Development of Expressed Sequence Tags(ESTs) from Korean Native Chicken cDNA Libraries

Sang Su Shin[1], Ki Duk Song[2], Jee Hye Shin[2], Sun Duck Lee[2], Young Mok Lee[1], Jin Kyu Kim[3] and Jae Yong Han[1]

[1]School of Agricultural Biotechnology, Seoul National University / [2]Avicore Biotechnology Institute, [3]Department of Microbiology, Chang Won National University, Korea

Abstract

한국 재래 닭의 유전적 특성 규명 및 기능 유전체 연구의 기초재료를 확보하고자 대량 EST 염기서열 결정 및 생물정보학 분석을 실시하였다. 대량 EST 염기서열 분석을 위한 첫 번째 단계로 재래 닭의 뇌, 비장, 정소, 배아 생식기를 이용하여 cDNA library를 구축하였다. 각각의 library로부터 총 15,121개의 클론을 선정하여 염기서열을 결정하였다. 생물정보학 분석결과 15,121개의 염기서열은 총 10,353개의 contig로 정리되었다.

이들 염기서열을 기존 데이터베이스를 대상으로 tBlastX(http://www.ncbi.nlm.nih.gov/BLAST) 분석을 실시한 결과, 염기서열 중 56 %가 기존 데이터베이스에 존재하는 유전자와의 상동성을 보였다. 상동성을 보이는 유전자들은 유전자의 구조 및 기능 분석에 이용될 것이고, 상동성을 보이지 않는 유전자들은 microarray와 같은 대량 유전자 발현분석 시스템을 이용하여 선별한 후 기능분석이 실시될 것이다.

(Key words : Mass sequencing, Korean Native Chicken, Expressed Sequence Tags(ESTs), Bioinformatics)

Introduction

Expressed Sequence Tags(ESTs) development is an essential step toward functional genomics study. Chicken ESTs are being mainly developed by Tirunagaru et al. in USA and European research groups deposited to GenBank. As the first step of genomics research in Korean Native Chicken(KNC), we have initiated cDNA library construction and single pass sequencing of KNC ESTs from four different organs; brain, spleen, testis, embryonic gonads. Using informatics analysis, we screened genes which were homologous to mammalian genes and to do in depth research about these genes. Functional genomics approach will be also performed and screened for identifying the useful genes with characterization of their functions.

Materials and Methods

Korean Native Chicken cDNA libraries were made using brain, spleen, testis and 6.5-day-old embryonic gonad. cDNAs from mRNA of each tissue were cloned into UNIZAP-XR lambda vector and packaged in vitro. Plasmids were converted by in vitro excision with helper phages from each

lambda library. Sequences of plasmids from each library were determined by automated sequencing using ABI-377 machine with T7 primer. Single passed sequences were preprocessed for bioinformatics analysis. After quality trimming including vector trimming, calling and removal of contaminating sequences, such as bacterial sequences, mitochondrial sequences etc., the processed sequences were subjected to BlastN and BlastX analysis toward public databases.

Result and Discussion

We have constructed cDNA libraries using KNC brain, spleen, testis and embryonic gonad. The titer of each library was greater than $2X10^5$ pfu/ml with insert size from 0.5-3.0kb and it was enough to perform phagemid conversion. After phagemid conversion, total 15,121 clones were picked and subjected to single pass sequencing. By bioinformatics analysis of our sequence data, total 10,353 contigs were assembled from 15,121 cDNA sequences. Fifty six percent of total contigs showed to be homologous to sequences in public databases such as GenBank and EMBL etc. Average length of sequence read was 675.7bp and read number per a contig was 1.34kb, which were compatible to Boardman et al.(2002) suggesting KNC cDNA libraries developed might be valuable resources for genomics study of chicken. Based on sequence homology, chicken sequences homologous to mammalian genes will be characterized and subjected to functional study. Function of non-homologous sequences will be studied after screening by expression profiles with microarray.

References

- Boardman PE, Sanz-Ezquerro J, Overton IM, Burt DW, Bosch E, Fong, WT, Tickle C, Brown WR, Wilson, S.A., and Hubbard, S.J. 2002. A comprehensive collection of chicken cDNAs. *Curr Biol* 12(22):1965-1969.
- Buerstedde JM, Arakawa H, Watahiki A, Carninci PP, Hayashizaki YY, Korn B, and Plachy J 2002. The DT40 web site: sampling and connecting the genes of a B cell line. *Nucleic Acids Res* 30(1):230-231.
- Tirunagaru VG, Sofer L, Cui J, and Burnside J 2000. An expressed sequence tag database of T-cell-enriched activated chicken splenocytes : sequence analysis of 5251 clones. *Genomics* 66(2):144-1451.