

슈퍼컴퓨터와 DNA 정보 활용

이성우 (E-mail: emoo@kju.ac.kr)
 동아대학교 생명공학부

I. 서 언

II. 생명공학과 슈퍼컴퓨터

III. 게놈연구의 산업화와 DNA 칩

IV. 유전자특허상식과 생명공학 상목시스템(BIOPASS)

V. 결 언

I. 서 언

세계의 식학들과 인문들은 21세기는 정보기술(IT)과 생명공학(BT)이 주도하는 고도 지식사회가 될 것이라 예측하고 있다. 마이크로소프트사의 빌 게이츠는 미래사회를 '하늘에는 정보통신산업, 땅에는 바이오산업'으로 표현했는데, 이는 공해를 일으키는 굴곡 산업을 대체하고 풍요롭고 건강한 사회를 만들어야 하는 인류의 임원과 정보산업의 선두에 선 입장에서 앞을 내다본 것으로 많은 시사점을 보여주고 있다.

정보기술을 일으킨 디지털정보가 '0'과 '1'의 두 가지 문자로 쓰여져 무한히 복제가 가능하고 전달이 용이하다면, 생명의 정보도 네 가지 염산 문자 즉, 네 종류의 DNA염기 'A', 'C', 'G', 'T'로 표시되고 무한히 복제되어 전달될 수 있다는 특징이 있다. 지구상에 존재하는 천만 종이 넘는 생물종들이 모두 이러한 DNA 염기서열 정보로 프로그램되어 다양성을 보이고 있다니 놀랍기 짝이 없다.

그렇다면 이 지구상에서 가장 고밀도의 정보를 가장 많이 축적하고 있는 것은 무엇일까? 단연 인간의 세포핵 속에 들어 있는 DNA로 구성된 '게놈'일 것이다.

인간은 하나의 수정란이 분할하여서 된 총 60조 개의 유전자로 볼 수 없는 작은 세포(1/10-1/100mm)로 구성되어 있다. 이들 각각의 세포의 핵 속에는 인간 전체를 만들고 운용할 수 있는 프로그램이 2m 정도의 DNA상에 두 세트를 저장되어 있으며 이를 게놈(genome, 유전체)이라 한다. 이 게놈 속에 저장된 정보의 축적도와 운용체제는 인간의 상상을 뛰어넘는 신비한 경지의 것이어서 21세기 내내 정보학자들과 생물학자들의 최대의 관심사이자 연구대제가 될 것이다.

또한 세포 각각에는 2m 정도의 DNA가 실다래처럼 응축 있으므로 60조 개의 세포에 포함된 DNA를 풀어서 전부 합치면 지구를 100번 이상 감을 수 있는 길이여 해당한다. 인간이 개발로 60조 개의 초슈퍼컴퓨터가 병렬로 오차 없이 조합되어 생명체를 이루는 최고급 정보의 응집합체라 할 수 있으며 이를 연결하고 운용하는 network system이 뇌와 신경망이다.

II. 생명공학과 슈퍼컴퓨터

최근 Human Genome Project의 진전에 따라 대량 생산되고 있는 유전 정보의 분석 및 활용을 위해 슈퍼컴퓨터가 많이 활용되고 있는데, 미국의 경우 최고 성능의 슈퍼컴퓨터와 2만 명 이상의 컴퓨터 프로그래머, 소프트웨어 전문가가 투입되고 있는 것으로 알려지고 있다.

특히 인간의 생명공학 현저화되고 게놈연구의 선봉에 있는 열매와 지노믹스의 관제자는 정보기술(IT)속 면에서 하나의 인간 게놈은 3GB의 DNA 시퀀스 데이터를 갖고 있어 인간 게놈은 주석이 포함된 70억자의 텍스트와 같으므로 최대 90TB 용량의 데이터 베이스를 필요로 하며 매년 9TB의 증가하는 데이터베이스를 관리하기 위하여 컴팩사의 슈퍼컴과 저장장치인 스토리지리스 도입이 불가피했다고 말하고 있다. 열매는 DNA 염기서열을 초고속으로 해석할 수 있는 300대의 ABI3700 자동분자기를 보유하고 이로부터 생산된 막대한 양의 유전정보를 저장하고 관리하기 위해 추후 극구장 절반 크기 규모의 데이터센터를 운영하고 있는데 이 센터는 단지 6평으로 운영되어 초당 1조 3,000억 부동

소수점 인산능력을 가진 컴택시의 알파64비트 프로세서 1,200개가 상호 접속된 시스템으로 되어 있다.

그런데 DNA에 담겨져 있는 방대한 양의 유전정보는 다양한 단락질을 생산하고 이 단락질들이 상호작용을 하면서 더욱 복잡한 정보를 만들어 내기 때문에 앞으로 유전정보를 바탕으로 전구체 생물학의 연구는 컴퓨터를 이용해야만 이들 정보의 저장과 활용이 가능하고, 반대로 이들 데이터를 저장하고 신속히 처리하기 위한 하드웨어 및 소프트웨어의 개발이 컴퓨터 및 정보통신 산업의 발전을 가져오는 상호 보완적인 관계를 유지하면서 발전하게 될 것이다.

또한 컴퓨터 과학자들은 정보 과학자들은 디지털 정보의 정밀을 살려 DNA와 단락질의 유전정보를 가꿈, 분석하는 고부가가치 산업을 만들어낼 뿐만 아니라 의학 및 생물학자들에게 합쳐 고도로 발전된 뇌와 신경망의 연구를 통해 인간에게 유용한 수많은 정보통신기술을 개발해낼 것으로 예상된다. 21세기에는 생명공학과 발전된 정보통신기술이 만나 고도의 지식산업을 활짝 꽃피게 될 것임에 틀림없다.

한편 계통연구에 필수적인 새로운 학문으로 각광을 받고 있는 생물정보학(Bioinformatics)은 생명공학과 정보학에 대한 고도의 전문지식이 동시에 요구되는 분야이나, 훈련된 전문인력 부족이 전 세계적인 심각한 실정이며, 파격적인 고임금을 조건으로 대학의 전문인력이 재임퇴사나 생명공학벤처회사로 스카웃되어 나가자 농사지를 씨앗까지 먹어 치워야 되겠다는 말까지 나오고 있다.

일부 전문기술은 앞으로 2-3년 이내에 아무리 큰 저약화이라도 이러한 계통연구 활용능력과 컴퓨팅 능력을 확보하지 못하면 그 기술을 상실할 것이라는 분석을 내놓고 있으나, 한국은 이에 대한 대비가 미흡하고 정

정직 있는 전문인력이 희귀한 실정으로서 바이오산업 발전에 큰 제약요소가 되고 있다.

미국의 경우 국립보건원 산하에 국립생명공학연구소(NCIB)를 설립하여 200여명의 박사급 전문기술이 매일 전세계 유전정보를 수집 분석하고 있으며, 유럽은 영국의 케임브리지 대학 근처에 유럽생물정보연구소(EBI)라는 통합된 중앙 시스템을 구축하여 유전자 데이터 교환 및 전문인력 양성에 나서고 있고, 일본도 생물산업 육성을 위한 고도 생물정보망 구축을 구상중에 있다.

이러한 공공적인 신뢰할 수 있는 생물정보망 등 지적 기반이 구축되어야 세계 각국 공동연구 수행 및 데이터의 교환이 가능하고, 또한 이러한 정보에 용이하게 접근할 수 있어야 원격 진료시스템이나 환경 모니터링 등 고임의 고용을 창출할 수 있는 수많은 지식산업들이 생겨날 수 있기 때문이다.

우리 나라도 뛰어난 슈퍼 컴퓨터를 보유할 수 있는 연구개발정보센터(KORDIC)에서 첨단과학정보 DB구축의 일환으로 유전자DB, 신약DB, 2차원 인체영상DB를 우선 구축하기로 한 것은 그나마 참으로 다행한 일이다.

아무튼 21세기 지식 사회에서 가장 방대하고 값진 천연의 정보가 유전자정보이고 이로부터 유용한 정보를 찾는 일은 현대판 금맥을 찾는 일에 비견될 전망이다. 그래서 인간계통해석 1차정보의 공개를 둘러싸고 세계인문들이 그렇게 떠들썩한 것이 아닐까.

III. 계통연구와 산업화와 DNA 칩

DNA칩이란 유전정보를 담고 있는 DNA조각을 등전 크로마틴 기판 위에 미리 고정되도록 실어놓아, 이에



■ 그림 1 열려진 슈퍼컴퓨터 정보센터 내부

기술·정책동향

부착하는 발암유전자들을 찾아낼 수 있도록 한 생화학 반도체를 말한다. 이 DNA칩의 원리는 DNA의 네 가지 염기(A, C, G, T)를, 구아닌(G-C)·시토신(C-G)과만 상보적으로 결합하는 생결합을 활용한 것으로 압동 결합의 전단계 유전자들의 기능 분석에 시간과 노력을 획기적으로 절약할 수 있으며 1회 사용할 수 있고 하나에 막대한 이상하는 고부가가치 제품이다. 향후 Human Genome Project가 완료되어 유전자의 기능 규명 및 산업화가 시도되면 급격히 수요가 증가될 것으로 예상되며 시장규모가 연평균 40%씩 성장하여 2010년에는 180억 불에 이를 전망이다.

경보산업과 함께 21세기를 이룰 핵심산업으로 등장하고 있는 생명공학산업에 키가 위해서는 DNA칩 기술의 확보가 꼭 필요한데 우리나라의 경우 반도체 메모리 생산 기술이 세계 수준이고 정보통신산업이 비교적 잘 발달되어 있어 이러한 강점을 활용하면 선진국을 따라잡을 수 있을 것으로 예상된다.

하지만 DNA칩은 유전자분석에 비해 가공 기술과 전자 제어 기술과 정보처리 기술이 접목된 지식집약적이고 복합 기술이 적용되므로 단순히 DNA칩의 제작 기술만으로는 산업화가 될 수 없고 앞서 살펴본 생물정보와 사회의 지적 자본이 갖추어져야 하므로 산업화를 위한 특별한 전략이 요구된다.

그 주요 기술을 살펴보면 다음과 같다.

DNA칩의 주요 기술

- 유전자공학기술 : DNA의 추출, 분석 가공 기술
- 미세가공기술 : 작은 기판 위에 수만 개의 DNA를 부착하는 기술
- 분석기술 : 부속된 DNA 조각을 인식하고 검출하는 기술
- 정보처리기술 : 검출 결과를 컴퓨터로 집계 처리하여 해석하는 기술

IV. 유전자 특허심사와 생명공학 검색 시스템 (BIOPASS)

최근 국제적으로 추진되고 있는 Human Genome

Project의 완료가 가시화됨에 따라 유전자의 기능규명을 통한 유전정보의 산업화가 급속히 진행될 것으로 예상되고 있으며, 이에 전산업의 관심은 막대한 부가가치를 창출하게 될 유전자의 특허권 확보에 집중되고 있다.

앞으로 생명공학은 생물의 유전정보에 기초하여 발전하게 되는데, 생물의 유전정보는 다른 분야의 정보와는 달리 매우 제한되어 있으며, 이후의 연구는 모두 이 제한된 정보를 이용하지 않을 수 없다는 특징을 가지고 있다. 따라서 이 제한된 유전정보를 선점하는 국가가 21C 생물산업을 지배하게 될 것이며, 이러한 유전정보를 비축하지 못한 나라는 기술선진국에 예속될 가능성이 있다.

따라서 선진국들은 유전자(Genome) 연구 결과 얻어진 새로운 유전자 서열과 그 기능의 연구결과를 특허권으로 선점함으로써 바이오벤처법, 제약법, 농업/식품법, 예치차량법 등이 이를 산업화 할 수 있는 기반을 마련하여 21세기 이끌어갈 부의 원천으로 삼고자 하고 있는 것이다.

최근 유전자의 특허 출원이 급증하면서 미국, 일본, 유럽 등 선진국 특허청을 중심으로 출원시, 컴퓨터 특허이 가능한 형태의 서열목록 전자파일을 제출하도록 하는 제도가 개발되어 활용되고 있다. 유전자 서열은 보통 수백 수천 개의 서열이 반복되고 있어 육안으로 자료를 찾거나 차이점을 대변하는 것이 실질적으로 불가능하므로 심사의 편의를 도모하고 서열 DB구축을 용이하게 하기 위해서이다.

우리 특허청에서도 이러한 세계적인 흐름에 적극 대처하여 1999년 1월 1일부터 유전자서열목록의 전자파일 제출을 의무화하는 한편, 특허청 내에 구축될 생명공학 특허서열 데이터베이스를 운영하고 외부 DB에 쉽게 접근할 수 있는 생명공학 검색시스템(BIOPASS, Biotechnology Patent Sequence Search System)을 모스데이타, 생명공학연구소, 포항공과대의 컨소시엄을 통해 개발하여 활용하고 있으며 올해 3월부터 KIPRIS의 웹사이트 (www.kipris.or.kr)를 통해 일반인들에게도 무료 공개하고 있다.

생명공학 검색시스템은 서열의 상동성 검색뿐 아니라 발명의 명칭, 출원번호 등 서지적 사항으로도 검색

