Review Articles

Application of expressed sequence tag (EST) for exploration of genetic diversity in domestic animals

Damri Darawiroj*

Department of Anatomy, Faculty of Veterinary Science, Chulalongkorn University Bangkok, Thailand,10330

***Corresponding author, E-mail address: ddamri@chula.ac.th**

Abstract

A short single-pass DNA sequence obtained from either end of cDNA clones called expressed sequence tag (EST) provides a rapid and reliable method for gene discovery and a resource for the large-scale analysis of gene expression of known and unknown genes in various organisms. Using of this high-throughput approach reveals partial or complete structure and function of interesting genes in both plants and animals. At present, gene information obtained from EST are globally collected in major public database such as the dbEST of GenBank and are widely used as major source for microarray analysis.This article summarized the application of EST in domestic animals during the past decades.

Keywords: expressed sequence tag, gene, domestic animals

การตรวจหาความหลากหลายทางพันธุกรรมของสัตว์เลี้ยง ด้วยวิธีการใช้เครื่องหมาย DNA ชนิด Expressed sequence tag (EST)

ดำริ ดาราวิโรจน์*

ภาควิชากายวิภาคศาสตร์ คณะสัตวแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพฯ ประเทศไทย 10330

***ผู้รับผิดชอบบทความ E-mail address: ddamri@chula.ac.th**

บทคัดย่อ

การใช้เครื่องหมาย DNA ชนิด expressed sequence tag(EST) ซึ่งเป็น DNA สายสั้นๆ ที่ได้จากการอ่านลำดับเบสเพียงรอบเดียวจากปลายสาย DNA ด้านใดด้านหนึ่งเป็นวิธีการที่ทำได้รวดเร็ว และมีผลที่เชื่อถือได้ในการค้นหายีนของสิ่งมีชีวิต รวมทั้งเป็นแหล่งข้อมูลที่ดีในการศึกษาลักษณะการแสดงออกของยีนที่สนใจทั้งยีนที่มีการศึกษามาแล้วและยีนที่ไม่เคยมีการศึกษามาก่อนของสิ่งมีชีวิตต่างๆ เทคนิค EST ยังเป็นวิธีการที่ได้ข้อมูลที่หลากหลายของทั้งโครงสร้างและการทำหน้าที่ของยีนซึ่งอาจจะเป็นเพียงบางส่วน หรือทั้งหมดของยีนที่สนใจไม่ว่าจะเป็นในพืชหรือสัตว์ก็ตาม ในปัจจุบันข้อมูลพันธุกรรมที่ได้จาก EST ได้ถูกเก็บบันทึกอยู่ในฐานข้อมูลสาธารณะขนาดใหญ่ เช่น dbEST ในฐานข้อมูล GenBank บทความนี้ได้รวบรวมข้อมูลการประยุกต์ใช้ EST ในสัตว์เลี้ยงในตลอดทศวรรษที่ผ่านมาจนถึงปัจจุบัน

คำสำคัญ : expressed sequence tag ยีนสัตว์เลี้ยง

Introduction to expressed sequence tag (EST)

From many genome projects, there are 15,000-50,000 genes in complex multicellular organisms. Nonetheless, only some of these genes are expressed as messenger RNA in a particular type of cell. In addition, each of these genes can be expressed at very different levels at any given time. It has been realized that the part of genome expressed as mature mRNA, collectively known as the transcriptome, contains much of the information of interest. An efficient approach called expressed sequence tag (EST) has been developed to identify a bulk of gene expression from particular tissues at different conditions. This high-throughput cDNA sequencing, was introduced in 1991 by a significant study from Venter and his colleagues. (Adam et al.1991, Okubo et al. 1992) ESTs refer to small pieces of DNA sequences (typically about 200 to 500 nucleotides) generated by sequencing either one or both ends of an expressed gene. Fabrication of EST is demonstrated in figure 1. EST sequences contain at least partial sequence of most mRNAs which present in the various tissues used for library construction. This single-pass sequencing is an important aspect of making the approach cost effective and less inaccurate (2% error, approximately) in the discovery of new genes and identification of coding regions of genomic sequences. Therefore, it has been used extensively as a source of information for the discovery of new genes whose function can be tentatively deduced from their sequence. It is also employed for gene mapping as marker-assisted selection or EST marker (Liu et al. 1999). Furthermore, information gained from genome sequencing EST that provide large sets of annotated clones and sequences are further used for the construction of microarray analysis (Johnston,1998). The most common use of ESTs, however, is gene identification by database homology search, but they have also been used for gene expression profiling in different tissues as well as for other genomic studies. The results of these comparisons are used as a guide to assign putative identifications to the cDNAs. Researchers may then search or browse through the putative identifications of the ESTs to determine which genes may be of interest for further study. An experimentally obtained partial protein or nucleic acid sequence can also be compared with the EST database to find out whether the cDNA for the gene has already been cloned. These methods have allowed the rapid identification of genes in a number of organisms and have accelerated research by providing genetic material for further investigation.

In 1992, a central repository for EST sequences, called dbEST, was established at NCBI (Boguski et al. 1993).The biggest contributors to the human and mouse EST sequencing projects have been the Washington University Merck EST sequencing project (Hillier et al. 1996) and the project funded by the Howard Hughes Medical Institute (Marra et al. 1999), respectively. Currently, ESTs generates greater information on eukaryotic genes than any other data source (Parsons and Rodriguez-Tome, 2000). The number of EST sequences in the dbEST division of GenBank was increased from 3 million ESTs in 1999 (Kimura et al. 2004) to nearly 74 million ESTs from over a thousand species in 2012 suggesting that the EST approach has been widely adopted. Nevertheless, EST information is available mostly for a few model species, which unable to be the representatives of the major taxa since model species have generally been selected for practical reasons, for instance small genome size, short life cycle or easily detected mutational phenotypes. The majority of reported animal ESTs is obtained from mouse followed by pig,cattle and zebrafish respectively (Table 1).

EST analysis in domestic animals

Pig

At present, number of pig EST collected in dbEST is 1,669,337 and is counted as the top of all food animal species. However, compared to human and mouse, the numbers of pig and other livestock species EST are

relatively low.The porcine gene mapping is promising scheme for porcine-human comparative traits. Conducting of porcine ESTs are a valuable resource of such a process. Consequently various porcine tissue are performed EST analysis such as white blood cells, small intestine, liver, ovary and nervous tissue. (Zhang et al. 2007, Claus et al. 1997, Tosser-Klopp et al. 1997). Moreover, genes from skeletal muscle (Davoli et al. 1999, 2002) and backfat (Ren et al. 2006) are also considered of relevant interest because they are the target tissues that determine meat production and quality. Biceps femoris (Zhang et al. 2007) and latissimus dorsi (Davoli et al. 1999) are selected as muscle samples for study.

Cattle

Most of bovine EST analysis aims at generating more comprehensive maps between the human and the cattle genomes (Stone et al. 2002). The utilization of human genomic sequence for integrating bovine ESTs into existing linkage maps by using SNPs as markers is positively practicable nowadays. Thus, it is possible to target most candidate genes or genetic intervals for SNP marker development and genotyping. Bovine EST derived from cDNA libraries of target tissues, for instance, mammary gland (Karall-Albrecht et al. 2000, Sonstegard et al. 2002), ovary (Ma et al. 1998), muscle and liver (Lim et al. 2010). Analyzing sequences from mammary gland tissue is of interest because genes expressed and the corresponding proteins in the lactating mammary gland could have an impacton milk production and health conditions. The generation of microsatellite markers using EST sequences has recently conducted in bovine because it can be rapidly developed at a low cost (Yan et al. 2008).

Horse

In the past decade, the problems of horse reproduction have been overcome by the application of genomic approach such as gene mapping, genome scanning including EST analysis. A genome sequence of the Thoroughbred mare was freely available in 2007, this resource will be useful to compare annotated human and mouse genes with the horse genome (Rothwell et al. 2004). A number of laboratories constructed horse ESTs based on cDNA cloned from variety of tissuesuch as skin (Lieto and Cothran, 2001) and white blood cell (Matia-Ovic, 2006). However, a few studies on horse EST are currently reported. EST number of horses in dbEST of NCBI is also scarce.

Domestic dog

The domestic dog *Canis familiaris*, bears approximately 360 diseases which are analogous to human diseases, consequently, the dog has been used as a valuable model for identification and study of disease loci in many heritable human diseases. Even though the complete genome sequence of the domestic dog isreported in 2005(Ostrander and Wayne, 2005), studies of canine genome analysis using EST technique was recently published (Kim et al. 2011, Kim et al. 2012). These studies were aimed at identification of single-nucleotide polymorphisms (SNPs) for the dog population. SNPs collected from EST are used as stable genetic markers to assess many individual genetic characteristics in canines such as; coat variation (Cadieu et al. 2009), behavior (Vage and Lingaas, 2008) and sensitive traits (Lesniak et al. 2008). Construction of cDNA libraries for EST analysis in domestic dog was obtained from a few organs such as brain, liver, kidney and testis (Kim et al. 2011, Palmer et al. 2003). At present, the number of dog ESTs recorded in dbEST is nearly 400000.

Domestic cat

Surprisingly, in contrast to another domestic animals, publications and public information of domestic cat EST is undersized. Despite the fact that the domestic catis not only a valuable model for toxoplasmosis and viral leukemia in human but also the model for the related endangered felids,only a research on cat EST derived from uterus is found. (Pathak and Kapil, 2006)

Chicken

Besides using for the generation of molecular marker for mapping immune-related genes responding to particular diseases and for developmental biology research as a model organism, the chicken ESTs are also constructed for monitoring and conservation on genetic diversity in this species. A recent study showed the possibility of using EST-derived microsatellite markers for investigating the *Gallus gallus* genome. (Bakhtiarizadeh et al. 2012) The chicken cDNA libraries are derived from various organs such as spleen, bursa, thymus, bone marrow, peripheral blood lymphocytes, pituitary, hypothalamus, pineal, abdominal fat, and oviduct (Carre et al. 2006) For identification of immune-relevance genes, ESTs are collected from both normal and infected organs of chicken. For instance, several cytokine gene sequences were obtained from EST-generated cecal tonsil cDNA (Rothwell et al. 2004).

Fish

A number of ESTs gernerated from various organs of several fish is recorded in ESTdb such as common carp (Zhu et al. 2012), zebrafish (Zeng and Gong, 2002), Atlantic salmon (Micallef et al. 2012), Japanese medaka (Hirono and Aoki, 1997) and yellowtail (Darawiroj et al. 2008). Most of all fish EST is constructed to collect gene information for improving the quality and quantity of commercial fish such as the expression profiling of disease resistance genes, metabolism-enhanced genes and genes related to growth and development of target organs. However, ESTs from ornamental fish were also found in the dbEST.

In conclusion, EST is the high-throughput methods for gene discovery and sequence diversity as well as being an important resource for study transcription profiling of genes of interest in various animals. The ESTs recorded in public database allow many scientists to exchange and share the biological data across species. Application of ESTs to new approach tools, for instance; microarray, has been widely conducted and resulted in more understanding mechanism, function, evolution and diversity of animal genetics.

Table 1 Summary of animal EST in dbEST of NCBI

Domestic animals	Number of EST
Mus musculus + domesticus (mouse)	4,853,570
Sus scrofa (pig)	1,669,337
Bos taurus (cattle)	1,559,495
Danio rerio (zebrafish)	1,488,275
Rattus norvegicus + sp. (rat)	1,162,136
Oryzias latipes (Japanese medaka)	666,891
Gallus gallus (chicken)	600,434
Salmo salar (Atlantic salmon)	498,245
Canis lupus familiaris (dog)	382,638
Ovis aries (sheep)	338,483

(www.ncbi.nlm.nih.gov/dbEST, as of January 2013)

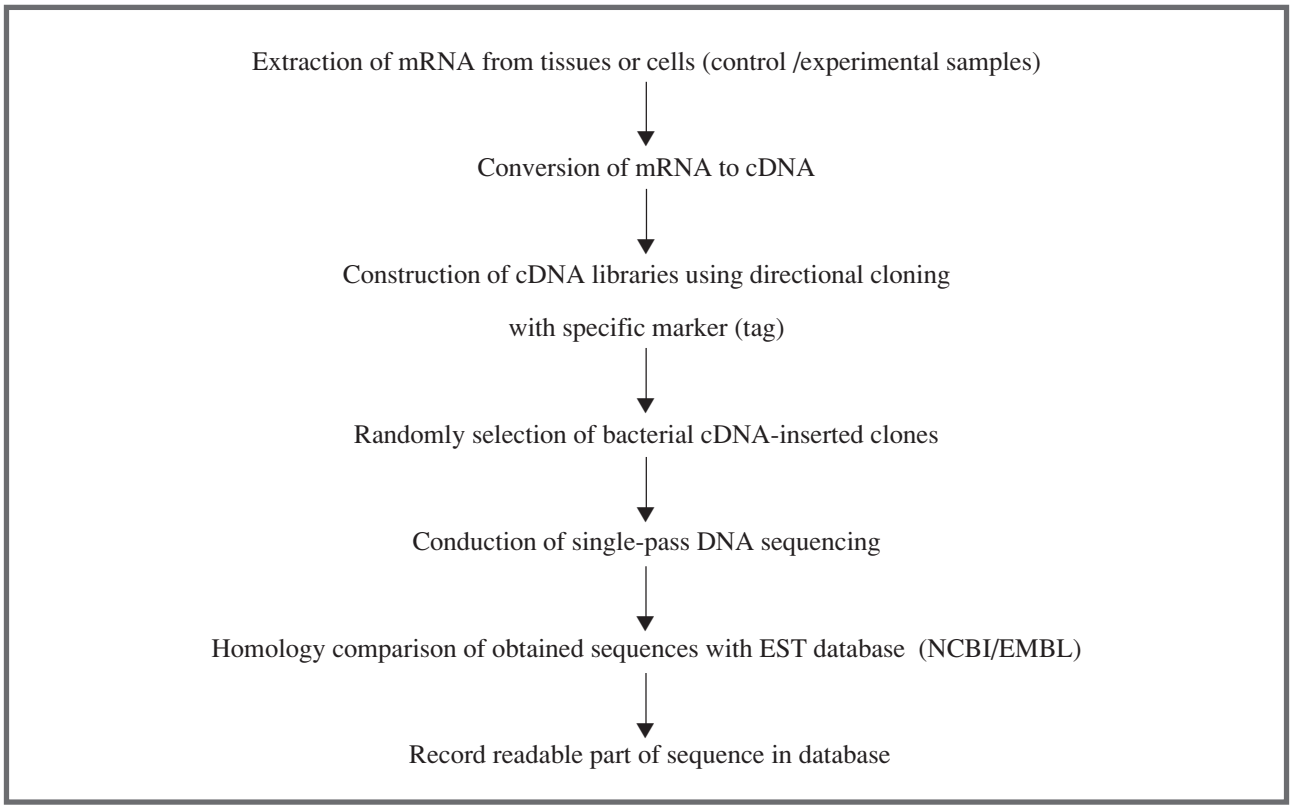

Figure 1 Diagram of EST construction

References

Adams M.D., Kelley J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B. and Moreno R.F. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science, 252. 1651-6

BakhtiarizadehM.R., Arefnejad B., Ebrahimie E. and Ebrahimi M. (2012). Application of functional genomic information to develop efficient EST-SSRs for the chicken *(Gallus gallus)*.Genet Mol Res, 11(2). 1558-74

Boguski M.S., Lowe T.M. and Tolstoshev C.M. (1993). dbEST-database for "expressed sequence tags". Nat Genet, 4. 332-3

Cadieu E., Neff M.W., Quignon P., Walsh K., Chase K., Parker H.G., Vonholdt B.M., Rhue A., Boyko A., Byers A., Wong A., Mosher D.S., Elkahloun A.G., Spady T.C., Andre C., Lark K.G., Cargill M., Bustamante C.D., WayneR.K. and Ostrander E.A. (2009). Coat variation in the domestic dog is governed by variants in three genes. Science, 326. 150-53

Carre W., Wang X., Porter T.E., Nys Y., Tang J., Bernberg E., Morgan R., Burnside J., Aggrey S.E., Simon J. and Cogburn L.A. (2006). Chicken genomics resource: sequencing and annotation of 35,407 ESTs from single and multiple tissue cDNA libraries and CAP3 assembly of a chicken gene index. Physiol Genomics, 25(3). 514-24

Claus B.J., Winter A.K., Yerle M. and Fredholm M. (1997). Mapping of 22 expressed sequence tags isolated from a porcine small intestine cDNA library.Mammalian Genome, 8. 423-27

Darawiroj D., Kondo H., Hirono I. and Aoki T. (2008). Immune-related gene expression profiling of yellowtail *(Seriola quinqueradiata)* kidneycells stimulated with ConA and LPS usingmicroarray analysis. Fish & Shellfish Immunol, 24. 260-66

Davoli R., Fontanesi L., Zambonelli P., Bigi D., Gellin J., Yerle M., Milc J., Braglia S., Cenci V., Cagnazzo M. and Russo V. (2002). Isolation of porcine expressed sequence tags for the construction of a first genomic transcript map of the skeletal muscle in pig. Animal Genetics, 33. 3-18

Davoli R., Zambonelli P., Bigi D.and Fontanesi L. (1999). Analysis of expressed sequence tags of porcine skeletal muscle. Gene, 233. 181-88

Hillier L.D., Lennon G., Becker M., Bonaldo M.F., Chiapelli B., Chissoe S., Dietrich N., DuBuque T., Favello A., Gish W., Hawkins M., Hultman M., Kucaba T., Lacy M., Le M., Le N., Mardis E., Moore B., Morris M., Parsons J., Prange C., Rifkin L., Rohlfing T., Schellenberg K. and Marra M. (1996). Generation and analysis of 280,000 human expressed sequence tags. Genome Res, 6. 807-28

Hirono I. and Aoki T.(1997). Expressed sequence tags of medaka *(Oryzias latipes)* liver mRNA. Mol Mar Biol Biotechnol,6(4). 345-50

Johnston M. (1998). Gene chips: array of hope for understanding gene regulation. Curr Biol, 8. R171-4

Karall-Albrecht C., Groenen M.A.M., van der Poel J.J., Barendse W., Womack J.E., Kalm E. and LooftC. (2000).Mapping of 16 ESTs expressed in the bovinemammarygland during lactation. Mammalian Genome, 11. 320ñ25

Kim J.Y., Park H.S., Lim D., Jang H.C., Park H.S., Lee K.T., Kim J.S., Oh S.I., Kweon M.S., Kim T.H. and Choi1 B.H. (2011). Functional analysis of expressed sequence tags from the liver and brain of Korean Jindo dog.BMB reports,44(4). 224-38

Kim R.N., Kim D.S., Choi S.H., Yoon B.H., Kang A., Nam S.H., Kim D.W., Kim J.J., Ha J.H., Toyoda A., Fujiyama A., Kim A., Kim M.Y., Park K.H., Lee K.S. and Park H.S. (2012).Genome analysis of the domestic dog (Korean Jindo) by massively parallel sequencing. DNA Res,19(3). 275-87

Kimura T., Jindo T., Narita T., Naruse K., Kobayashi D., Shin I.T., Kitagawa T., Sakaguchi T., Mitani H., Shima A., Kohara Y. and Takeda H. (2004). Large-scale isolation of ESTs from medaka embryos and its application to medaka developmental genetics.Mech Dev, 121. 915-32

Lesniak A., Walczak M., Jezierski T., Sacharczuk M., Gawkowski M. and Jaszczak K. (2008). Canine olfactory receptor gene polymorphism and its relation to odor detection performance by sniffer dogs. J Hered, 99. 518-27

Lieto L.D. and Cothran E.G.(2001). Characterization of expressed sequence tags generated from skin cDNA clones of *Equus caballus* by single pass sequencing. Ani Biotechnol, 12(1). 87-97

Lim D.J., Lee S.H., Cho Y.M., Yoon D.H., Shin Y.H., Kim K.W., Park H.S and Kim H.B. (2010). Transcript profiling of expressed sequence tags from intramuscular fat, longissimus dorsi muscle and liver in Korean cattle (Hanwoo). BMB reports, 43(2). 115-121

Liu Z., Karsi A. and Dunham R.A. (1999). Development of polymorphic EST markers suitable for genetic linkage mapping of catfish.Mar Biotechnol (NY), 1. 437-47

MaR.Z., van Eijk M.J., Beever J.E., Guerin G., Mummery C.L. and Lewin H.A. (1998). Comparative analysis of 82 expressed sequence tags from a cattle ovary cDNA library. Mamm Genome, 9. 545-49

Marra M., Hillier L., Kucaba T., Allen M., Barstead R., Beck C., Blistain A., Bonaldo M., Bowers Y., Bowles L., Cardenas M., Chamberlain A., Chappell J., Clifton S., Favello A., Geisel S., Gibbons M., Harvey N., Hill F., Jackson Y., Kohn S., Lennon G., Mardis E., Martin J., Mila L., McCann R., Morales R., Pape D., Person B., Prange C., Ritter E., Soares M., Schurk R., Shin T., Steptoe M., Swaller T., Theising B., Underwood K., Wylie T., Yount T., Wilson R. and Waterston R.(1999). An encyclopedia of mouse genes. Nat Genet, 21. 191-4

Matia-Ovic J., Kralik P., Leva L., Faldyna M. and Horin P. (2006). Suppressive subtraction hybridization on stimulated primary horse macrophages. Acta Vet Brno, 75. 337-41

Micallef G., Bickerdike R., Reiff C., Fernandes J.M., Bowman A.S. and Martin S.A. (2012). Exploring the transcriptome of Atlantic salmon *(Salmo salar)* skin, a major defense organ.Mar Biotechnol (NY), 14(5). 559-69

Okubo K., Hori N., Matoba R., Niiyama T., Fukushima A., Kojima Y. and Matsubara K. (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nat Gene, 2. 173-9

Ostrander E. A. and Wayne R. K. (2005) The canine genome. Genome Res, 15. 1706-16

Palmer L.E., O'Shaughnessy A.L., Preston R.R., Santos L., Balija V.S., Nascimento L.U., Zutavern T.L., Henthorn P.S., Hannon G. J. and McCombie W.R. (2003). A survey of canine expressed sequence tags and a display of their annotations through a flexible web-based interface. J ered, 94 (1). 15-22

Parsons J.D. and Rodriguez-Tome P. (2000). JESAM: CORBA software components to create and publish EST alignments and clusters. Bioinformatics, 16. 313-25

Pathak S. and Kapil S. (2006). Expressed sequence tags from feline uterine library DNA. Sequence, 17(2). 87-93

Ren Z.Q., Xiong Y.Z., Deng C.Y., Lei M.G., Zuo B., LiF. E., Zheng R. and Xu D.Q. (2006). Isolation, cloning, and identification of expressed sequence tags from the backfat tissue in Duroc and Tongcheng pigs by mRNA differential display. Agricultural Sciences in China, 5(2). 141-145

Rothwell L., Young J.R., Zoorob R., Whittaker C.A., Hesketh P., Archer A., Smith A.L. and Kaiser P. (2004). Cloning and characterization of chicken IL-10 and its role in the immune response to *Eimeria maxima*. J Immunol, 173(4). 2675-82

Sonstegard T.S., Capuco A.V., White J., van Tassell C.P., Connor E.E., Cho J., Sultana R., Shade L., Wray J.E., Wells K.D. and Quackenbush J. (2002). Analysis of bovine mammary gland EST and functional annotation of the *Bos taurus* gene index. Mammalian Genome, 13(7). 373-79

Stone R.T., Grosse W.M., Casas E., Timothy P.L., Smith J.W., Keele and Bennett G.L. (2002).Use of bovine EST data and human genomic sequences to map 100gene-specific bovine markers. Mammalian Genome, 13. 211ñ15

Tosser-Klopp G., Benne F., Bonnet A., Mulsant P., Gasser F. and Hatey F. (1997). A first catalog of genes involved in pig ovarian follicular differentiation. Mamm. Genome, 8. 250-54

Vage J. and Lingaas F. (2008). Single nucleotide polymorphisms (SNPs) in coding regions of canine dopamine- and serotonin-related genes. BMC Genet, 9. 10

Yan Q., Zhang Y., Li H.B., Wei C.H., Niu L., Guan L., Li S.G. and Du L.X. (2008). Identification of microsatellites in cattle unigenes. J Genet Genom, 35(5). 261-66

Zeng S. and Gong Z. (2002).Expressed sequence tag analysis of expression profiles of zebrafish testis and ovary. Gene, 294(1-2). 45-53

Zhang L., Tao L., Ye L., He L., Zhu Y.Z., Zhu Y.D. and Liang Y.Z. (2007).Alternative splicing and expression profile analysis of expressed sequence tags in domestic pig.Genomics, Proteomics & Bioinformatics, 5(1). 25-34

Zhu C., Cheng L., Tong J. and Yu X. (2012). Development and characterization of new single nucleotide polymorphism markers from expressed sequence tags in common carp *(Cyprinus carpio)*. Int J Mol Sci, 13(6). 7343-53