The Journal of Animal & Plant Sciences, 25(1): 2015, Page: 139-145
ISSN: 1018-7081

PRIMARY ANALYSIS OF THE EXPRESSED SEQUENCE TAGS FROM A FULL-LENGTH ENRICHED cDNA LIBRARY OF SIBERIAN TIGER (*PANTHERA TIGRIS ALTAICA*)

Y. Guo[1*], C.Q. Liu[1*], L.S. Xu[1], T.F. Lu[2], D. Liu[3], X.C. Li[2], J.Z. Ma[4], Y.H. Ma[2], and W.J. Guan[2]

[1]Bioscience Department, Department of Biochemistry and Molecular Biology, Bengbu Medical College, Bengbu, 233030, China
[2]Institute of Animal Science, Chinese Academy of Agricultural Science, Beijing, 100194, China
3 The Northeast Tiger Wooden Land of Heilongjiang, Harbin, 150028, China
4 College of Wildlife Resources, Northeast Forestry University, Harbin, 150028, China
[*] Equal contribution to the paper. Both may be considered as the first author.
Corresponding author E-mail address: wjguan86@iascaas.net.cn

ABSTRACT

Tiger is a special species only found in Asia and considered as one of the world's most endangered species. In this study, a full-length enriched cDNA library was successfully constructed from Siberian tiger. The titers of primary and amplified libraries were 1.28×10^6 pfu/ml and 1.59×10^{10} pfu/ml respectively. The proportion of recombinants from unamplified library was 91.3% and average length of exogenous inserts was 1.06 kb. A total of 365 individual ESTs with sizes ranging from 264 to 1,378bps were then analyzed. Furthermore, 305 unigenes were successfully annotated and involved in 47 functions of the GO classification, Cell (278, 91.1%), Cellular process (302, 82.6%) and Binding (250, 82.0%) are the dominant terms. A total of 234 unigenes were assigned to 244 KEGG pathways, and the pathways with most representation are metabolic pathways (40, 17.1%). Among the 25 COG categories, general function prediction only cluster (52, 14.2%) represents the largest group, followed by translation, ribosomal structure and biogenesis (41, 11.2%), replication, recombination and repair (30, 8.2%), and only 9.9% ESTs (36) classified as novel genes. Moreover, this library provided a useful platform for the functional genome and transcriptome research for the *P. tigris* and other felid animals in future.

Key words: Siberian tiger; Fibroblast cell; SMART cDNA library; Expressed sequence tags.

INTRODUCTION

The tiger (*Panthera tigris* Linnaeus, 1758), the largest felid species and a widely recognized symbol of wildlife conservation, is one of the world's most endangered species (Cho *et al*., 2013). Historically tigers inhabited much of Asia, including the regions between the Caspian and Aral Seas, southeastern Russia, and the Sunda islands, forming the nine genetically validated subspecies (Luo *et al*., 2004). Four of these went extinct in the wild during the last century (Javan, Balinese, South China and Caspian tigers), leaving five extant subspecies (Amur, Bengal, Indochinese, Malayan and Sumatran tigers). The Siberian tiger (*P. t. altaica*), also known as Amur tiger originated in northeast Asia, now mainly distributed in the Russian far east, northeast of China and north of Korea. Today, approximately 500 wild Siberian tiger remained, while only 20 or so left in China, mainly in eastern mountain area of Heilongjiang and Jilin provinces (Liu *et al*., 2010). Siberian tiger is the largest subspecies of the extant species and the only subspecies inhabiting snow-covered regions. The tiger is warranted the highest level of protection by the Convention on International Trade in Endangered Species of Wild Fauna & Flora (CITES). In 1989, the Chinese government placed Siberian tiger into the highest National Protected Animal category.

Previous genetic studies using mitochondrial and a comparative analysis among Panthera whole genome sequences provided valuable information on genome organization, evolutionary divergence and overall endemic diversity (Cho *et al*., 2013). However, only few functional genes, mitochondrial sequences and 57 annotated ESTs of *P. tigris* have been cloned and partially studied according to the latest data of NCBI. To identify more genes of Siberian tiger, including the characterization of specific expressed, new or unknown functional genes and further study of their functions, construction of full-length cDNA libraries of Siberian tiger is an efficient method (Ying, 2004; Li *et al*., 2008). In this paper, we report the construction of a normalized full-length enriched cDNA library and primary analysis of 365 ESTs from female Siberian tiger fibroblast cells.

MATERIALS AND METHODS

Animal materials collection and Cell culture: Siberian tiger ear tissue samples (16 male and 18 female) were sampled from The Siberian Tiger Park of Heilongjiang. Ear tissue samples (about 1 cm^2 in size) were chopped

into 1 mm^3 pieces and added DMEM medium with 10% fetal bovine serum in a 37°C incubator with 5% CO_2. A fibroblast cell line with normal karyotype was obtained using primary explanting techniques and cell cryogenic preservation technology. The characteristic of cell line was tested for viability, microorganism detection and chromosome euploidy according to Liu *et al* (Liu *et al*., 2013).

cDNA library construction: Cells were harvested and total RNA was extracted using Trizol reagent (Invitrogen, Carlsbad, CA, USA) when they were in the period of passage 3. The quality of total RNA was estimated using the Agilent RNA 6000 nano Reagents Port 1. First and double-strand cDNAs were synthesized according to the protocol of the SMART cDNA Library Construction kit (Clontech, Palo Alto, CA, USA). Subsequently approximately 2μL of first strand cDNA sample was amplified using long distance PCR (LD-PCR). The first four peak fractions containing cDNA (>500bp) were pooled together using column chromatograph with CHROMA SPIN-400 medium. The cDNA was ligated to λTriplEx2 vector (1:2) and the ligation was packaged with Gigapack Ⅲ Gold Packaging extract.

Titration of the primary library: The titer of the primary library was estimated by the number of independent phage and independent clones in the library according to the formula: pfu/mL = number of plaques × dilution factor × 10^3μL/mL (μL of diluted phage plated). The different dilution of packaging extracts (1:5 to 1:20) for an unamplified λ lysate was mixed with 500μL of the *E.coli* XL1-Blue (OD_{600}=4.0). The recombination efficiency was identified by blue/white screening in *E.coli* XL1-Blue. Colony PCR was used to confirm the size of inserted fragments in the library by the primers (cDNA F: CCATTGTGTTGGTACCCGG; cDNA R: ATACGACTCACTATAGGGCGAATT). After amplification, the completed cDNA libraries were stored in 7% dimethyl sulfoxide (DMSO) at -80°C.

Sequencing and sequence analysis: cDNA clones were selected randomly from the cDNA library and sequenced on an ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). A large-scale EST sequencing project for Siberian tiger was initiated to identify and functionally annotate as many unique transcripts as possible. The processed cDNA sequences were used to perform the BLASTX search and annotation against NCBI non-redundant (nr) database and Swiss-Port to compare all available ESTs and genes to date. BLASTX results with bit scores greater than 80 and E-values of less than 10^{-10} were generally regarded as significant match (Wheeler *et al*., 2008; Thanh *et al*., 2011; Phukon *et al*., 2012). Functional annotation by gene ontology terms (GO) was analyzed by Blast2go and WEGO software (Gotz *et al*., 2008). The COG and KEGG pathway annotation were performed against Cluster of Orthologous Groups database and Kyoto Encyclopedia of Genes and Genomes database, respectively (Kanehisa *et al*., 2008; Sultan *et al*., 2008).

RESULTS

Cell culture and Characteristic Tests: We used a primary explanting technique and cell cryogenic preservation technology to establish the Siberian tiger fibroblast cell line and proceeded to biological and genetic detection. The culture conditions were optimal, and the cells were healthy (Fig. 1A–D). Because we wanted to construct a cDNA library to conserve genome character of Siberian tiger, the fibroblasts must maintain diploid character similar to *in vivo*. Chromosome analysis showed that the frequency of fibroblast cell chromosome number of 2n = 38 was 90.6-92.2% in passage 1 to 4, which indicated that the cell line was still primarily diploid (Fig. 1F). The test results of the bacteria, virus and mycoplasma screens were negative (Fig. 1E).

Total RNA extraction and LD-PCR: The ratio of OD_{260}/OD_{280} to the total RNA was approximately 2.04 and the concentration was 2.238μg/μl (Fig. 2A). As shown in Fig. 2B, two bright bands of 18S rRNA and 28S rRNA can be seen clearly, indicating that the total RNA is integrated and stable enough for cDNA library construction. Two micrograms of total RNA were subjected to reverse transcription for synthesis of the first and double-stranded cDNAs for LD-PCR, and the ds-DNA appeared as a smear of bands of 0.5 - 4 kb on the gel (Fig. 2B).

Characterization of cDNA library and Generation of expressed sequence tags: The cDNA size fractionation was carried out using CHROMA SPIN-400 column (Fig. 2C). The titers of primary and amplified libraries were 1.28×10^6 pfu/ml and 1.59×10^{10} pfu/ml respectively. The recombination efficiency of the amplified library was 91.3%. The insert ratio and the average length of inserted fragments were measured by PCR, as shown in Fig. 2C. The average size was approximately 1.06 kb in average, 1-2 kb in 62.3% and 0.5-1.0 kb in 37.7%. Four hundred and fifty-six white clones were picked randomly for EST sequencing. After removal of the vector sequences and low-quality sequences, 426 effective sequences were obtained. Moreover, a total of 365 individual ESTs ranged from 264 to 1,378 nucleotides in length were analyzed, and partly deposited in the GenBank under accession No. from KF896402 to KF896586, KJ635578 to KJ635637.

Gene Ontology and Pathway KEGG analysis: A total of 365 unigenes of Siberian tiger were assigned for GO analysis based on matches with sequences whose functions were known previously using BLASTX

(KEGG, Swissprot and TrEMBL). As shown in Fig. 3A, 305 unigenes were successfully annotated and involved in 47 functions, and classified into three major groups: cellular-components, biological-processes and molecular-functions. In each of the three main categories of the GO classification, Cell (278, 91.1%), Cellular process (302, 82.6%), Binding (250, 82.0%), organelle (219, 71.8%) and Metabolic process (190, 62.3%) are the dominant terms (Fig. 3A). To investigate further the expression characteristics, we also examined the gene ontology (GO) analysis using PANTHER expression analysis tool. Most unigenes are associated with transporter (PC00227, 17.60%), nucleic acid binding (PC00171, 15.70%), hydrolase (PC00121, 15.40%) and transferase (PC00220, 11.80%), and these occupied more than 60% of the total related function (Fig. 3B), which are processes typically found in the tissue function and metabolism.

To further explore the function of the annotated genes, pathway analysis was performed using KEGG. A total of 234 unigenes (64.1%) were assigned to 244 KEGG pathways, and the numbers of transcripts in different pathways ranged from 1 to 40. The top 25 pathways with EST numbers are shown in Fig. 4B, and the highest of the number of transcripts is involved in the metabolic pathways. The pathways with most representation by the unique sequences are Metabolic pathways (40 members, 17.1%), PI3K-Akt signaling pathway (28, 12.0%), Ras signaling pathway (19, 8.1%), and Focal adhesion (18 members, 7.7%). Nine of the top-25 pathways were mapped to the pathways associated with signaling pathways (indicated with red box). In addition, four of top-25 pathways were mapped to the pathways associated with cytoskeleton and cell junction (indicated with green box), such as Focal adhesion, Regulation of actin cytoskeleton.

COG analysis and Function classification: The distribution of ESTs from *P. t. altaica* cDNA library revealed that 149 (40.8%) of them were classified as strong matches to sequences in the non-redundant protein database (Nr) for the highest match with a *E*-value less than 10^{-25}, meanwhile 180 (49.3%) ESTs were nominal with *E*-value for the highest match between 10^{-10} and 10^{-25}, and 36 (9.9%) ESTs were weak with *E*-value for the highest match greater than 10^{-10} or no significant similarity to sequences in the database. Among the 25 COG categories, general function prediction only cluster (52, 14.2%) represents the largest group, followed by translation, ribosomal structure and biogenesis (41, 11.2%), replication, recombination and repair (30, 8.2%), and cell cycle control, cell division, chromosome partitioning (28, 7.7%) , and only 9.9% ESTs (36) classified as novel genes (Fig. 4B).

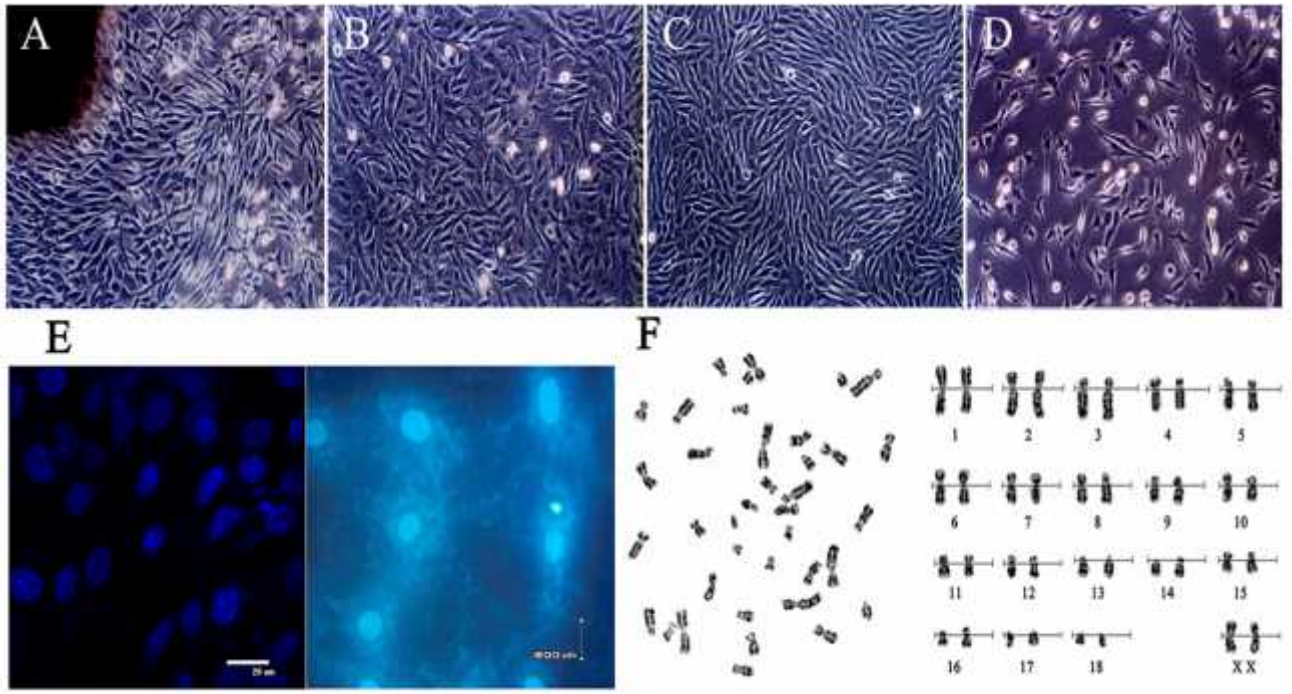

Fig. 1. Morphology, Mycoplasma contamination and karyotype of Siberian tiger cell line. A: Primary cells (100×); B: Subcultured cells (100×). C: Cells before cryopreservation (100×); D: Cells after recovery (100×); E: Mycoplasma contamination Stained with Hoechst33258 and Positive control (400×); F: Chromosome at metaphase (left) and karyotype (right) (♀, 1,000×).

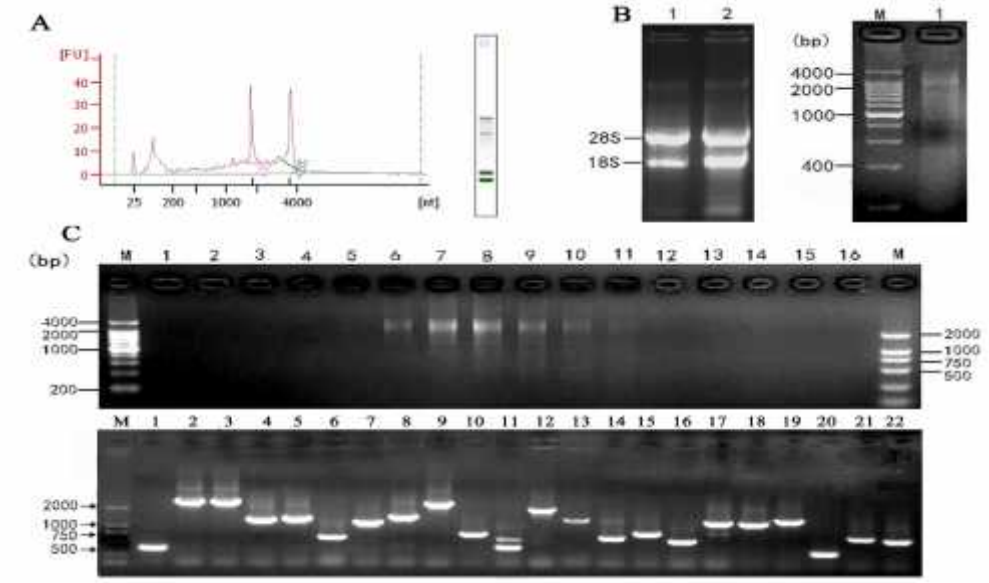

Fig. 2. Construction of the Siberian tiger fibroblast cells cDNA library.
A: The quality of total RNA was estimated by examining a sample on the Agilent 2100 Bioanalyzer; B: Total RNA from fibroblast cells of Siberian tiger and LD-PCR with 22 cycles. C: cDNA size fractionation by CHROMA SPIN-400 and Recombinant clones screening within the library.

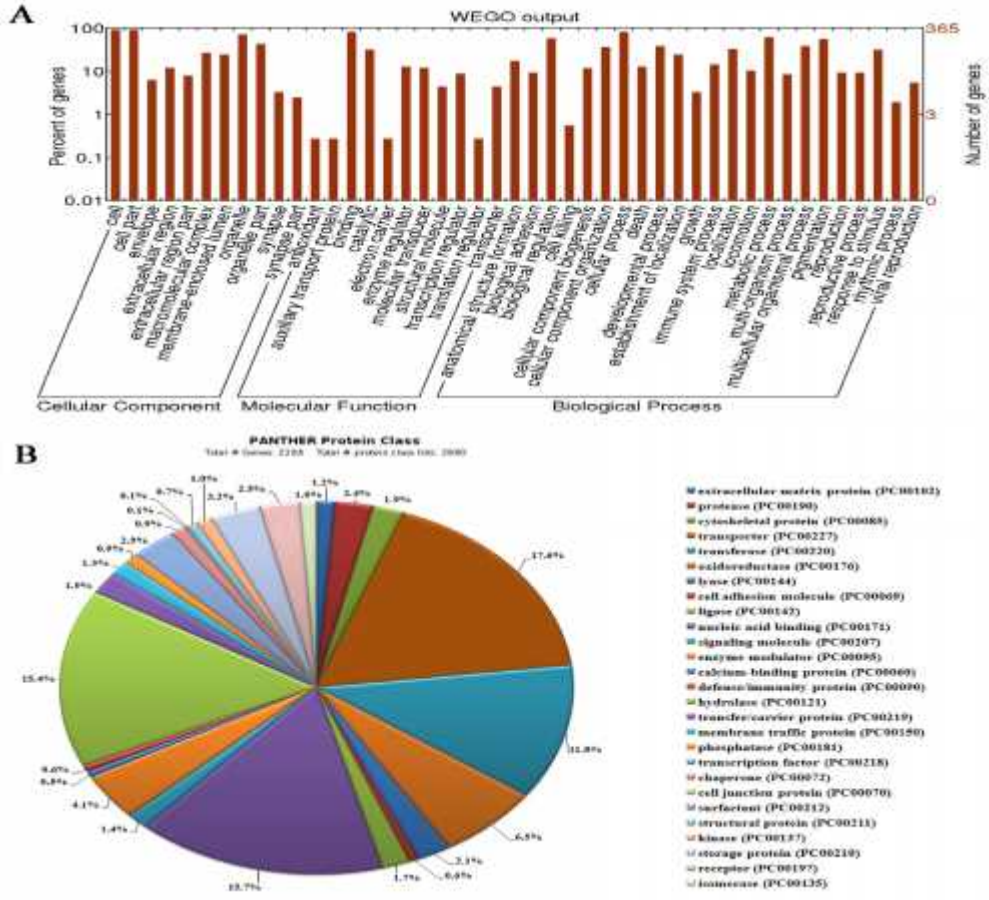

Fig. 3. Functional annotation of GO analysis of the Siberian tiger unigenes. A: GO categories at the 2nd level are plotted, are grouped into three main ontologies: biological process, cellular component and molecular function; B: GO classification for Siberian tiger UniGenes using PANTHER expression analysis tool.

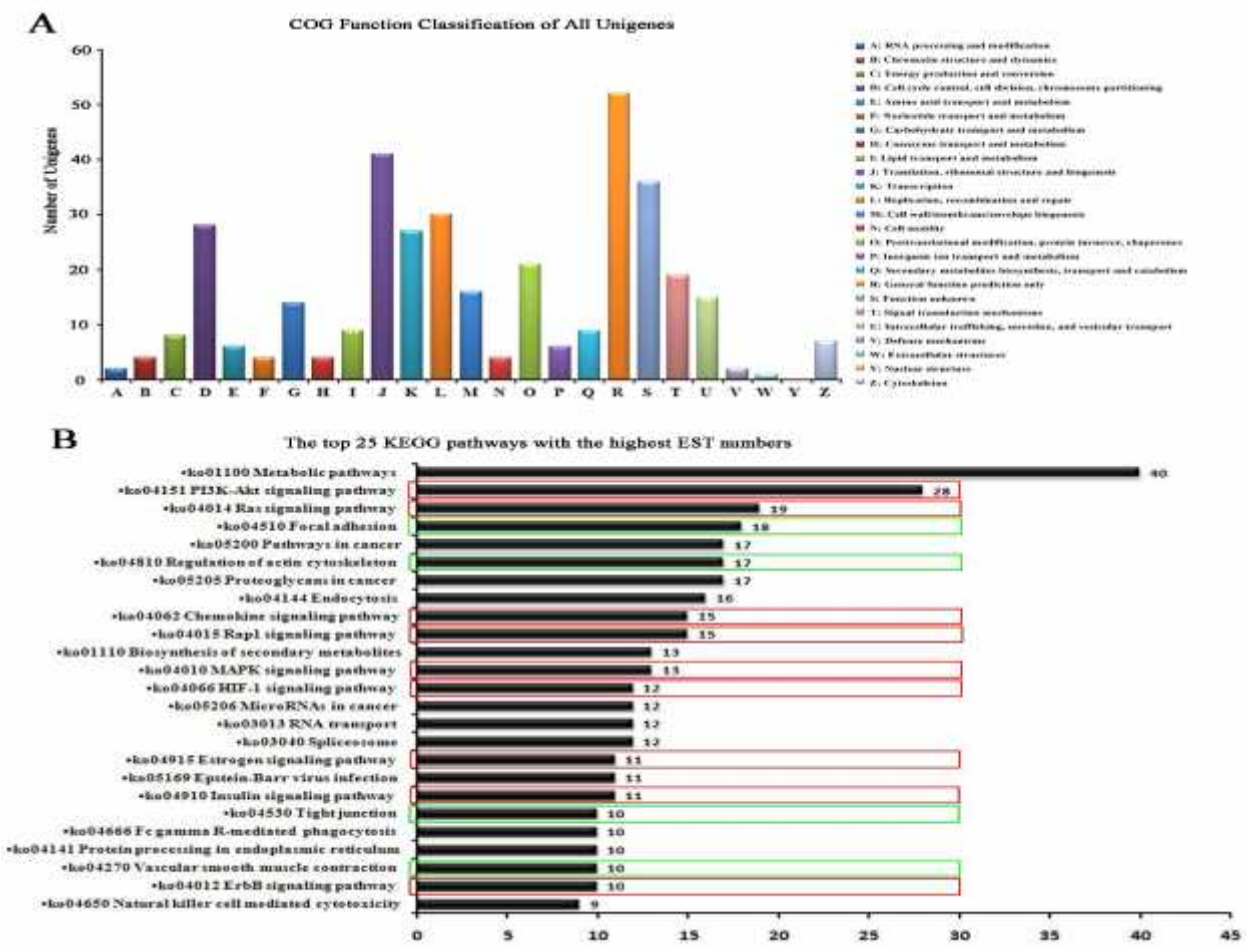

Fig. 4. Clusters of COG classification and KEGG pathways analysis of the Siberian tiger unigenes. A: COG categories, 365 unigenes were assigned to 25 categories in the COG classification; B: 234 unigenes were assigned to 244 KEGG pathways, the top-25 pathways was shown. Nine of the top-25 pathways were mapped to the pathways associated with signaling pathways (indicated with red box). In addition, four of top-25 pathways were mapped to the pathways associated with cytoskeleton and cell junction (indicated with green box).

DISCUSSION

Over the past years, cDNA library construction and analysis is considered to be an indispensable tool for functional genome analysis as it provides much more detailed information on the genomic mechanisms underlying diverse processes of the organism (Wellenreuther *et al*., 2004; Ling *et al*., 2007; Li *et al*., 2009; Shao *et al*., 2009; Al-Taweel *et al*., 2011). Generation of ESTs is an excellent and unique approach in molecular studies as it allows both expression and measurements, and the discovery of new genes to be conducted at the same time (Blair *et al*., 2011). Recently, emphasis on EST sequencing has waned due to the advent of next generation sequencing techniques that can quickly dissect a transcriptome. However, full length cDNA sequences will remain useful in the discovery of alternative splice sites and for unraveling paralogs within gene families.

Previously, only few tiger ESTs had been identified in NCBI database despite it being an important wild animal. In our study, the primary cDNA library, instead of the amplified library, was used for generation of ESTs to reduce the redundancy of cDNA clones as only a small number of ESTs were targeted through random selection. Four hundred and five-six white clones were picked randomly for EST sequencing. After removal of the vector sequences and low-quality sequences, 426 effective sequences from the total cDNA sequences were obtained, a total of 365 individual ESTs were analyzed and they ranged from 264 to 1,378 nucleotides in length. For species distribution, 57.3% of the unigenes sequences have top matches (first hit) trained with sequences from the domestic cat, which suggests that the relationships between tiger and domestic

cat.

A total of 365 unigenes of Siberian tiger were assigned for GO analysis, and 305 unigenes were successfully annotated (about 83.6% of the assembled unigenes) and involved in 47 functions. In each of the three main categories of the GO classification, Cell, Cellular process, Binding, organelle and Metabolic process are the dominant terms. A total of 234 unigenes were assigned to 244 KEGG pathways, and the numbers of transcripts in different pathways ranged from 1 to 40. Among the 25 COG categories, "General function prediction only cluster" represents the largest group. Consequently, analysis of the expression of a large number of genes combined with the knowledge of their functions can facilitate the understanding and allows us to take a glimpse of the overall picture of biological processes in *P. t. altaica* fibroblast cells. In this study, approximately 90.1% of the ESTs generated were sequences with known or putative functions, while the remainders were unknown proteins or sequences with no similarities to the databases.

In conclusion, we constructed a full-length cDNA library of Siberian tiger using the cultured cell *in vitro* to save the genetic resources of nationally protected Siberian tiger at cell and molecular level. This important genomic resource offers an efficient way to identify more genes and information of this majestic species. We obtained and characterized the high-quality Siberian tiger EST sequences, and the number of genes identified in tiger achieved a major leap from 57 to 422.

Acknowledgements: This research was supported by the National Natural Scientific Foundation of China (grant no. 81301333), the Ministry of Agriculture of China for Transgenic Research Program (grant no. 2013ZX08009-003-006), Science and Technological Fund of Anhui Province for Outstanding Youth (grant no. 1408085QH155) and the Key Natural Science Project of Anhui Provincial Education Department (grant no. KJ2012A199).

REFERENCES

Al-Taweel, K., W.G. Dilantha Fernando, and A.L. Brule-Babel (2011). Construction and Characterization of a cDNA Library from Wheat Infected with Fusarium graminearum Fg 2. Int. J. Mol. Sci. 12(1): 613-626.

Blair, M.W., A.C. Fernandez, M. Ishitani, D. Moreta, M. Seki, S. Ayling, and K. Shinozaki (2011). Construction and EST sequencing of full-length, drought stress cDNA libraries for common beans (Phaseolus vulgaris L.). BMC Plant. Biol. 11: 171.

Cho, Y.S., L. Hu, H. Hou, H. Lee, J. Xu, S. Kwon, S. Oh, H.M. Kim, S. Jho, S. Kim, Y.A. Shin, B.C. Kim, H. Kim, C.U. Kim, S.J. Luo, W.E. Johnson, K.P. Koepfli, A. Schmidt-Kuntzel, J.A. Turner, L. Marker, C. Harper, S.M. Miller, W. Jacobs, L.D. Bertola, T.H. Kim, S. Lee, Q. Zhou, H.J. Jung, X. Xu, P. Gadhvi, P. Xu, Y. Xiong, Y. Luo, S. Pan, C. Gou, X. Chu, J. Zhang, S. Liu, J. He, Y. Chen, L. Yang, Y. Yang, J. He, S. Liu, J. Wang, C.H. Kim, H. Kwak, J.S. Kim, S. Hwang, J. Ko, C.B. Kim, S. Kim, D. Bayarlkhagva, W.K. Paek, S.J. Kim, S.J. O'Brien, J. Wang, and J. Bhak (2013). The tiger genome and comparative analysis with lion and snow leopard genomes. Nat. Commun. 4: 2433.

Gotz, S., J.M. Garcia-Gomez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, M. Robles, M. Talon, J. Dopazo, and A. Conesa (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36(10): 3420-3435.

Kanehisa, M., M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi (2008). KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36: D480-484.

Li, J.Y., H.Y. Wang, J .Liu, Q. Liu, J.S. Zhang, F.C. Wan, F.J. Liu, S.H. Jin, and Y.L. Zhang (2008). Transcriptome analysis of a cDNA library from adult human epididymis. DNA Res. 15(3): 115-122.

Li, Y.P., R.X. Xia, H. Wang, X.S. Li, Y.Q. Liu, Z.J. Wei, C. Lu, and Z.H. Xiang (2009). Construction of a full-length cDNA Library from Chinese oak silkworm pupa and identification of a KK-42-binding protein gene in relation to pupa-diapause termination. Int. J. Biol. Sci. 5(5): 451-457.

Ling, P., M. Wang, X. Chen, and K.G. Campbell (2007). Construction and characterization of a full-length cDNA library for the wheat stripe rust pathogen (Puccinia striiformis f. sp. tritici). BMC Genomics, 8: 145.

Liu, C., D. Liu, Y. Guo, T. Lu, X. Li, M. Zhang, J. Ma, Y. Ma, and W. Guan (2013). Construction of a Full-Length Enriched cDNA Library and Preliminary Analysis of Expressed Sequence Tags from Bengal Tiger Panthera tigris tigris. Int. J. Mol. Sci. 14(6): 11072-11083.

Liu, C.Q., T.F. Lu, B.G. Feng, D. Liu, W.J. Guan, and Y.H. Ma (2010). Construction of cDNA library and preliminary analysis of expressed sequence tags from Siberian tiger. Int. J. Biol. Sci. 6(6): 584-589.

Luo, S.J., J.H. Kim, W.E. Johnson, J. van der Walt, J. Martenson, N. Yuhki, D.G. Miquelle, O. Uphyrkina, J.M. Goodrich, H.B. Quigley, R.

Tilson, G. Brady, P. Martelli, V. Subramaniam, C. McDougal, S. Hean, S.Q. Huang, W. Pan, U.K. Karanth, M. Sunquist, J.L. Smith, and S.J. O'Brien (2004). Phylogeography and genetic ancestry of tigers (Panthera tigris). PLoS Biol. 2(12): e442.

Phukon, M., R. Namdev, D. Deka, M.K. Modi, and P. Sen (2012). Construction of cDNA library and preliminary analysis of expressed sequence tags from tea plant [Camellia sinensis (L) O. Kuntze]. Gene, 506(1): 202-206.

Shao, Z.T., X. Cong, J.D. Yuan, G.W. Yang, Y. Chen, J. Pan, and L.G. An (2009). Construction and characterization of a cDNA library from head kidney of Japanese sea bass (Lateolabrax japonicus). Mol. Biol. Rep. 36(7): 2031-2037.

Sultan, M., M.H. Schulz, H. Richard, A. Magen, A. Klingenhoff, M. Scherf, M. Seifert, T. Borodina, A. Soldatov, D. Parkhomchuk, D. Schmidt, S. O'Keeffe, S. Haas, M. Vingron, H. Lehrach, and M.L. Yaspo (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science, 321(5891): 956-960.

Thanh, T., V.T. Chi, M.P. Abdullah, H. Omar, M. Noroozi, H. Ky, and S. Napis (2011). Construction of cDNA library and preliminary analysis of expressed sequence tags from green microalga Ankistrodesmus convolutus Corda. Mol. Biol. Rep. 38(1): 177-182.

Wellenreuther, R., I. Schupp, A. Poustka, S. Wiemann, and cDNA.C. German (2004). SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. BMC Genomics, 5(1): 36.

Wheeler, D.L., T. Barrett, D.A. Benson, S.H. Bryant, K. Canese, V. Chetvernin, D.M. Church, M. Dicuccio, R. Edgar, S. Federhen, M. Feolo, L.Y. Geer, W. Helmberg, Y. Kapustin, O. Khovayko, D. Landsman, D.J. Lipman, T.L. Madden, D.R. Maglott, V. Miller, J. Ostell, K.D. Pruitt, G.D. Schuler, M. Shumway, E. Sequeira, S.T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, R.L. Tatusov, T.A. Tatusova, L. Wagner, and E. Yaschenko (2008). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 36: D13-21.

Ying, S.Y. (2004). Complementary DNA libraries: an overview. Mol. Biotechnol. 27(3): 245-252.