



Construction of a full-length cDNA library and preliminary analysis of expressed sequence tags from lymphocytes of half-pipe snowboarding athletes

Y.H. Zhao^{1*}, Z.B. Zhang^{1*}, C.Q. Zhao¹, Y. Zhang¹, Y.F. Wang¹, W.J. Guan² and Z.Q. Zhu¹

¹Harbin Institute of Physical Education, Nangang District, Harbin, Heilongjiang Province, China

²Department of Animal Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Haidian District, Beijing, China

*These authors contributed equally to this study.

Corresponding authors: Z.Q. Zhu / W.J. Guan

E-mail: zhiqiangzhu999@163.com / weijunguan301@gmail.com

Genet. Mol. Res. 14 (4): 12921-12930 (2015)

Received April 17, 2015

Accepted July 30, 2015

Published October 21, 2015

DOI <http://dx.doi.org/10.4238/2015.October.21.13>

ABSTRACT. The genes of top athletes are a valuable genetic resource for the human race, and could be exploited to identify novel genes related to sports ability, as well as other functions. We analyzed the expressed sequence tags from top half-pipe snowboarding athletes using the SMART complementary DNA (cDNA) library construction method to elucidate the characteristics of the athlete genome and the differential expression of the genes it contains. Overall, we established a full-length cDNA library from the lymphocytes of half-pipe snowboarding athletes and analyzed the inserted gene fragments. We also classified those genes according to molecular function, biological characteristics, cellular composition, protein types, and signal paths. A total of 201 functional genes were noted, which were distributed in 27 pathways. *TXN*, *MDH1*, *ARL1*, *ARPC3*, *ACTG1*, and

other genes measured in sequence may be associated with physical ability. This suggests that the SMART cDNA library constructed from the genetic material from top athletes is an effective tool for preserving genetic sports resources and providing genetic markers of physical ability for athlete selection.

Key words: Half-pipe snowboarding; Full-length cDNA library; Lymphocyte; Expressed sequence tags

INTRODUCTION

Half-pipe snowboarding became an official event at the 1998 Winter Olympics in Nagano, Japan (Turnbull et al., 2011). In recent years, research has shown that genetic factors make a significant contribution to the differences in athletic ability between individuals. As the study of molecular biology continues, a large number of genes associated with athletic ability have been identified, and the study of athletes using biotechnological techniques is currently a popular focus area for research (Williams and Folland, 2008). A greater understanding of the effects of gene expression regulation in athletes is affected by two issues: the limited genetic resources available and the ongoing research into functional genes. Constructing a cDNA library has become the first choice for resolving these issues.

Since the mid-1970s, when the results of the first cDNA cloning were published, cDNA library construction has become one of the basic tools of functional gene research (Zhu et al., 2001). Construction of a cDNA library involves using all the mRNA in cells as templates for reverse transcription and converting host bacteria by combining vector DNA with the products of transcription, thereby forming new recombinant DNA clones. The technique is now widely used for gene recognition and to splice variants, and can be an effective resource for full-length cloning (Wiemann et al., 2003; Shao et al., 2009). cDNA library construction is the most effective way of determining information about the complete genome, and is a rapid, economic pathway for functional research; it greatly promotes the study of functional genomes. A review of the research conducted on full-length cDNA library construction methods investigated the use of SMART, Cap-trapper, Oligo-capping, Cap-ture, and Cap-jumping resources (Zhu et al., 2001). This research uses SMART (switching mechanism at the 5' end of the RNA transcript) to construct a cDNA library of snowboarding athletes. The advantage lies in reducing the amount of the initial-stage RNA, whilst rapidly and easily obtaining a higher percentage of full-length cDNA (Chen et al., 2005).

MATERIAL AND METHODS

Sample preparation

Athlete-derived blood (from 25 athletes) and Epstein-Barr virus (EBV)-transformed B-cell lines (from 25 athletes) were obtained from the Institute of Animal Sciences at the Chinese Academy of Agricultural Sciences. The athletes were associated with the LG 2010/2011 International Snowboard World Cup China National Team. All sampling procedures were approved by the local ethics committee, and the subjects gave their written informed consent for the procedures.

Construction of the cDNA library

Lymphocyte cells were harvested and total RNA was extracted with TRIzol reagent (Invitrogen, Carlsbad, CA, USA). First and double-stranded cDNAs were synthesized according to the protocol of the SMART cDNA Library Construction kit (Clontech, Palo Alto, CA, USA). Subsequently, a first-strand cDNA sample (approximately 3 μ L) was amplified using long-distance polymerase chain reaction (LD-PCR). The first four peak fractions containing cDNAs (N 500 bp) were pooled together using column chromatography with CHROMA SPIN-400 medium. The cDNAs were ligated to λ TriplEx2 vector (1:2) and packaged with Gigapack III Gold Packaging extract.

Titering the unamplified library

The potential concentration of the plaque-forming units (pfu/ μ L) of the unamplified library was estimated by the number of independent phages and independent clones in the library. Different dilutions of the packaging extract (1:5 to 1:20) for an unamplified λ lysate were mixed with 500 μ L *Escherichia coli* VCS257 (optical density at 600 nm (OD_{600}) = 4.0), and the mixture was incubated at 37°C for 20 min. Melted LB/MgSO₄ top agar (2 mL) was then added. The mixture was then quickly inverted and immediately poured onto 90-mm LB/MgSO₄ plates that had been pre-warmed to 37°C. When the top agar had solidified after cooling the plates at room temperature for 10 min, the plates were inverted and incubated at 37°C for 6-18 h. The plates were periodically checked to ensure that plaques were developing. The plaques were then counted and the titer of the unamplified library was calculated. pfu/mL = (number of plaques x dilution factor x 10³ μ L/mL) / (diluted phage plated μ L).

Determining the percentage of recombinant clones

The recombination efficiency was evaluated by performing blue/white screening of the *E. coli* VCS257. The titer of the unamplified library was tested on the LB/MgSO₄ plates and before plating, the phage + bacteria mixtures, isopropyl β -D-1-thiogalactopyranoside (IPTG), and X-gal were added to the melted top agar. We used 50 μ L IPTG and X-gal stock solution for every 2 mL melted top agar to form 500-1000 plaques/90-mm plate. The plates were incubated at 37°C for 6-18 h to develop blue color in the plaques. The ratio of white (recombinant) and blue (non-recombinant) plaques was used to estimate the recombination efficiency of the cDNA library.

Library amplification

The number of plates required depends on how many independent clones will be amplified in the library. A general guideline is to aim for 6-7 x 10⁴ clones per 150-mm plate when working with λ TriplEx2. Therefore, a library of 1 x 10⁶ clones requires 20 LB/MgSO₄ 90-mm plates to amplify the unamplified library. With the objective of forming 4-5 x 10⁴ plaques per 90-mm plate, the unamplified library was mixed with 500 μ L *E. coli* VCS257 (OD_{600} = 4.0), and the mixture was incubated at 37°C for 20 min. The melted LB/MgSO₄ top agar (2 mL per plate) was then combined with the mixture by quickly inverting and immediately pouring onto 90-mm LB/MgSO₄ plates that had been pre-warmed to 37°C. The plates were cooled at room temperature for 10 min until the top agar solidified. The plates were then inverted and incubated at 37°C for

6-18 h. SM buffer (12 mL) was added to each plate and the plates were stored at 4°C overnight in order to form an amplified library lysate. After incubating the plates on a platform shaker (50 rpm) for 1-2 h at room temperature, the λ -phage lysates were poured into a sterile beaker. This amplified-library lysate was then titered using the amplified-library blow method. The amplified library can be stored at 4°C for 6 months. The completed cDNA libraries were stored in 7% dimethyl sulfoxide at -80°C for long-term storage.

Titering the amplified library

The titer (pfu/ μ L) of the amplified library was estimated by the number of independent phages and independent clones in the library. The different dilutions of the unamplified library (1:500 to 1:1000) were mixed with 500 μ L *E. coli* VCS257 ($OD_{600} = 2.0$), and the mixture was incubated at 37°C for 15 min. The melted LB/MgSO₄ top agar (2 mL) was then added to the mixture, and the contents were quickly mixed and immediately poured onto 90-mm LB/MgSO₄ plates that had been pre-warmed to 37°C. The plates were quickly swirled after pouring to promote an even distribution of the melted agar. The plates were then cooled at room temperature for 10 min to allow the top agar to solidify. The plates were incubated at 37°C for 6-18 h, during which they were periodically checked to ensure that plaques were developing. The plaques were counted and the titer of the amplified library was calculated as follows: pfu/mL = (number of plaques x dilution factor x 10³ μ L/mL)/(diluted phage plated μ L). The dilution factor = 1 x 10⁴ (Liu et al., 2010).

Determination of the cDNA insert size by PCR

Phage clones (96) were randomly selected to determine the cDNA insert size by PCR. The phage clones were placed in 100 μ L SM buffer and incubated at 37°C for 3-5 h while shaking at 150 rpm. The primers (cDNA F: CCATTGTGTTGGTACCCGG; cDNA R: ATACGACTCACTATAGGGCGAATT) were used to determine the cDNA insert size, and the phage clones served as templates for PCR. The PCR product (3-5 μ L) was analyzed on a 1.2% agarose gel.

Generation of expressed sequence tags and sequence analysis

After the λ TriplEx2 vectors had been converted to pTriplEx2, the cDNA clones were selected randomly from the cDNA library and single-pass sequenced (5' sequencing primer: CTCGGGAAGCGCGCCATTGTGTTGGT) at the 5' end on an ABI 3730 Genetic Analyzer (Applied Biosystems). The low-quality sequences and the empty vectors were then neglected. All sequences shorter than 100 bp were removed (Li et al., 2008). All the unigenes were annotated using BLASTN (NCBI) (E-value, 10⁻¹⁰). Gene ontology (GO) analysis was used to investigate the major function of the genes on the basis of the key functional GO classifications in the Swiss-Prot database. Generally, the GO category was classified by Fisher's exact test, and the false discovery rate was calculated to correct the P value. Similarly, pathway analysis was used to determine which pathways the genes participated in, according to the Kyoto encyclopedia of genes and genomes (KEGG). The Fisher's exact test was used to select important pathways; the critical value of significance was defined by the P value and false discovery rate, and the degree of enrichment was estimated (Ying, 2004; Guo et al., 2014; Li et al., 2014).

RESULTS

Library construction

Total RNA was extracted and tested; the OD_{260}/OD_{230} ratio was 2.10, and the OD_{260}/OD_{280} ratio was 1.95, which indicates that the RNA extracted was not contaminated by DNA, proteins, or other impurities. The total RNA amounted to 105 μg , and the concentration was 3450 $\text{ng}/\mu\text{L}$. As can be seen in Figure 1A, two bright bands attributable to 18S rRNA and 28S rRNA are clearly visible, indicating that the total RNA was integrated and stable enough for cDNA library construction. LD-PCR was carried out and the double-stranded DNA produced a collection of bands representing molecules of 0.5-4 kb, as shown in Figure 1B. The methods used are described in detail by Hu et al. (2014).

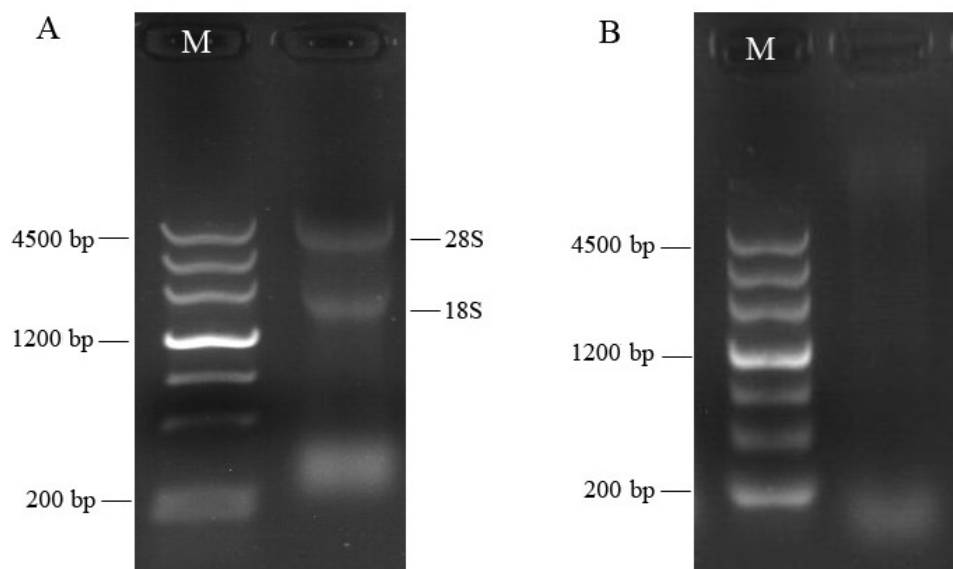


Figure 1. Total RNA from athletes and LD-PCR. **A.** Total RNA from half-pipe snowboarding athlete lymphocytes. **B.** LD-PCR amplification products analyzed by agarose gel electrophoresis.

Characterization of cDNA library

A CHROMA SPIN-400 column was used to purify the cDNAs, which were up to 4500 bp in length (Figure 2A). cDNA was collected in tubes 8, 9, and 10. Large pieces of cDNA were connected to the vector and used in phage transformation. The VSC257 host bacteria were selected and screened, and the inserted fragment sizes and titers were determined. The titers of primary and amplified libraries were 1.28×10^6 pfu/mL and 1.70×10^{10} pfu/mL, respectively. The recombination efficiency of the amplified libraries was 91.2%. The insert ratio and the average length of inserted fragments were measured by PCR. The size of the inserted fragments ranged from 100 to 1100 bp (Figure 2B) which suggests that the insertion fragments covered most of the mRNAs and reached the requirement for further studies on gene structure, translation, expression, and function.

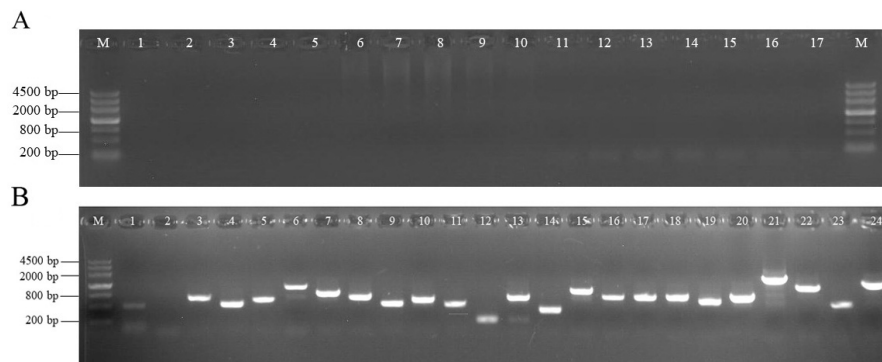


Figure 2. Construction of the half-pipe snowboarding athlete lymphocyte cDNA library. **A.** cDNA size fractionation by CHROMA SPIN-400. **B.** Recombinant clones screening within the library.

Library amplification

The vectors that contained cDNAs were transfected to VCS257 bacteria in lysogeny broth culture medium and the bacteriophage plaque method was used to select the amplified cDNAs. The purified cDNA library can be stored at 4°C for 6 months.

Sequencing and sequence assembly

In total, 406 positive clones were randomly selected for sequencing. Of these, 360 sequences were effective, and 201 sequences were calibrated and assembled. Sequence lengths ranged from 100 to 1100 bp, and the average length was 606 bp (Table 1, Figure 3).

Table 1. Summary of the analysis of the half-pipe snowboarding athlete cDNA library.

Description	Number
Total number of sequences	406
Number of valid sequences	360
Average expressed sequence tag (EST) length (nt)	606
Number of unigenes	201
Number of annotated genes	143

Generation of expressed sequence tags and sequence analysis

To further explore the function of the annotated genes, pathway analysis was performed using KEGG. A total of 143 unigenes were mapped to 39 KEGG pathways, and the genes functioned in 27 different aspects. Table 2 lists the top 10 annotated pathways (Figure 4).

In order to obtain more information, 143 unigenes were successfully annotated and classified according to their functions by GO analysis. As shown in Figure 5, the 143 annotated unigenes were involved in 31 functions, and could be classified into three major groups: cellular components (29.0%), biological processes (25.8%), and molecular functions (45.2%). Within the category "Molecular Function", the subcategories "structural molecule", "binding", and "catalytic" contained the highest number of unigenes, and the majority of gene functions were related to the regulation of protein activity.

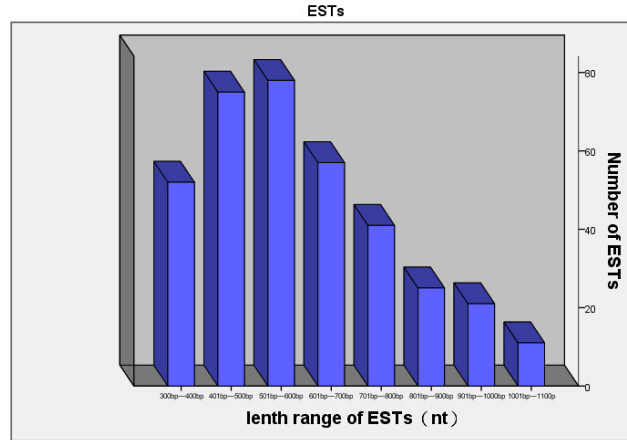


Figure 3. Distribution of the sequence length of the half-pipe snowboarding athlete lymphocyte cDNA library.

Table 2. Top 10 pathways in the half-pipe snowboarding athletes' cDNA library, mapped using KEGG pathway analysis.

KEGG pathway ID	No. of clusters	Pathway	Genes of pathway
P00034	4 (10.3%)	Integrin signaling pathway	ARL1, ARPC3, ACTG1, ITGAE
P00031	4 (10.3%)	Inflammation mediated by chemokine and cytokine signaling pathway	ARPC3, CCL3L1, ACTG1, SOCS4
P00012	3 (7.7%)	Cadherin signaling pathway	FER, CSNK2A1, ACTG1
P00016	2 (5.1%)	Cytoskeletal regulation by Rho GTPase	ARPC3, ACTG1
P02740	2 (5.1%)	De novo pyrimidine ribonucleotides biosynthesis	ACTG1, NME1
P00057	2 (5.1%)	Wnt signaling pathway	CSNK2A1, SMARCA2
P00010	1 (2.6%)	B cell activation	BLNK
P00030	1 (2.6%)	Hypoxia response via HIF activation	TXN
P00046	1 (2.6%)	Oxidative stress response	TXN
P00051	1 (2.6%)	TCA cycle	MDH1

KEGG = Kyoto Encyclopedia of Genes and Genomes.

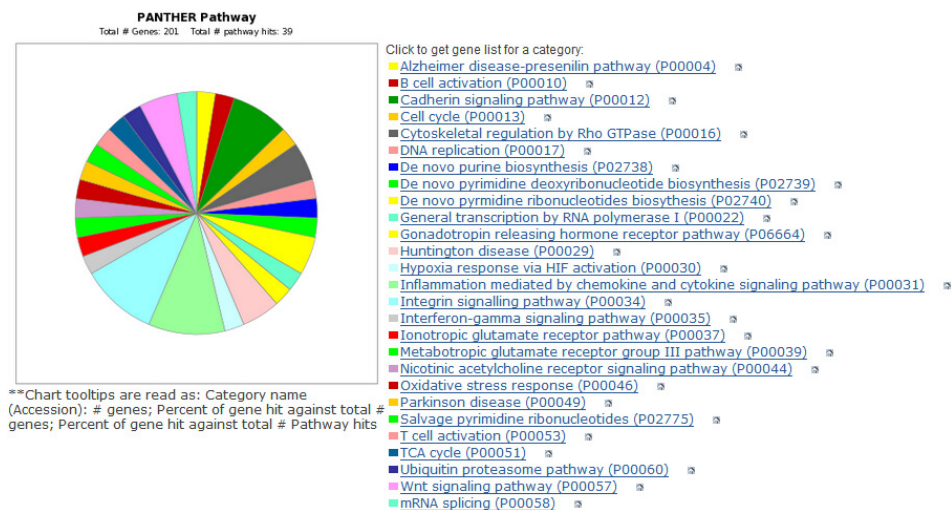


Figure 4. KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway Analysis.

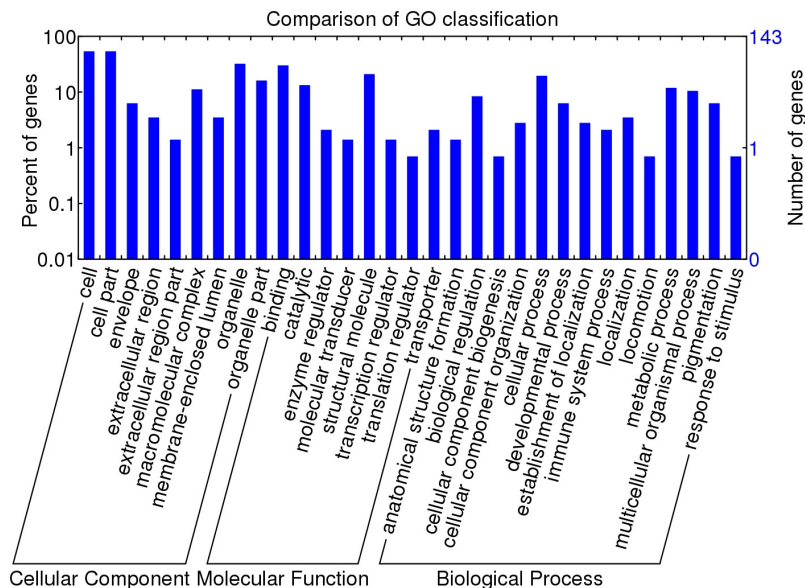


Figure 5. Functional annotation of the athlete cDNA library unigenes. Gene Ontology categories at the 2nd level are plotted; "Biological process", "Cellular component" and "Molecular function" are categorized independently in this ontology.

DISCUSSION

Previous research has shown that a cDNA library is an indispensable tool for functional genomics analysis, and provides detailed information on the mechanism of diversification of gene groups in organisms. With the advent of the appropriate technology, the construction of cDNA libraries is becoming increasingly refined, and has led to a few valuable research discoveries in the fields of agricultural, biology, and medicine. However, few discoveries have been made that are relevant to sport. The construction of a cDNA library effectively preserves the genetic information of top athletes and separates full-length gene function; it also provides an important resource for information about the genes of athletes at the transcriptional level. This information is useful for research on biological phenomena including ontogeny, cell differentiation, cell cycle control, and cell aging, and it has other broader applications (Wheeler et al., 2008). Because the cDNA library is easy to clone and express, a gene screened objectively using the cDNA library can be used for direct expression, and cDNA libraries are the most commonly used gene expression libraries in research (Thanh et al., 2011).

Several chief aspects can be used to identify the quality of a cDNA library. According to the Clarke-Carbon formula, a cDNA library should contain at least 1.7×10^5 independent clones to ensure that the 99% of low-abundance mRNAs are present in the library. High recombination efficiency is another index of library quality. The average length of inserted cDNA should be no less than 1.0 kb to ensure the integrity of the cDNA. Finally, the recombination rate, which is an important indicator of the quality of the evaluation library, should be at least 90% (Liu et al., 2010; Liu et al., 2013). In the present study, we obtained unamplified titers of 1.28×10^6 pfu/mL and 1.70

$\times 10^{10}$ pfu/mL, and inserts between 100 and 1100; 91.2% of the recombinant cDNA library indicated that good quality can provide a basis for the in-depth study of the transcriptional levels of athletes. Another study demonstrated expressed sequence tags (EST) analysis for the study of functional genomics in terms of nucleotide sequences greater than 300 bp that represent information (Pratt et al., 2005). A total of 406 clones were sequenced after the removal of randomly selected sequences of less than 300 bp, leaving 360 sequences of lengths ranging from 300 to 1100, with an average length of 606 bp, which further indicates that the library was well constructed. This study could be used to build a library for EST sequencing. In this study, we constructed veneer U of the full-length cDNA library constructed using the samples from the athletes, in line with the requirements of a standard library.

Adams et al. (1991) put forward the concept that ESTs adopted from a cDNA library can be used for large-scale functional gene studies. In the present study, we sequenced 201 functional genes using ESTs obtained from KEGG pathway analysis, and found that the genes were distributed on 27 pathways including those involved in the following: integrin signaling, inflammation mediated by chemokine and cytokine signaling, cytoskeletal regulation by Rho GTPase, the cadherin signaling pathway, de novo pyrimidine ribonucleotides biosynthesis, oxidative stress response, the TCA cycle, hypoxia response via HIF activation, the integrin signaling pathway, and other channels involving metabolism, oxidative stress, and heart function associated with the vascular system. We found that the gene library comprising *TXN*, *MDH1*, *ARL1*, *ARPC3*, and *ACTG1* played a role in each of those pathways. We therefore infer that those roles may be involved in athletic-related functions. Finally, we used EST GO to classify 143 functional genes involving 30 different functions, mostly related to metabolism and protein regulation, including cellular components, molecular functions, and biological processes in cellular components; molecular functional response is a separate function of the gene product, and biological processes are a comprehensive reflection of the complexity of various life activities. ESTs can provide information on physical function, exercise capacity, and related information for follow-up research on the performance of athletes.

Using cDNA library screening, Bellinge et al. (2005) studied the myostatin (*MSTN*) gene, which affects mouse muscles. They used a human cDNA library to identify the *MSTN* gene, which has a strong negative effect on skeletal muscle strength. In this experiment, we used peripheral blood lymphocytes because they are readily available and carry large amounts of legacy information. Previous studies have shown that human lymphocyte gene expression profiles are associated with various diseases. This is an important area of research. Lymphocytes are immune cells that secrete interleukins and other important cytokines, which are not only involved in immune function, but also with cellular signal transduction, growth, blood cells, and nerve cell regeneration. These functions are closely related and have a wide range of applications, as described by Mezey et al. (2000). In this study, we constructed a cDNA library of peripheral blood lymphocytes from athletes. Our aim was to create a functional gene library based on a large number of ESTs isolated from elite athletes in order to confirm an association between genes and athletic ability. The technique may also be useful and convenient for studying gene expression in relation to the lymphatic and nervous systems of athletes. Therefore, a cDNA library constructed from the peripheral blood lymphocytes of athletes may have a wide range of applications.

CONCLUSION

The effective preservation of genetic information from top athletes is essential before that information can be fully understood. Building a high-quality cDNA library will provide abundant

materials for further research and a realistic basis for the selection of elite talent and top athletes at the molecular level.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Agricultural Science and Technology Innovation Program (ASTIP) (#cxgc-ias-01).

REFERENCES

- Adams MD, Kelley JM, Gocayne JD, Dubnick M, et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252: 1651-1656.
- Bellinge RH, Liberles DA, Iaschi SP, O'Brien PA, et al. (2005). Myostatin and its implications on animal breeding: a review. *Anim. Genet.* 36: 1-6.
- Chen XH, Chen Z, Yao HP, Chen F, et al. (2005). Construction and characterization of a cDNA library from human liver tissue with chronic hepatitis B. *J. Zhejiang Univ. Sci. B* 6: 288-294.
- Guo Y, Liu C, Lu T, Liu D, et al. (2014). Generation and analysis of a large-scale expressed sequence tags from a full-length enriched cDNA library of Siberian tiger (*Panthera tigris altaica*). *Gene* 541: 75-81.
- Hu PF, Li XC, Liu HK, Guan WJ, et al. (2014). Construction and characterization of a cDNA expression library from the endangered Hu sheep. *Genet. Mol. Res.* 13: 9019-9023.
- Li JY, Wang HY, Liu J, Liu Q, et al. (2008). Transcriptome analysis of a cDNA library from adult human epididymis. *DNA Res.* 15: 115-122.
- Li Q, Lu TF, Liu D, Hu PF, et al. (2014). Screening and analyzing genes associated with Amur tiger placental development. *Genet. Mol. Res.* 13: 7869-7878.
- Liu CQ, Lu TF, Feng BG, Liu D, et al. (2010). Construction of cDNA library and preliminary analysis of expressed sequence tags from Siberian tiger. *Int. J. Biol. Sci.* 6: 584-589.
- Liu C, Liu D, Guo Y, Lu T, et al. (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: 11072-11083.
- Mezey E, Chandross KJ, Harta G, Maki RA, et al. (2000). Turning blood into brain: cells bearing neuronal antigens generated *in vivo* from bone marrow. *Science* 290: 1779-1782.
- Pratt LH, Liang C, Shah M, Sun F, et al. (2005). Sorghum expressed sequence tags identify signature genes for drought, pathogenesis, and skotomorphogenesis from a milestone set of 16,801 unique transcripts. *Plant Physiol.* 139: 869-884.
- Shao ZT, Cong X, Yuan JD, Yang GW, et al. (2009). Construction and characterization of a cDNA library from head kidney of Japanese sea bass (*Lateolabrax japonicus*). *Mol. Biol. Rep.* 36: 2031-2037.
- Thanh T, Chi VT, Abdullah MP, Omar H, et al. (2011). Construction of cDNA library and preliminary analysis of expressed sequence tags from green microalga *Ankistrodesmus convolutus* Corda. *Mol. Biol. Rep.* 38: 177-182.
- Turnbull J, Keogh JW and Kilding AE (2011). Strength and conditioning considerations for elite snowboard half pipe. *Open Sports Med. J.* 5: 1-11.
- Wheeler DL, Barrett T, Benson DA, Bryant SH, et al. (2008). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 36: D13-D21.
- Wiemann S, Mehrle A, Bechtel S, Wellenreuther R, et al. (2003). CDNAs for functional genomics and proteomics: the German Consortium. *C. R. Biol.* 326: 1003-1009.
- Williams AG and Folland JP (2008). Similarity of polygenic profiles limits the potential for elite human physical performance. *J. Physiol.* 586: 113-121.
- Ying SY (2004). Complementary DNA libraries: an overview. *Mol. Biotechnol.* 27: 245-252.
- Zhu YY, Machleder EM, Chenchik A, Li R, et al. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* 30: 892-897.