

메타게놈 분석 시스템 설계 및 구현 Implementation of MetaGenome Analysis System

김도완, 최한석*, 김동욱**

농업과학원 농업생명자원부, 목포대학교 컴퓨터공학*,
(주)올비티**

Do-Wan Kim, Han Suk Choi*, Dong-Wook Kim**

National Institute of Agricultural Science,
Department of Computer Engineering Mokpo
National University*, AllBTco, Ltd**

요약

본 논문은 유전체 분석 연구에서 국내 최초 메타게놈 Data를 윈도우상에서 효율적으로 저장하고 분석할 수 있는 시스템으로서 향후 국내외 유전체 시장에서 메타게놈 Data 분석 및 저장을 선도할 수 있는 시스템이 될 것이다. 또한, 개발된 메타게놈 유전체 분석 시스템을 이용한다면 국내외 메타게놈 유전체 연구에 많은 도움이 될 것이다.

I. 서론

메타게놈(metagenome)이란 용어는 Handelsman 등에 의해서 “특정 자연환경에 존재하는 모든 미생물의 유전체 집합”으로 정의되고 있다. 그러나 최근 환경시료로부터 추출한 유전체 또는 유전자를 포함하는 클론을 총칭하여 메타게놈이라고 부르고 있으며 이러한 메타게놈에 관련된 일련의 연구를 메타게노믹스(metagenomics)라고 부르기도 한다. 학자들은 자연계에 존재하는 대부분의 미생물들이 왜 실험실에서 인공적으로 분리되거나 배양되기 어려운가에 대한 몇 가지 가설들을 내놓고 있다. 이러한 가설들 중 가장 중요한 이유로 꼽히는 것은 전통적인 평판배양방법으로는 미생물들이 존재하는 원래의 자연환경과 유사한 조건을 제공해 주기가 어렵기 때문에 대부분의 미생물들이 실험실에서 배양이 되지 않는다는 것이다. 따라서 자연계에 존재하는 대다수 미생물들의 다양성과 이들의 생태학적인 역할에 대한 올바른 이해와 함께 이들의 효과적인 활용을 위해서 다른 거시적인 접근방법이 필요하게 되었다. 이런 중에 도입되기 시작한 방법이 메타게놈 연구로 비교적 연구의 역사는 길지 않은 편이지만 최근 거시적인 연구 성과들이 발표되고 있다.

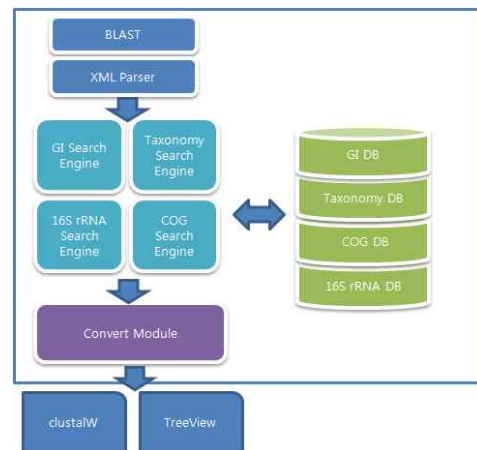
이러한 메타게놈을 분석하기 위해서 이미 다양한 분석 프로그램이 나와 있다. 그러나 일반 메타게놈 분석학자들이 쉽게 사용할 수 있는 파이프라인 형태의 프로그램은 전무한 실정이다. 또한, 이러한 분석 파이프라인을 설치한다고 하여도 컴퓨터에 깊은 지식이 없는 사용자들은 사용하기가 쉽지 않은 실정이다[1][2].

이런 문제점들을 해결하고자 많은 사용자들이 쉽게 설치하고 사용할 수 있는 윈도우 기반 분석 파이프라인 시스템을 설계하고 구현하였다.

II. 시스템 설계 및 구현

1. 시스템 설계

본 논문에서는 META Genome 분석을 윈도우 환경에서 사용하기 쉽도록 사용자관점에서 설계하였고 인터페이스 및 뷰어를 구현하였다.

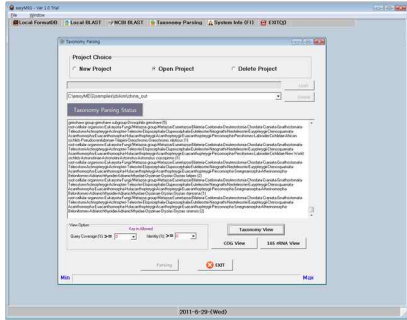


▶▶ 그림 1. 시스템 구조

easyMEG 시스템은 <그림 1>과 같이 sequence 입력을 받아 NCBI DB 또는 로컬 DB를 통해 BLAST 검색 후 자동으로 분석되어 메타게놈 분석에 필요한 Taxonomy Tree, COG검색 16S rRNA검색등이 자동으로 이루어질 수 있도록 설계되었다.

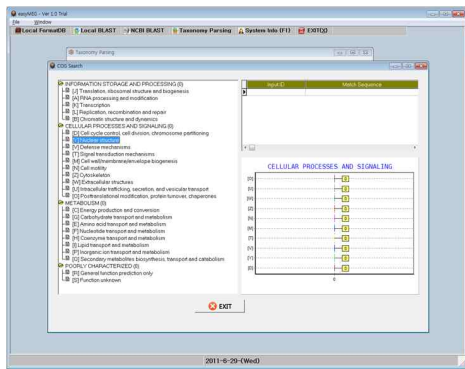
2. 시스템 구현

easyMEG는 delphi7.0을 이용하여 구현하였으며 윈도우 환경에서 실행 되도록 개발하였다. 또한, 복잡한 프로그램 셋팅을 최소화 하기위하여 윈도우 기반 무료 DB인 MS사의 AccessDB를 활용하여 구현 되었다.



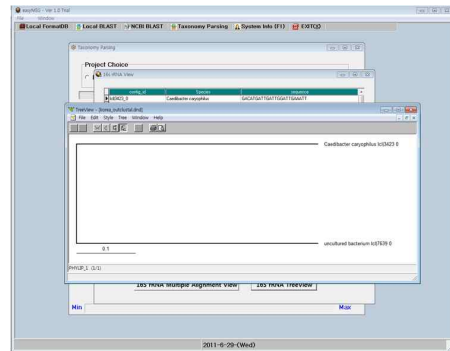
▶▶ 그림 2. project 관리 및 Taxonomy tree 분석화면

〈그림 2〉는 easyMEG의 project 관리 및 Taxonomy tree 분석 과정을 보여주고 있다. BLAST결과를 자동으로 파싱하여 시스템 자체에 내장되어있는 Taxonomy DB를 검색하여 Tree를 보여주고 있다. 상세 분석을 위하여 분석옵션을 선택하게 할 수 있는 기능이 구현되어 있다. 한번 분석된 데이터는 자동으로 저장되어 향후 분석을 위하여 재 로드할 때는 추가적인 분석과정 없이 실시간으로 분석자료를 확인 할 수 있도록 구현 되었으며 저장된 project는 삭제 가능 하도록 하여 필요 없는 자료가 공간을 차지하지 않도록 설계되었다.



▶▶ 그림 3. COG 분석 화면

〈그림 3〉은 검색된 Taxonomy를 이용하여 COG DB를 검색하여 분석자료를 보여주고 있다. 사용자가 왼쪽의 Tree를 선택 하면 연관된 시퀀스와 각 모듈별 결과를 상세하게 보여줄 수 있도록 구현 되었다.



▶▶ 그림 4. 16S rRNA를 이용한 계통도 분석화면

본 시스템은 〈그림 4〉와 같이 BLAST 파싱 결과 중 16S rRNA 관련 시퀀스를 자동저장하여 clustalW로 계통도를 자동분석 하게 하였으며 이 결과를 Treeview 프로그램을 이용하여 쉽게 확인 할 수 있도록 구현 되었다.

III. 결론

구현된 시스템은 메타게놈 분석을 위한 필수 패키지를 사용자가 사용하기 편리한 window PC기반으로 구현하였으며 네트워크의 연결 없이 자체 내장된 DB만으로도 분석이 가능하며 설치 및 관리가 편리하도록 윈도우 인스톨 프로그램을 사용하여 설치 할 수 있도록 구성되었다. 본 시스템을 사용하면 컴퓨터에 많은 지식이 없는 연구자도 쉽게 메타게놈 연구를 진행 할 수 있을 것이다.

본 성과물은(논문, 산업재산권, 품종보호권 등)은 **농촌진흥청 포스트게놈다부처유전체연구사업 (PJ01369302018)의 지원에 의해 이루어진 것임**

■ 참고 문헌 ■

- [1] Liu F, Chen P, Cui SJ, Wang ZQ, Han ZG: SJPdb: integrated transcriptome and proteome database and analysis platform for Schistosoma japonicum, BMC Genomics 2008, 9:304.
- [2] Aurecochea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, et al: PlasmoDB: a functional genomic database for malaria parasites, Nucleic Acids Res 2009, 37:D539-543.