J Cell Sci 2002; 115: 4993-5002.
18. Peng Z, Omaruddin R, Bateman E. Stable transfection of *Acanthamoeba castellanii*. Biochim Biophys Acta 2005; 1743: 93-100.