



바이오포매틱스부문 (생물 정보학)

한국의 생물정보학 연구와 KISTI의 역할



임용표 교수 **서울대학교 전염학교수**

yplim@cnu.ac.kr

- I. 유전체연구의 현황
- II. Bioinformatics의 중요성과 세계적 현황
- III. 생물정보학 분야의 국내환경과 현황
- IV. 우리나라 Bioinformatics의 문제점
- V. 생물정보학분야에서의 KISTI의 역할

21세기 들어 IT 기술의 급격한 발전을 통한 컴퓨터의 이용 확산은 생물학 분야에 대단한 영향을 미치게 되었고 이러한 현상은 생물정보학이라는 새로운 학문으로 탄생하게 되었다. 생물정보학은 유전자 및 단백질의 정보를 포함한 생물체의 모든 정보를 효율적으로 처리하기 위하여 유용한 정보를 창출해 내는 새로운 융합학문이다. 이러한 생물정보학은 특히 1990년대 이후 발전한 생명과학 및 유전체 관련 지식이 폭발적으로 늘어남에 따라 이러한 정보를 효과적으로 관리하기 위하여 급격하게 성장하게 되었으며, 이미 생물산업 분야의 핵심응용 기술로 인정받고 있다. 본고에서는 이러한 생물정보학의 국내외 현황과 흐름을 분석하였고 KISTI의 역할 및 앞으로의 방향에 대해 검토해 보고자 한다.

I. 유전체연구의 현황

생물정보학 분야의 우리나라 연구 현황을 살펴보기 위하여 우선적으로 국내외의 Genomics 연구현황을 살펴볼 필요가 있다고 생각한다. 특히 Human, mouse, Arabidopsis, rice 등 주요 생물을 상대로 한 유전체 사업의 효과는 대용량의 차원에서 유전정보 해석 및 유전자 획득이 가능하고, 연구 모델로서 다양한 organism의 유전체 연구를 시도할 수 있어 학문적인 획기적 발전을 도모할 수 있을 것이며, 인간 질병의 근원적 예방과 치료가 가능해지고, 식량 문제 해결 방안을 궁극적으로 제시할 것으로 기대되고 있다.

따라서 한국에서도 과거부 프론티어연구등을 통하여 생명공학연구원의 인간 유전체 기능 연구 사업단을 중심으로 한 인간의 위암/간암 관련 유전체 연구를 통하여 cDNA microarray와 프로테옴 분석을 통한 위암·간암 유전자 및 단백질의 대량 분석 및 한국인 특이 SNP 발굴 등의 연구를 수행하고 있으며, 식물의 경우 벼, 고추에 대한 연구를 작물 유전체 사업단에서, 인삼의 경우 자생식물사업단에서 연구를 시작하였다. 농진청을 중심으로 진행해오던 벼 유전체 연구는 일본과 함께 1번 임색체의 전 염기서열을 분석 완료한 바 있고, 중국에서 Indica type의 벼 전염기서열을 분석한 바 있으며, Monsanto와 Syngenta에서도 염기서열을 분



석한 바 있다. 또한 배추 유전체 연구는 충남대학교 유전체 연구 센터와 농촌진흥청 농업생명공학 연구원을 중심으로 세계 10여개국이 참여한 Multinational Brassica Genome Sequencing Project을 출범시켜 연구가 시작되고 있다.

II. Bioinformatics 의 중요성과 세계적 현황

인간유전체 사업이 착수된 1990년부터 이후 GenBank에 유전체 관련 데이터양의 기하급수적인 증가로 인하여 생물정보학 분야의 발전이 필요하게 되었고 이에 따라 생물정보관련 연구자들에 의해 고성능서열분석 시스템이 고안되고 이는 유전자 기능 예측 및 유전자 발급에 지대한 영향을 미치게 되었다.

이에 대한 대표적인 예로서 생물정보학의 가장 기초적이며 중요한 분야로서 첫째로 유전자 서열간의 상동성 검색 (homology searching) 을 위한 데이터베이스를 들 수 있다. 상동성 검색 프로그램 (FASTA, BLAST, Smith-Waterman search) 등을 이용한 DNA (nucleotide sequences) 및 단백질 (amino acid sequence) 데이터베이스 검색하고, 밝혀진 유전자 및 단백질에 대해 민감성(sensitivity)과 특이성(specificity)을 고려하여 서열간의 유사성 및 상동성을 추정한다. 통계적으로 의미있는 오차범위 내에서 빠른 속도로 유전자 및 단백질 검색이 가능하게 되었다. 둘째로는 인공지능 기법, 즉 신경망(neural networks, 예로 GRAIL)과 확률그래프모델(Markov model, 예로 GENSCAN)을 이용하여 유전자의 구조를 예측하는 예를 들수 있다. 이는 유전자뿐만 아니라 이들을 이용하여 단백질의 2차 구조의 예측이 가능하다.

외국의 경우 유전체 사업의 가속화와 산출물에 대한 산업화를 위해 Bioinformatics 분야에 많은 노력을 기울이게 되는데 그 예로 TGR (The Institute of Genome Research) 에 의해 주요 생물체 (미생물 17종, 진핵생물 14종)의 EST (Expressed Sequence Tags)의 대단위 발굴 및 분석 시스템과 데이터베이스 구축이 이루어 졌으며, Sanger Institute에서는 Human, mouse, worm 중심의

genome sequencing project 수행과 이들을 위한 분석 소프트웨어의 개발 (<http://www.sanger.ac.uk/Software/>) 및 데이터베이스 (ACEDB 및 Ensembl 등 유전체 정보의 상호참조(cross-referencing)가 가능하도록 한 객체지향의 데이터베이스)의 구축을 하였고, Stanford 대학의 TAIR(The Arabidopsis Information Resource)는 Arabidopsis 유전체에 대한 정보(map, sequence, 유전자 및 단백질(gene ontology 및 패턴 찾기 등), biochemical pathway)를 제공하고 있다. 특히 DNA 및 mutant line seed stock을 운영하여 유용유전자원을 보급하기도 한다.

기능유전체학 측면에서 보면, 1995년 미국 Stanford 대학의 Pat Brown 연구실에서 세계 최초로 분자생물학 및 기계공학, 컴퓨터공학의 기술을 접목하여 1cm²의 slide glass 공간에 약 2000개의 효모 유전자를 집적시킨 cDNA chip 개발에 성공한 이후, 보다 고밀도의 집적기술이 계속 개발되어 Affymetrix 사에서는 40만개의 oligonucleotide를 함유하는 DNA chip을 개발 및 시판 중에 있다. cDNA chip의 등장으로 인해 특이적으로 발현되는 유전자들 단 수 시간 내에 전체 유전자를 대상으로 초고속 탐색기술이 개발되어 게놈기능연구의 시대를 예고하고 있다.

이러한 게놈연구의 후속연구의 한 예로서 단일 염기의 변이 (SNP, single nucleotide polymorphism)에 의한 개체변이 분석 기술을 들 수 있는데 특히 인간게놈구조가 밝혀짐에 따라 염기의 개체변이에 큰 관심이 집중되고 있다. SNP는 특정 질병의 유전적 진단수단으로 사용되고 환자 개개인에 따른 최적화된 약물치료의 선택이 가능해 지리라 기대되고 있다. 미국의 경우 NIH에서 98년부터 3년간 100,000개의 SNP발굴을 목표로 3,000만 달러를 투자하여 2001년 초반 150만개의 SNP을 발굴한 것으로 보고한 바 있다. (<http://www.ncbi.nlm.nih.gov/SNP/>). 또한 일본의 경우 100,000-150,000개의 SNP발굴을 목표로 약 600억원을 투입 할 예정이다.

이러한 국내외 생물정보학 분야의 현황을 살펴보면 외국의 Bioinformatic는 생물학 및 유전체학과 상호

긴밀함 속에서 함께 발전
함을 알 수 있다.

III. 생물정보학 분 야의 국내한 경과 현황

국내 생물정보학
환경은 수요시장이
아직 형성되어 있지
않아 동기부여가 미약
할 뿐 아니라, 대부분의 연
구가 단순한 연구수준에 머물
러 있다.

또한 대량의 원시 데이터
를 생산, 저장, 관리하는 곳
이 적다는 것이 장애요인
중 하나이다. 국내 IT 기술이
세계수준임에도 불구하고 생
물정보학 관련기술은 열악한 상
태다.

국내 생물정보학 분야 전문가의 추정치를 살펴보면
질현유전자 초고속발굴분야 97명, 유전자기능분석분
야 105명, 신약개발분야 43명, 인프라분야 110명 정
도의 수준이며 이는 앞으로 국가차원에서 전문인력
양성을 위한 조치가 필요한 부분이다.

국내 생물정보학을 사실상 이끌어 왔고 이끌어 가
고 있는 주체가 바이오벤처라고 할 수 있다. 2000년
을 정점으로 IMF 시대를 맞이한 이후 정부의 적극적인
인 벤처 육성정책으로 생겨났다. 기술력으로 승부하
는 벤처의 이미지가 사실상 아주 잘 맞아떨어지는 데
마였기 때문에 많은 바이오 벤처들이 연구를 해오고
있다.

국내 바이오 벤처들은 최근 경기불안으로 무자가 줄
어들어 경영에 어려움을 겪고 있으나, 정부의 관심과
투자 그리고 벤처들의 적극적인 노력이 집중된다면
장족의 발전이 있을 것이다.

국내 대기업은 생물정보학 분야에 있어서는 철저하
게 방관자 역할을 했다고 할 수 있다. 기업의 육성상

제산성 문제를 우선적으로 검토해야
하지만 이 분야가 파급효과는 크지만
장기적인 투자가 이루어져야 한다
는 것, 산업화 아이템부족 그리고
연구인력확보의 어려움이 있어
투자를 미루고 있는 것으로 보
인다.

국내 주요기업들의 바이오 및
생물정보학 관련 투자계획을 보
면 삼성 2조원(2005년), 현대
1000억원(2002년), 포항제철
5300억원(2005년), LG 5000
억원(2005년), SK 4110억
원(장기플랜), 한화 500억
원, 녹십자 1200억원
(2005년) 등이다.

교육기관 및 연구 인력
을 살펴보면 현재 국내 생물
정보학 교육 시스템구축은 생
물정보학 활성화를 위해 필요한 인

프라 중 가장 중요한 것으로 범 국가적인 차
원에서 대책을 수립해야 할 것이다. 국내 생물정보학
교육관련 단체들을 보면 대학 5개, 대학원 6개, 정부
기관 2개, 기타 5개 등이다.

프로그램 개발현황을 보면 열악한 개발환경에서도
국내에서 벤처들에 의해 생물정보학 프로그램이 개
발되었고, 일부 제품은 국외로 판매되기도 하였다.
국내에서 개발된 주요 생물정보학 관련 유전체 분석
SW로는 연술테크의 Gene-Master, Hyper-Blast,
스몰소프트의 통합솔루션(미생물유전체분석),
아이디알의 IDXViewer(Gene analysis),
IDProTMPProStrATM(단백질 구조 분석)
IDPharoTM(단백질 3차원구조분석 및 예측), 바이
오인포메틱스 BASE series 정도일 것이다.

IV. 우리나라 Bioinformatics 의 문제점

우리나라의 생물정보학연구의 문제점을 살펴보면,
관중연 연구사업에 대한 Bioinformatics 연구에 초점





을 두게 됨으로서 Bioinformatic 에 대한 이해와 개발의 확장에 한계성을 주게 되었다. 예를 들면 Post-genom e를 주장하며 기능유전체학(cDNA chip, proteomics 중심)과 관련된 Bioinformatics 의 분위기 조성에 중점을 두었으므로 기초적이며 기반분야인 database 구축분야, 구조유전체분야에 대한 무자가 거의 이루어져 있지 않으며, 이로 인해 구조 및 기능유전체학의 상호적 발전을 저해시키고 있고, 실제 구조유전체학 연구는 국내에서 매우 열악한 기반을 가지고 있다.

또한 Biotechnology 연구자와 informatics 연구자간의 이해 부족으로 인해 국내 개발 저해 요인, 국내에서 개발한 것보다는 외국에서 개발한 public software 를 사용하게 되었으며, 대학, 연구소, 기업간의 공동 연구가 대단히 부족한 실정이다. 따라서 앞으로 정부의 체계적이고 지속적인 상호연계 속에서 연구 기반이 형성되도록 노력해야할 필요성이 요구되고 있다.

이런 상호 협력적 기반조성이 이루어지지 않음에 따라 생명과학 관련 사업의 결과들이 체계적으로 잘 정리되지 않아 Bioinformatics 를 할 만한 일차적인 자료가 정리되지 못한 실정이며, 이는 생물학적 데이터 특성을 이해하고 데이터의 표준화에 매우 어려움을 가지게 되었다. 더불어 Bioinformatics 분야의 체계적인 커리큘럼의 미확립과 전문인력 양성 교육기관의 부족하며 국내는 informatic 연구자로 편중된 Bioinformatics 연구 분위기(학회 및 실제 연구기관의 연구자 전공을 고려했을 경우 임)가 조성됨으로서 생물의 다양성과 com puter의 조화가 잘 이루어지지 않고 있는 것이 현실이다.

V. 생물정보학분야에서의 KISTI의 역할

끝으로 국가기관으로서 KISTI의 역할을 살펴보면 KISTI 출범의 기본 목표는 국가 과학기술의 정보관리 및 데이터베이스의 운영 및 제공에 있다고 생각되는데, 그동안 이루어 온 많은 업적과 실적을 토대로 생물정보학 분야에서 KISTI가 앞장서야 할 일들을 살펴보면 앞으로 많은 도전과 노력이 필요할 것으

로 생각한다.

첫째, 생물정보학 분야에서 분석시스템의 개발에 대한 주력보다는 기존의 국내외 생물학 정보들을 축적하고 표준화 체계를 확립하여 국제화하도록 해야 할 것이며 이를 위해 다각적인 노력이 필요할 것이다. 구조유전체, 기능유전체, 단백질구조 및 기능, 구조생물학 등에 대한 보다 폭넓은 분야의 정보를 운용할 수 있는 시스템 구축이 필요하며, 국외의 타 생물정보학 및 유전체 연구기관과 협력을 통해 데이터의 확장을 통한 국제화와 국내 연구자에 대한 서비스체계의 설립이 필요하다.

둘째, 국내 고유의 생물 유전체 데이터베이스 구축 지원 및 서비스 운영이 필요하다. 이를 위해 각 분야의 생물학 데이터를 제공하는 국내 연구기관에 대한 적극적인 지원과 협력, 그리고 상호 데이터교환 시스템이 요구될 것이다. 이것의 목적은 국외 기관(예를 들어, GenBank, DDBJ, EMBL 등)과 국내 고유의 데이터베이스의 통합을 통해 국내 생물학 및 생물정보학의 연구역량을 높이는 데 있다.

셋째, 연구자들 사이의 정보 교환을 위한 연동 체계 시스템 운영이 필수적이다. 즉 생물학 분야의 다양성을 통합하여 이를 연구시키기 위해 상호 정보교류와 상호간의 불편한 점을 이해하고 해결할 수 있는 시스템의 운영이 필요하다. 즉 각 분야 데이터의 공공화 및 표준화 확립이 절대적으로 요구되며 이를 기반으로 전문가들을 위한 생물정보학 교육 프로그램 개발 및 운영을 위한 다각적 노력이 필요하다. 현재 국내의 경우 대부분의 연구기관의 데이터의 사유화가 대부분이므로 생물정보학의 한계성을 극복하지 못하고 있다.

따라서 이러한 지원이 있을 때 생물학 연구분야의 연구자들도 KISTI에 정보를 제공하고 이를 즐겁게 이용할 수 있게 될 것이라 생각한다.

마지막으로, 생물정보학 분야 관련 산업체의 기술 관리와 지원에 더욱 적극적이어야 할 것이다. 정보부 뿐만 아니라 산자부와 연계하여 IT와 BT 사이에서 교량적 역할을 해주면서 새로운 기술의 개발 및 이의 관리와 이를 산업화할 수 있는 지인시스템을 구축하여 지원해야 할 것이다. ³⁶⁾