

# 국내 Bioinformatics 발전을 위한 전산학의 역할

허철 구(herk@hankyong.ac.kr, hcb@h.y.ac.kr)  
생명공학연구소 선임연구원

## I. 서론

### II. 주요 연구·개발 분야

### III. 국내 생물정보 발전을 위한 대책

## IV. 결론

## I. 서론

인간 유전체에 대한 전체 서열분석이 되는 시점을 거기로 국내외적으로 기초연구에 대한 논의가 본격화됨과 동시에 국내에서는 인문을 통한 생물정보에 대한 관심이 학계를 중심으로 매우 높아지고 있다. 사실 생물정보(Bioinformatics)는 미국, 유럽을 중심으로 한 1990년 인간 유전체 연구를 시작으로 각종 생물에 대한 기초 연구가 본격화되면서 육성되어 왔다. 이에 따라 염기서열과 단백질 서열 등 각종 생물학 연구 결과들이 생산되면서 이를 관리하고, 체계적으로 분석할 수 있는 프로그램을 개발하여 왔다. 미국의 경우 주로 10여년 전부터 염기서열 분석 관련 SW개발을 주도하여 왔고, 1990년대 중반부터는 DNA Chip이라는 획기적인 유전자 분석 기법이 나오면서 본격적인 전산학의 자질을 필요로 하는 생물정보 분야의 새로운 장을 열었다. 생물정보학을 단순히 정의하자면은 매우 어려운 분야인데 이것은 생명공학 연구 분야가 다양해지고 생물정보학 자체도 너무 방대하기 때문에 포괄적인 지식을 요구하는 특수한 분야라고 판단하기 때문이다. 즉, 수학을 바탕으로 통계학, 전산학, 생물학 등 자연과학의 거의 모든 분야를 필요로 하는 융합학문과 같이 느껴지기 때문이다. 이런 특수성 때문에 인력 유성이 쉽지 않은 것은 다른 나라도 사정은 마찬가지라고 생각된다. 그러나 미국과 유럽 및 일본 같은 경우 20여년이 걸쳐 꾸준히 투자를 해 온 결과 오늘날 국내에서

합계 연구·개발하기 어려운 부분도 상당하다. 하지만 21세기 생명공학을 주도할 생물정보에 대한 투자와 연구를 지금부터 소홀히 한다면 경쟁력 상실은 당연한 것이나 다름없다. 다르게 생각하면 우리 나라만큼 훌륭한 생물정보학에 투자할 여건이 좋은 국가도 없을 것이다. 전산관련 분야에 대한 투자가 매우 많이 되어 있고, 생명공학관련 전문 인력이 많이 있는 여건을 감안한다면 아직부터라도 체계적으로 접근하여 선진국을 따라잡을 수 있는 기틀을 마련하여야 한다고 생각되는 것은 비단 혼자만의 생각은 아닐 것이다.

그러나 국내에서 생물정보에 대한 발전을 이루어민 선명되어야 하는 몇 가지 과제와 현재 우리가 안고 있는 문제점을 파악하고 해결책이 제시되어야 한다. 이를 위해서 해외 기술 동향과 국내 현황 그리고 주요 연구·개발 분야, 문제점 및 대책에 관하여 나름대로 논와하고자 한다.

## II. 주요 연구·개발 분야

### 1. 해외 연구기관 현황

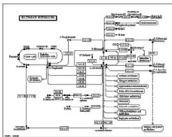
미국을 중심으로 유럽, 일본 등에서 투자되고 있는 정보 분야는 전인 유전체 연구 기관과 대학별로 너무 방대하고 그 범위가 넓으므로 대표적인 기관의 사례만 들어보기로 한다. 우선 미국의 NCI를 살펴보면 선진국의 수준을 짐작할 수 있다.





■ 그림 3 스위스 생물정보연구소의 ExPASy 시스템 서비스

물정보 연구 경험을 살려 생물정보 합동 교육 과정을 실시하였으며, 경북대학교, 충북대학교, 건국대학교, 세종대학교, 한림대학교, 서울대학교 등에서 활발하게 진행중이다. 이외에도 상당수 대학교에서 관심을 가지고 새로운 분야를 개척하기 위한 노력이 이루어지고 있다. 이는 매우 고무적이고 바람직한 현상이라고 생각된다. 하지만 생명공학분야 발전의 추세에 비하면 아직 턱없이 부족하다. 이런 현실에서 나타나는 문제점은 생물학 관련 연구자들이 생물 프로그램 위주로 연구비를 투입하는 경향이다. 일부 필요한 프로그램은 얻을 수 없이 외국 상용 제품으



■ 그림 4 KEGG 시스템에서 제공하는 정보의 예

로 운영하지만, 시스템 구축에 중요한 분석용 프로그램은 국내에서 안전한 기술을 확보하여야 한다. 국내에서 개발될 시점까지 실험자들은 상용 프로그램을 구입하여 부분적으로 활용하고, 전산 관련 학과를 중심으로 시간을 두고 국내에서 개발할 수 있는 토양을 마련하여야 한다.

이런 과정을 거치지 않고서는 생물정보 관련 경쟁력이 완전히 상실되는 사태가 오는 것은 명백하기 때문에 국내 몇몇 되지 않는 생물정보 전문가들이 끊임없이 노력하고 있다. 국내에서는 안전한 기술을 계속 확보하기 위해서 전산관련 연구자들의 관심이 촉구되는 현실에서이다.

### 3. 생물정보학 주요 분야

생물정보 분야를 크게 나누어 보면 서열 비교·분석 분야, 데이터베이스 분야, 단백질 기능 예측 분야, DNA Chip 관련 정보 분야, SNP 분야, 프로테오믹스 분야 등으로 최근 급변하는 추세이다. 아무리 생물정보 분야가 발달하더라도 가장 기본이 되는 분야는 역시 서열 비교 분야이다. 이는 서열 비교를 통한 상동성 검색으로서 생물학적으로 유사 기능을 추정하기 위한 중요한 분석을 위한 기본 분야이다. 이가운데 다중 정렬에 관한 알고리즘과 프로그램은 그리 쉬운 일이 아닌 고난도의 기술을 요구하여, 아직 표준화된 방법은 없고 개발된 도구가 갈수록 각각 다르게 나오기 때문에 효율성 면에서도 계속 연구되어야 하는 분야이다. 이런 비교 정렬 알고리즘을 바탕으로 계통도 분석을 위한 도구도 개발되고 있으며 고성능 컴퓨터를 사용하는 생물학의 주요 분석 방법이다. 단백질 서열 비교는 물론 Metabolics을 위한 알고리즘 개발 분야는 매우 중요한 분야에 속한다. 유전자의 기능을 밝히기 위하여 단백질 서열을 알고 난 후 기능을 알아보는 데 있어 매우 중요한 역할을 하기 때문이다. 다만, 실험적으로 증명된 기능이 아닌 예측된 결과이기 때문에 반드시 실험을 거쳐야 하는 과정은 필수적이다. 즉, 생물정보학 자체가 실험자들의 실험

**비교·분석 분야 데이터베이스 분야 단백질 기능 예측 분야 DNA Chip 관련 정보 분야 SNP 분야 프로테오믹스 분야 등으로 최근 급변하는 추세이다**

화수를 줄여주고 예측 시스템인 한 정확도를 높여주는 데 큰 목적이 있는 점을 시사하는 것이다. 암기 서열에 관한 분야도 방대하고 단락질을 분석하기 위한 방법은 또한 매우 넓은 범위에 걸쳐 필요하다. 최근 DNA Chip과 SNP분야 프로테오믹스 분야가 대두되면서 점차 종합적이고 체계적인 시스템 구축이 필요한 방향으로 정보의 흐름이 바뀌고 있다. cDNA를 심고 분석하는 DNA Chip의 경우를 살펴보면 cDNA 서열 판독을 위한 데이터베이스 관리, cDNA 가장 긴 서열 분석 및 기능 유추, 각종 인타페이스 개발, 이미지 분석 등 생물정보학의 종합학문이 되고 있고, SNP의 경우나 프로테오믹스 분야의 경우도 기존의 생물정보학의 전분야를 필요로 한다.

### III. 국내 생물정보 발전을 위한 대책

#### 1. 생명공학 분야 연구, 개발에 따른 생물정보 인력 부족

최근 1999년 인건유전자연구단이 발족되

고, 삼성, LG, SK, 제일제당을 비롯한 대기업의 신규 투자 및 (주)바이오나이를 중심으로 한 중소기업과 시론은 한층 기업들이 계속 탄생하고 있다. 이런 추세는 당분간 계속 될 것으로 보이며 이와 함께 매우 많은 생물학 관련 데이터가 생성될 것으로 예상된다. 이는 곧

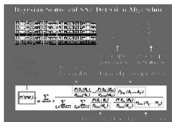
생물정보를 이해하고 연구·개발할 수 있는 전문 인력 수요가 증대되게 될 것이며 향후 2~3년후 심각한 연구·개발 인력의 불균형 현상을 초래하게 될 것으로 보인다. 이는 생명공학 분야의 전문 인력과 발전에 따른 정보 인력의 절대 부족을 의미한다. 예를 들어보면 국내 어느 연구 분야에서 대량 유전자 발현 사업을 시작한다고 가정하자. 우선 EST 단편 서열을 인간 6만건 정도 생산한다고 했을 때 정보 관련 업무를 파악하여 보면 전문 인력의 수요를 알 수 있다. 첫번째, EST 서열을 Clone 번호와 일치시켜 저장한 1차 데이터베이스를 구축하는 일이다.

이 데이터베이스 같은 경우 단순 FASTA 형식으로 작성한 구조와 서열들을 NCBI의 BLAST를 이용하여 나온 결과에서 필요한 정보를 추출한 자료를 저장하는 구조, 각종 Library, Clone 등을 연결한 정보 등이 필요하다. 간단하게 보아도 데이터베이스 전문기 및 웹을

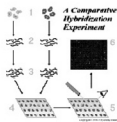
통한 자료 관리를 위한 전문 프로그램머들이 우선 투입되어야 제대로 관리할 수 있음을 알 수 있다. 두 번째, EST 단편

을 가지고 긴 cDNA를 만들어 유전자의 기능을 분석한다고 했을 때 필요한 일이다. 우선 *in silico* cloning 작업을 수행하는 것이다. 즉, EST Blast 차의 절차가 필요하다. 우선 EST 서열을 가지고 BLAST 작업을 수행하고 서로 연결하는 작업을 통하여 가끔씩 긴 EST를

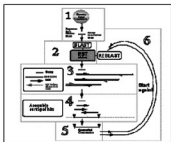
**최근 1999년 인건유전자연구단이 발족되고, 삼성, LG, SK, 제일제당을 비롯한 대기업의 신규투자 및 (주)바이오나이를 중심으로 한...**



■ 그림 5 DNAChip을 이용한 유전자분석 화면



■ 그림 6 SNP분석 방법(Washington University 자료 제공)



■그림 7 ESTBlast 절차 (Glaxo Wellcome사 제공)

만들어 consensus 서열을 찾아낸 다음 reblast 작업을 반복 수행한다. 세 번째, 알린의 의미 있는 긴 EST가 만들어지면 다시 EST Clustering 작업을 통하여 중복

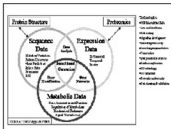
**데이터베이스 전문가, 인터페이스 개발자, 유전자 예측 시스템을 운영하면서 관련 프로그램을 개발할 수 있는 생물정보 전문가 등 하나의 시스템을 개발·운영할 수 있는 일련의 전문가들이 필요하다**

된 후보 유전자를 재차하는 작업을 통하여 중요한 유전자를 결정한다. 네 번째, 이런 작업의 결과를 DB화시키고 단락질 서열과 비교 분석을 통하여 기능을 추정하는 작업을 수행하여야 한다. 이후 DNA Chip 관련 작업에 들어갈 경우 Clustering 유전자 발현 분석을 할 수 있는 시스템 개발 등 정보 전문가들이 하여야 할 역할이 크다. 위에서 살펴본 바와 같이 데이터베이스 전문가, 인터페이스 개발자, 유전자 예측 시스템을 운영하면서 관련 프로그램을 개발할 수 있는 생물정보 전문가 등 하나의 시스템을 개발·운영할 수 있는 일련의 전문가들이 필요하다. 이런 전문가 육성이 그동안 국내 전산학이나 생명공학관련 학계에서 체계적으로 이루어지지 못한 점은 아쉽지만(이는 데이터 생산의 한계 때문여 생긴 상황이지만) 아래부터는 국내에서도 상황이 많이 변할 것으로 기대 된다(그림 7참조).

2. 전산학과 생물학 분야의 공동 연구

현재 생물정보학을 전공하기 위하여 기본적으로 확률론, 통계학 등이 필요하고 Programming, Data structure, Algorithm, 데이터베이스 등 전산학의 기본적인 분야가 필수적이다. 또 Pairwise alignment, Multialignment, Phylogeny analysis tool, Motif analysis, Gene prediction, Homology search 등 기본적인 서열 분석 부분 이외에도 유전자 기능 분석 분야가 세로이 등장하면서 DNA Chip 관련 아미지 분석 기술, 유전자 발현 분석 기술, SNP 분석 분야, 프로테옴 분야, 유전자 비교분석 분야 등 방대하고 새로운 분야가 계속 생기고 있다. 이는 전산 분야만 가지고 접근할 수 있는 분야도 중요하지만 상호 있는 생물학 지식을 접목시킨 분야가 갈수록 요구되고 있다는 것을 의미하기도 한다. 순수 전산학 분야도 아니고 순수 생물학 분야도 아닌 과학 등을 이수하기 위해서는 반드시

기본적인 전산 지식과 생물학 지식을 겸비한 인력이 절실히 필요한 것이다. 그러나 실재적으로 전 분야를 접목시킨 전문 인력을 짧은 시간에 육성하면 사실상 힘



■그림 8 생명과학분야와 연구개발에 필요한 전산분야

들다. 하지만 전산학과 생물학적인 학과간에 협동 과정을 개설하고 학업 교류를 통한 공동 연구를 꾸준히 진행한다면 장기적으로 매우 수준 높은 인력이 배출 될 수 있을 것으로 기대된다. 이 외에 전체 조건의 현재 각 전공별로 나와야 할 분야를 어느 정도 온합한 형태의 교육 과정을 허용하는 교육 시스템을 개발하고 전문 교수수를 시급히 발굴하여 체계적인 교육을 시작해야 한다는 점이다. 이런 인재 육성을 체계적으로 시뮬러 육성하지 않을 경우 다가오는 생명공학 시대에 국가 경쟁력은 떨어질 수밖에 없다. 또 생명공학분야 데이터가 계속 많아지는 추세를 감안한다면 이런 생물학을 전공하고 분석 프로그램을 활용하여 전문적으로 데이터를 분석할 수 있는 인력 또한 필요한 것도 사실이다(그림 6 참조).

### 3. 전문 연구기관 필요성

최근 국내에서 가장 많은 인력을 배출한 기관은 생명

학을 선두로 각 정부 부처마다 특성적인 생물정보에 대한 투자 계획을 하고 있어서 향후 2~3년 후에는 국가 경쟁력을 갖춘 전문 기관이 탄생할 것으로 예상된다. 이런 중요한 시점에 연구개발정보센터(소장: 포영화) 박사와 같은 슈퍼컴퓨터와 연구전선망을 갖춘 기관에서 대학과 연계하여 국가 공동 전문 기관으로 운영을 하고, 각 부처에서 실행한 정보 선진화는, 특성화된 형태의 기관 운영이 된다면 가까운 장래에 무리나 해도 선진국과 경쟁할 수 있는 기초가 마련될 것으로 판단된다.

## IV. 결 론

앞서 논의한 바와 같이 생물정보학의 중요성과 시급성은 이제 강조할 사항이 아닌 본격적인 연구 개발을 시작하여야 할 때다. 우수한 전산 전문 인력과 생명공

**연구개발정보센터에서 슈퍼컴퓨터와 연구전선망을 갖추고 생명공학관련 정보를 집대성하여 서비스할 때, 각 대학별로 특성 있는 정보를 꾸준히 개발하여 간다면 일본의 KEGG 시스템과 스위스의 ExPASy 시스템 같은 우수한 기관이 탄생할 것이다.**

공학연구소(소장: 복성태) 박사와 모함금대 생물학적인 연구센터(연다장: 남승길) 교수이다. 두 기관에서 생물정보를 시작할 당시만 해도 크게 호응을 받지 못한 상황이었다. 그 시급성이나 중요성에 비추어볼 때 위에서 언급한 바와 같이 데이터 생산량이나 관련 연구를 큰 단위로 수행하는 기관이 필요 있었던 상황을 이해하여야 한다. 하지만 현재 국내에서도 중요성을 깊이 인식하고 있는 만큼 조속한 시일 내에 체계적으로 연구 개발할 수 있는 기관이 설립될 수 있을 것으로 기대된다. 이런 전문 연구 기관이 설립되어야 대학교에서 인력 육성에 관한 기구제로 작용할 수 있기 때문이다. 다만, 분야가 너무 방대한 만큼 각 기관마다 전문화, 특성화된 모습으로 설립되기를 간절히 바란다. 물론 상호 유기적인 결합이 가능한 형태가 되어야 하는 것은 당연한 모습이라 하겠다. 다양한 제작 기술부의 존재가 사

학적인 인재들이 모여서 뜻을 같이 한다면 앞으로 국내 생물정보 발전은 매우 고무적이 되고 판단된다. 이런 시각에서 본다면 연구개발정보센터에서 슈퍼컴퓨터와 연구전선망을 갖추고 생명공학관련 정보를 집대성하여 서비스를 하고, 각 대학별로 특성 있는 정보를 꾸준히 개발하여 간다면 일본의 KEGG 시스템과 스위스의 ExPASy 시스템 같은 우수한 기관이 탄생할 것이며 이를 바탕으로 생명공학 관련 기업의 경쟁력 또한 강화될 수 있다고 생각된다. 다가오는 생명공학 시대의 전산관련 학과의 집중적인 지원이 생명공학 분야에 기대된다.

