

## 생물정보학(Bioinformatics) II - 해외동향

미국, 유럽선진국, 일본의 생물정보학에 대한 동향을 살펴봄으로서 미래 한국의 생물정보학의 방향을 알아보고자 한다.

### <미국의 생물정보학>

생물과 생명의 특징을 결정짓는 것은 바로 DNA 염기서열이다라는 사실을 학자들간에 널리 인정하게 되면서 이에 대한 연구는 가속이 붙어져 갔으며, 인간 유전체사업을 수행하면서 발생하는 방대한 정보의 처리를 위해 생물정보학의 필요성은 본격적으로 강조되어 갔다.

1986년 7월 하워드 휴즈 의학연구소에 의해 국립보건원내에서 인간 유전체에 관한 최초의 국제회의가 이루어지고, 1990년 미국 에너지성과 국립보건연구소에 의해 인간유전체사업의 발족되면서 인간 유전체사업은 본격적으로 가속이 붙게되어 2000년 6월말 성공적으로 인간 유전체의 초안이 작성되었다. 이는 16개 컨소시엄 참가 연구소로부터의 전체보고서로 이중 미국 단독으로 8개 이상의 연구소 (텍사스 휴스턴 Baylor 의과대학, 메사츄세츠 지놈치료법인, 캘리포니아 US에너지국 연합유전체연구소, 스탠포드 DNA 시퀀싱 기술개발센터, 시애틀 워싱턴대학 유전체센터, 멀티메가베이스 시퀀싱 센터, MIT 바이오메디컬 Whitehead 연구소, 미주리 워싱턴 대학 유전체 시퀀싱센터)가 참여하였다. 이에 1988년에 분자생물정보에 관한 미국의 국가적 자원으로써 설립된 National Center for Biotechnology Information (NCBI)는 1) 분자생물학·생화학·유전학에 대한 지식의 저장·분석하는 자동화된 시스템 생산, 2) 생물학적으로 중요한 분자들의 구조와 기능을 분석하기 위한 컴퓨터에 기반하는 정보를 다루는 발전된 방법 연구, 3) 데이터베이스와 소프트웨어사용 지원, 4) 전세계 생명공학정보 취합 등의 주요역할을 수행하며, 다양한 의료·과학단체에 생성한 데이터베이스들을 제공·지원하고 있다.

NCBI가 제공하는 데이터베이스 및 프로그램으로는 Online Mendelian Inheritance in Man (OMIM), 3D 단백질 구조의 Molecular Modeling Database (MMDB), Unique Human Gene Sequence Collection (UniGene), 인간의 유전자 지도, 데이터베이스로부터의 정보를 통합하는 검색·복구 시스템인 Entrez, 유전자와 단백질 데이터베이스의 분석을 지원하기 위해 디자인된 동일 서열 검색 프로그램으로 유전자 특징을 식별하는데 도움을 주는 BLAST 등이 있으며 그림. 1에 생물정보학과 관련된 데이터의 흐름도를 나타내었다. 특히, Entrezsms 사용자들에게 데이터의 순서, mapping 분류법, 구조의 통합적인 접근을 제공하며 서열과 염색체 지도를 그래픽으로 보여준다.

지금까지의 인간유전체 사업이 유전체의 구조를 밝히는 구조유전학의 연구였다면, 앞으로의 연구는 그 유전체의 기능을 밝히는 기능성 유전학(functional

genomics)에 집중될 전망이다. 이에 대한 미국소재 유전체 기관을 살펴보면, 로스웰 파크 암연구소내에 있는 BacPac Resource Center, Research Genetics사, Celera Genomics사를 들 수 있다. 이중 Celera Genomics사는 1998년 5월에 설립된 생명공학 전문기업으로 미국의 민간연구업체 중 대표적으로 꼽을 수 있다. 이들은 인간·쥐·초파리·쌀·Arabidopsis 등의 유전체연구를 통하여 미래의 의약과 농생식물의 기반을 창출하는 것을 목표로 하며, 1년동안 작업한 결과를 얻게 된 데이터의 양은 무려 15TB에 이른다. 또한 Celera사의 시설은 일반 유전체 회사보다 10 내지 20개 이상을 보유하고 있으며 세계전반의 유전체 관련시설보다 더 많은 우수한 시설을 보유하고 있다. 특히, 세계에서 두 번째로 강력한 슈퍼컴퓨터를 보유하고 있음은 괄목할 만한 사실이다.

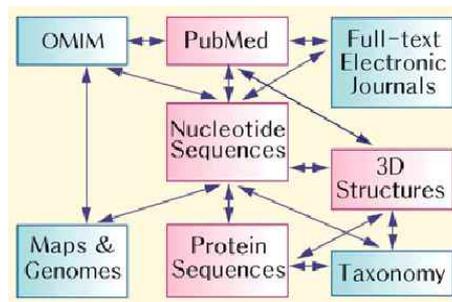


그림.1 생물정보학에 관련된 데이터의 흐름도

### <유럽의 생물정보학>

대표적으로 유럽의 생물정보학을 지원하는 European Molecular Biology Laboratory (EMBL)은 1974년 설립되어 유럽의 16개국으로부터 지원을 받고 있다. 이들의 주된 역할은 1) 분자생물학의 기초 연구, 2) 회원국의 과학자들에게 생물정보학 서비스의 제공, 3) 기능적 훈련, 4) 생물학 연구의 새로운 방법 모색 등을 들 수 있다.

European Bioinformatics Institute (EBI)는 EMBL의 연구 지부로 생물정보학에 대한 연구·서비스를 담당하는 기관으로 핵산, 단백질 서열, 거대 분자의 구조에 대한 생물학적 데이터베이스를 관리한다. EBI에서 제공되는 데이터 베이스와 분석도구로는 FASTA, BLAST, CLUSTALW 등과 3차원상에서 단백질 구조를 비교하고 생물학적으로 관심있는 상동성을 보여주는 DALI라는 도구, GeneQuiz라는 단백질 서열에 대해 생화학적 기능을 예측하는 자동화된 분석 시스템, 텍스트 정보의 데이터 베이스를 indexing cross-referencing하는 프로그램은 SRS 등이 있다. 또한 EBI에서 개발하여 제공하는 분자생물학 정보에 관한 주된 데이터베이스는 EMBL Nucleotide Sequence Database, TrEMBL, SWISS-PROT protein sequence database, 생물학적 거대

분자의 3차원적 구조 좌표 데이터베이스인 Macromolecular Structure Database (EBI-MSD), 방사선 지도에 관한 Radiation Hybrid Database (RHdb) 등이 있다. 그림. 2에 연도별 핵산 데이터베이스 증가와 생물종에 대한 데이터베이스 수의 비교를 나타내었다.

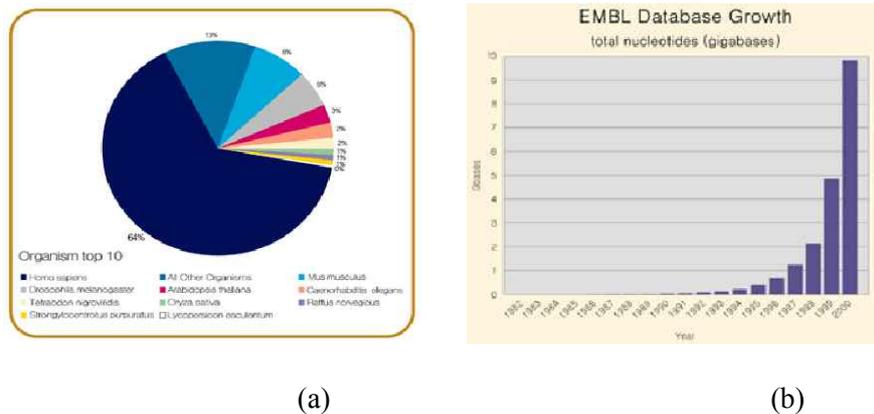


그림. 2에 생물종에 대한 데이터베이스 수의 비교(a)와 연도별 핵산 데이터베이스 증가(b)

스위스의 SWISSPROT 단백질 서열 데이터베이스는 1986년 제네바 대한 Swiss Institute for Bioinformatics (SIB)에서 시작되어 1987년 이후에는 EBI와 공동 작업으로 구축되었다. ExPASy (Expert Protein Analysis System) proteomics server는 SIB에서 단백질 서열 분석과 구조 분석을 위한 서버이며, 단백질의 2D, 3D에 대한 서비스 뿐만 아니라 효소 명명법에 관한 데이터베이스와 서열분석을 위한 서지 참조 데이터베이스도 제공하고 있다.

### <일본의 생물정보학>

DNA Data Bank of Japan (DDBJ)은 유럽의 EBI/EMBL 및 미국의 NCBI/GenBank와 밀접하게 연계하여 ‘DDBJ/EMBL/GenBank 국제염기배열데이터베이스’를 구축하고 있는 3대 DNA 데이터베이스의 하나로 일본의 국립유전학연구소 생명정보연구센터 내에 설치되어 있으며 다음과 같은 역할을 수행하고 있다.

1. 국제DNA 데이터베이스 공동 구축 및 운영  
; 국제DNA 데이터베이스 공동 구축도를 그림 3.에 나타내었다. 전세계의 연구자들이 실험으로 결정한 DNA(또는 RNA) 염기배열데이터를 DDBJ/EMBL/GenBank 국제염기배열데이터뱅크가 결정한 데이터 구축 규범에 따라 수집·가공하여 컴퓨터파일의 형태로 제공한다.
2. 소프트웨어의 개발 및 홍보활동
3. 국립유전학연구소 컴퓨터시스템의 관리 및 운영

#### 4. DNA 데이터베이스의 온라인 이용

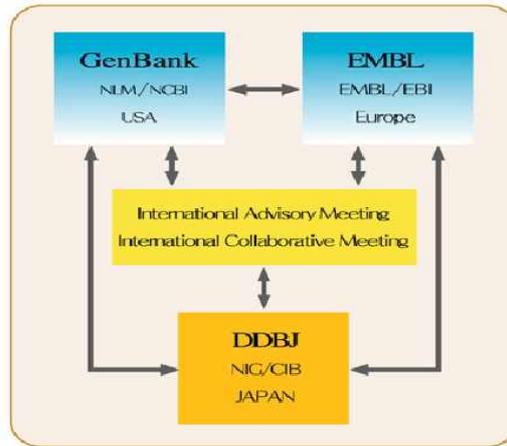


그림. 3 국제염기배열데이터베이스 구축도

일본의 생명정보연구센터(CIB: Center for Information Biology)는 시즈오카현 미시마시에 있는 국립유전학연구소 부설로 1995년에 설립되었으며, 유전정보분석 연구실, 유전자기능 연구실, 대량유전정보 연구실, 분자분류 연구실의 4개로 연구실로 구성되어 국제염기배열데이터베이스의 하나인 일본 DNA 데이터뱅크를 운영하고 있다. 이곳에서의 주요 목표는 컴퓨터를 사용한 데이터 해석을 위한 환경과 인력을 갖추고 생물정보학 활동에서 중요한 역할을 담당하는데 있다.

유전정보분석 연구실에서는 컴퓨터를 사용하여 DNA 염기배열이나 아미노산 배열의 배열을 해석하여 유전자나 생물의 진화를 해명하는 것을 목표로 분자진화학, 집단유전학, 유전체데이터의 해석 및 실험에 의한 배열 결정 등을 한다. 대량유전정보 연구실에서는 유전자의 산물인 단백질의 해석을 주로 컴퓨터를 이용하여 수행하고 있다. 아미노산배열데이터를 컴퓨터에 입력하고 계산하여 단백질의 입체구조를 예측하기란 쉽지 않았으나 최근 데이터베이스에 구축된 배열이나 입체구조의 정보를 구사하여 이 문제를 해결하는 방법인 3D-1D법 (structure(3D)-sequence(1D) compatibility evaluation)을 고안하여 몇몇 단백질에 대해 성공한 사례가 있다. 분자분류 연구실에서는 다양한 생명현상과 생물을 다각적인 관점에서 분류하는 방법을 연구 개발하여 생물다양성(Biodiversity)의 본질에 다가서는 것을 목표로 분자진화 이론에 기초한 진화계통분석이나 통계학에 기초한 수치분류 혹은 가시화 방법 등을 이용하여 연구를 진행하고 있다.

생물정보학에의 관심과 지원은 국가·민족 전체의 생사가 달린 문제로 이미 우리는 유전체의 종속의 위협 속에 살고 있는 것이다. 이런 점에서 생물정보학의 발전을 위한 국가적인 차원의 노력이 필요하다고 사료된다.