

# 연체동물 전용 BLAST 서버 구축에 관한 연구

이 용 석

한국생명공학연구원 인간유전체연구실

## 서론

바이오테크놀로지와 관련된 기계, 전자 산업의 급속한 발전에 따라 생물학데이터는 대량화 되고 있으며, 이 대량화된 정보를 처리하기 위해 컴퓨터의 이용은 필수적인 요소가 되어 버렸다. 이러한 데이터는 유전체 서열정보(genome), 유전자 서열정보(gene), 아미노산 서열 정보를 비롯하여, 단백질의 3차 구조와 관련된 문헌정보 까지도 종합되고 있다. 이러한 정보들은 NCBI, EMBL 등을 통해 현재 모두 공공 웹 사이트들을 통해 무상으로 공개하고 있다. 이러한 공공 웹사이트 들에는 연체동물에 관련한 자료들도 다수 포함하고 있으나, 다른 동물 카테고리에 비하여 자료가 그리 많지 않다. 특히 사람, 마우스, 래트, 복어 등 게놈프로젝트가 완성된 동물들의 경우에는 게놈정보가 이미 공개되어 있어, 연구자들이 어떠한 유전자에 대한 연구를 할 때 많은 도움을 받을 수 있으나, 연체동물을 포함한 다른 동물의 경우엔 그렇지 못한 실정이다. 또한 가장 많이 이용하는 blast 프로그램 이용 시 카테고리 상에서 멀리 떨어진 생물들과 비교를 하게 되므로 검색이 그리 쉽지 않으며, 전 세계의 연구자들이 공동으로 사용하고 있어 속도가 그리 빠르지 않은 단점을 가지고 있다. 이에 대부분의 공공 웹사이트 및 기타 생물연구자 들은 단독으로 데이터베이스를 구축하여 사용하는 경우가 많다. (ex. ANODB ; *Anopheles gambiae* 관련 DB) 이에 연체동물을 단독으로 한 BLAST 서버는 필연적으로 요구되어진다고 할 수 있다.

## 방법

사용된 PC 서버는 펜티엄4 2.4 celeron cpu를 사용하였으며, Linux(Redhat 7.3)를 기반으로 하여 구축하였다. 운영체제 설치 후 apache 웹서버 기반에서 cgi(common gate interface)를 사용할 수 있도록 환경설정을 한 후 NCBI wwwblast package를 이용하였다. (<http://bioinfo.sch.ac.kr>)

NCBI 에 등록되어 있는 연체동물과 관련된 게놈정보(미토콘드리아 게놈정보), 유전자서열 정보, 아미노산 서열정보를 모두 다운 받은 후, 멀티파스타 형태의 정보로 만든후 NCBI 에서 제공하는 formatdb 프로그램을 사용하여 데이터베이스화 한 후 cgi 서버를 이용하여 STAND ALONE BLAST 서버를 구축하였다. 또한 실험 후 데이터 확인 시 필요한 벡터 및 *e. coli* 시퀀스를 모두 입력하여, 실험한 데이터를 검증 시 용이하도록 하였으며, est 연구 시 용이하도록 est 및 est\_others 데이터를 이용한 blast 검색이 가능하도록 하였다. 그리고 primer3 등 실험시 부가적으로 필요한 웹용 프로그램을 인스톨 하여 연구자들이 편리하게 이용하도록 하였다.

## 결과

BLAST 검색 시 연체동물관련 뉴클레오타이드 정보와 일치된 결과만을 따로 얻을 수 있었으며, vector 및 *e. coli* 시퀀스와 동시에 검색이 가능하므로 library 제작 시 데이터 검증에 용이하며, repeat sequence 검색도 동시에 가능하도록 되어 있어, 데이터 분석 시 용이하였다. 또한 EST 데이터들을 이용한 검색이 가능하므로, 유전자의 function 연구 및 차후 진행 가능성이 가장 많은 연체동물 EST 실험이 진행될시 보조 자료를 얻을 수 있는 시스템을 갖추었다.

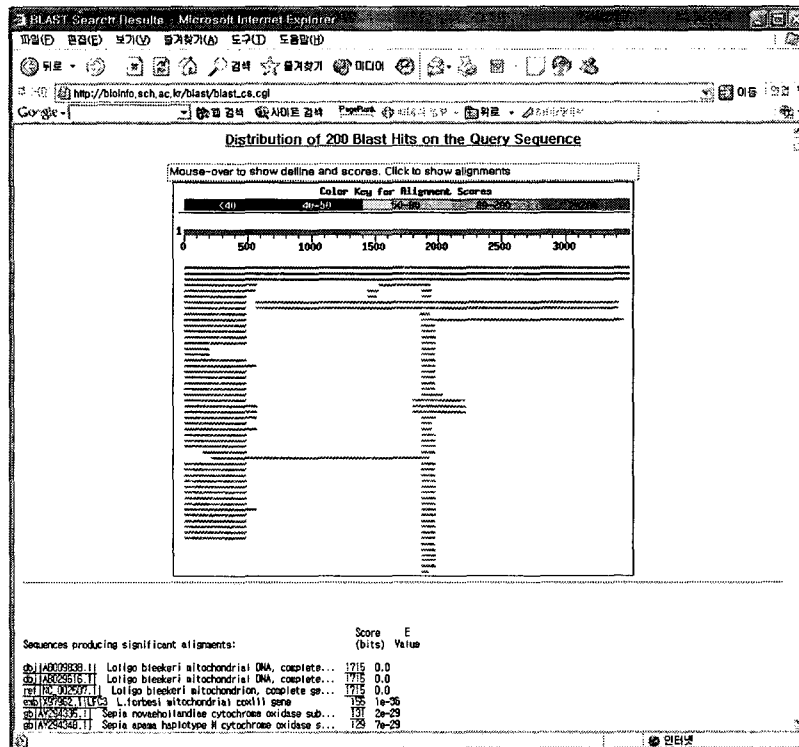


그림 2 ) 유전자 서열을 BLAST 한 결과

## 참고문헌 및 참고 웹사이트

1. Mount, David W, 2001 Bioinformatics : Sequence and Genome Analysis, CSHL Press.
2. NCBI, 2002. NCBI hand book, National Library of Medicine,
3. <http://ncbi.nih.gov>