



웹기반 유전체 분석시스템 개발

스몰소프트 박기정 · 서화정 · 태홍석
김기봉 · 남혜원 · 이대상

1. 서 론

1.1 연구개발의 목표와 의의

유전체 연구는 현재의 생물학 연구의 기반이 되는 연구분야이고, 생물학 응용 산업의 가장 핵심적인 연구분야이다[1]. 이는 식량개발, 신약개발이나 신물질 개발과 같은 국가적인 부가가치 산업이 유전정보에 절대적으로 의존할 것으로 예상되고 있기 때문이다. 한편, 이 분야 연구에서 생물정보학이 가지는 비중은 절대적이다. 유전체 서열 결정 이후 서열분석을 위한 시스템은, 다양한 분석 방법을 수행하기 위한 분석 모듈 [2][3]과 데이터베이스 [4]가 관여하므로, 이를 효과적으로 활용하여 유전체 정보 분석을 효과적으로 수행하여 분석 정보를 효과적으로 정리하고 검색할 수 있는 시스템은, 전체적인 유전체 연구의 효율에 결정적인 영향을 미치게 된다. 유전체 연구를 위한 생물정보학적 방법과 시스템 개발에 관련된 분야를 유전체 정보학(genome informatics)이라고 한다[5].

유전체 정보 분석시스템은 현재 그 전용 시스템이 선진국을 중심으로 개발되고 있으며, 개발된 제품은 유전체 연구의 필수 소프트웨어로 대단히 고가의 소프트웨어로 판매될 전망이다. 한편, 국내 연구자의 연구환경에 적합한 사용자 인터페이스를 갖춘 시스템의 개발이 시급한 상황이며, 이러한 개발이 이루어지면, 변화하는 연구환경에 맞춘 시스템 개선과 보정이 지속적으로 이루어질 수 있게 된다. 따라서, 국내 유전체 연구자의 생물정보에 대한 고급 분석 욕구 충족에 기여하고, 기반이 되는 유전정보 분석의 핵심기술을 확보하며 동시에 방대한 유전정보 분석 소프트웨어 시장을 선점하기 위해 이러한 통합 패키지의 개발을 선도하는 것이 시급하다. 이러한 취지에서 유전

체 서열에 대한 기능 분석(annotation) 시스템의 개발을 목표로 본 연구개발을 시작하게 되었다.

1.2 해외의 연구개발 현황

유전체 기능 분석의 주요 분석 도구로서는, 상동성분석, 유전자 기능부위 예측, 단백질 motif 분석(이후 motif 분석), 프로모터 분석 및 gene ontology 분석 등을 위한 도구를 들 수 있다.

상동성 검색(homology search)을 위한 alignment 알고리즘에 대한 연구는, 1970년 Needleman & Wunsch 의 dynamic programming 기법에 의한 pairwise alignment algorithm이 발표된 이후, Smith & Waterman 등이 일반적인 길이의 gap에 대해 이 algorithm을 확장함으로써, 두 sequence 간의 alignment를 위한 실용 가능한 프로그램들이 구현되기 시작하였다. 이들 연구에 기반하여, 전산학의 graph algorithm 이론이나, pattern matching algorithm 등이 주로 생물학 패턴의 검색에 많이 응용되어 여러 응용 알고리즘들이 개발되어 왔으며, 데이터베이스의 개발과 함께 대량의 데이터를 대상으로 하여 빠른 실행 시간이 요구되어 여러 heuristic algorithm들이 개발되었다. Lipman, Wilbur, Pearson 등이 hashing과 window 설정의 방법을 사용하여 고안한 FASTP/FASTN는 local homology alignment를 포함한 FASTA로 발전하여 데이터베이스 대상의 homology search에 사용되었고, Altschul과 Karlin 이 homologous candidate fragment 선정과 extension 방법을 사용하여 개발한 BLAST는 데이터베이스 대상의 local homology search를 위해 가장 많이 활용되어 왔으며, 1990년대에는 BLAST의 결정적인 단점으로 gap 처리가 안되던 것을 해결한 BLAST2 [6]가

발표되어 BLAST를 대체하여 현재까지 가장 많이 활용되고 있다. 같은 목적으로 MPP 기종에서 병렬 프로그래밍된 MPsrch가 상용 프로그램으로 개발되어 있다. 현재 유전체 프로젝트의 결과로 폭증한 데이터로 인한 상동성 검색의 시간문제와 보다 나은 sensitivity의 향상을 위한 노력들이 이루어지고 있다. Myers가 최근에 개발하여 일부에서 활용되고 있는 MEGA-BLAST의 경우 속도는 빠르나 실제로 논문에서는 소개되지 않은 sensitivity 상의 문제가 있으며, IBM에서 개발한 FLASH와 같은 알고리즘은 DNA 데이터에 대해서는 적용이 되나 단백질 서열의 경우에는 심각한 sensitivity의 문제를 안고 있어 실용화와는 거리가 멀다. 한편, 클러스터링 서버의 발전 등으로 알고리즘 상의 개선과 함께 하드웨어의 아키텍처를 활용한 구현상의 개선을 위한 노력도 이루어지고 있다. 현재 생물학 연구자들의 분석요구를 수용하기 위해서는 알고리즘 면에서 보다 개선되고 사용자 인터페이스가 개선된 새로운 상동성 검색 프로그램들이 개발되어야 할 것이다.

유전자 기능부위 예측(gene prediction)을 위한 알고리즘 [7][8][9][10]은, 단순한 ORF 예측 알고리즘을 시작으로 1970년대부터 지속적으로 개선되고 있다. 1980년대의 GRAIL과 GRAIL II를 기점으로 실용적으로 활용될 수 있는 유용한 프로그램들이 소개되고 있으며, 기능유전체 분석을 위한 핵심적인 도구로서 인식되어 최근 급속히 발전되고 있다. 이들 알고리즘은 search by content 와 search by signal 의 두 가지 방식을 기본 방법으로 하여 두 가지를 적절히 혼용하여 알고리즘을 구성하고 있다. GenQuest의 경우는 상동성 분석을 기반으로 하여 비교적 정교한 결과를 산출한 알고리즘으로 실용적으로 활용되기도 하였다. GRAIL을 기점으로 기계학습 알고리즘을 응용하여 neural network이나 HMM 등을 사용한 방법이 주류를 이루고 있으며, 최근에는 HMM을 사용한 알고리즘들이 대세를 이루고 있다. Prokaryote 유전체의 분석을 위한 알고리즘은 TIGR나 Sanger 센터와 같은 초기의 각 유전체 연구단에서 자체적으로 개발한 몇 가지가 활용되어 왔는데, 최근에는 간단한 목적으로 NCBI의 ORFfinder를 사용하거나, 실용적인 프로젝트를 위해서는 TIGR에서 개발한 Glimmer가 가장 보편적으로 활용되고 있다. Eukaryote 유전체 분석은 일반적인 도구는 인간 유전체 프로젝트 [11]를 위해 사용된 도구를 대표적으로 활용하고 있다.

GRAIL이 1990년대에 일반적으로 사용되어 오다가 최근에는 GenScan이 인간 유전체 프로젝트의 실용적인 도구로서 사용되면서 본격적으로 활용되고 있다. GenScan은 기계학습을 위한 HMM의 파라미터를 비교적 정교하게 구성하여 그 결과의 sensitivity가 이전에 비해 개선되었다고 평가되고 있으며, 무엇보다 multiple gene에 대한 예측을 그 이전의 프로그램들에 비해 탁월하게 수행하고 있어 유전체 프로젝트를 위한 실용적인 프로그램으로 사용되고 있다. HMM을 비롯한 기계학습 알고리즘들에 대한 시험이 계속되고 있고 파라미터에 대한 설계가 다양하게 시도되고 있으므로 이전의 알고리즘들의 단점을 개선하고 GenScan과 같은 실용성을 갖춘 알고리즘들이 지속적으로 개발될 것이다. 한편, 유전체 프로젝트를 위해 사용자 인터페이스가 강화된 프로그램들이 개발될 것이다.

Motif 검색을 위한 알고리즘은 Prosite 와 같은 데이터베이스를 기반으로 하는 패턴검색을 위주로 이루어져 왔으며, 현재 정확성 면에서는 신뢰도가 매우 낮은 분야이나, 상동성 검색외의 기능 분석의 주요 방법으로 그 중요성은 매우 크게 평가받고 있다. Motif 구성을 위한 알고리즘은 information theory나 HMM을 활용하여 통계적이나 기계학습적인 방법으로 패턴을 구성하여 비교적 정교한 motif 프로파일을 구성하기 위한 방법들이 개발되고 있다. Motif와 유사하나 확장된 개념으로 ProDom, BLOCKS, Pfam 등과 같이 domain과 block에 대한 데이터베이스들이 최근에 개발되어 활용되고 있다. 유전체 자료들이 축적되고 기능부위에 대한 자료가 정확해 지면서 이들 데이터베이스의 양적인 증가가 급속히 이루어지고 있어, 이를 활용한 패턴 분석과 이에 근거한 기능 분석의 활용도가 증가하고 있다.

Promoter 분석도 motif 분석과 함께 상동성 검색외의 기능 분석의 주요 방법으로 그 중요성은 매우 크게 평가받고 있다. Motif 데이터베이스와 유사한 개념으로 알려진 promoter와 관련 데이터를 수집하고 정리한 데이터베이스로 TFD, TRRD, TransFac, EPD, RegulonDB 등이 최근에 개발되어 오고 있다. Promoter 패턴 구성 알고리즘도 HMM이나 information theory를 활용한 방법들이 많이 사용되고 있으며, 그 외에 decision tree를 사용한 방법 등과 같이 인공지능에서 사용하는 알고리즘을 활용한 방법들도 사용되고 있다.

Gene ontology 분석을 위한 노력은 유전체 분석이 활발해 지면서 최근에 급속히 진행되고 있으며, 표준화를 위한 노력이 특히 활발히 전개되고 있다. 분석을 위한 중요한 데이터베이스로서 NCBI에서 개발한 COG와 UniGene이 각각 prokaryote와 Eukaryote의 유전체 분석을 위해 활용될 수 있으며, 이들 데이터베이스는 양적으로 급속히 증가하고 있다. 양적인 증가와 함께 현재의 분류의 개선과 unknown 또는 unidentified 범주의 기능이 새로운 분류항목으로 다소 교체되어 더욱 복잡하고 세분화된 분류체계가 이루어질 것이다.

유전자 기능 분석(gene annotation) [12][13][14]을 위한 도구는 NCBI나 LANL과 같이 주요 데이터베이스 관리기관에서 자체적인 용도로 약 10여년 전부터 개발하여 사용해 왔으나, 유전체를 단위로 한 기능 분석의 도구는 최근에 개발되고 있다. Celera와 같은 거대 유전체 회사에서는 자회사를 통해 개발된 이러한 도구의 상용화를 추진하고 있으며, Informax, DoubleTwist 등과 같이 잘 알려진 생물정보 기업들이 자체적으로 이에 상응하는 시스템을 개발한 바 있으며, 현재 다수의 생물정보 기업들이 이의 연구개발에 참가하고 있다.

1.3 국내의 연구개발 현황

국내에서는 본 연구팀이 서열정렬과 유전자 기능 부위 예측 등에 대한 알고리즘 및 프로그램 개발연구를 수년간 수행하였으며, 최근 수년간, 전산학과와 생물관련학과를 중심으로 대학에서 활발한 연구가 시작되고 있고, 연구소와 생물정보학 벤처 일부에서도 관심을 가지고 기술과학과 연구개발을 하고 있는 상황이다.

본 연구팀에서는, Windows 기반의 개별 프로그램으로 다수의 모듈을 개발한 바 있으며 [15][16], pairwise alignment나 multialignment 등의 알고리즘 개발과 promoter 분석 알고리즘 개발 [17], 유전자 기능 부위 예측 알고리즘 개발 [18] 등에 대한 연구를 다년간 수행해 오고 있다. 국내에서는 적절한 규모를 갖춘 패키지의 개발이 최근에 소규모로 이루어지기 시작하고 있으며, 분석 모듈에 대한 개발이 소수의 생물정보 벤처에서 발표된 바 있다.

전반적으로 국내에서는 본 연구의 목표와 같은 시스템을 개발하기 위한 전체적인 연구와 각 요소기술에 대한 기술경험이 약하고 사용자의 사용경험이 빈

약하여, 현장연구를 효과적으로 반영하기 위한 사용자 요구분석 반영이 취약할 수 있다. 본 연구팀은 다년간 생물학 연구자와의 다양한 유전정보 분석 공동 연구를 통해 이러한 경험을 개별적으로 갖추고 있어 사용자 요구분석에 대한 노우하우를 갖추고 있어, 이러한 제약요인에 대처하면서 연구개발을 수행하였다.

2. 본 론

2.1 접근 및 구현 방법

그림 1은 유전체 프로젝트 진행을 위한 전체 생물정보 분석시스템의 일반적인 구조를 나타내며, 본 연구개발의 목표와 범위는 유전자 예측 분석시스템과 관련 데이터베이스 시스템의 개발을 포함한다.

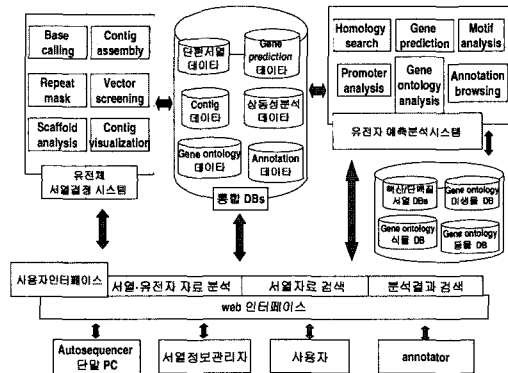


그림 1 유전체 프로젝트 수행을 위한 정보 분석시스템 구조

본 연구에서 개발된 시스템은 WeGAS(Web-based Genome Annotation System)으로 명명하였다. WeGAS는 gene prediction 및 상동성 검색에 기반하여 기능을 1차적으로 예측하고, 상세한 기능예측 또는 기능이 예측되지 않은 유전자의 기능 예측을 위해 motif 분석을 보조적으로 시행하고, 유전체 전체에 대해 promoter 분석을 하여, 유전체 전체에 대한 annotation 결과를 생성하도록 구성하였다. 사용자 인터페이스와 drawing 및 formatting 도구를 구현하였다.

Gene prediction 모듈 개발에서는, prokaryotes과 eukaryote에 대해 각각 Glimmer와 GenScan을 사용하여 입력한 유전체 서열 혹은 contig 서열에 대해 분석할 수 있도록 시스템을 구축하였다. 추후의 연구

개발을 통해, prokaryotes과 eukaryotes의 두 가지 모델에 대해 gene prediction 프로그램 모듈을 각각 개발한다. HMM을 사용하여 모델을 구성하고, prokaryotes과 eukaryotes의 종별로 training 결과를 분석하여 파라미터를 설계한다. GC 컨텐트에 대한 dependency 등을 분석하여 파라미터에 반영한다.

프로모터 분석은 eukaryotes와 prokaryotes에 대해 별도의 정보와 분석 방법이 필요하므로, 각각 PromEC와 EPD의 promoter 패턴 데이터를 검색할 수 있도록 시스템을 구축하였다. 패턴 검색을 위해 BLAST를 사용하였고, 검색 대상은 코딩 유전자의 프로모터 가능 지역으로 하고, 검색 결과는 각 유전자별로 저장되도록 하였다.

Motif 분석은 eukaryotes와 prokaryotes의 구분 없이 이루어지며, 예측된 각 유전자의 단백질 서열에 대해 PROSITE의 regular expression 패턴을 검색하도록 하였다. 검색된 결과는 각 유전자별로 저장되도록 하였다.

Gene ontology 분석 시스템에서는, prokaryotes

분석은 COG를, eukaryotes 분석은 UniGene을 기반으로 하였다[19]. 상동성 분석과 GenBank [20][21] 및 LocusLink 등을 활용하여 각 유전자의 기능을 지정할 수 있는 분석 시스템을 구현하였다.

전체 시스템의 입력 서열로부터 분석 과정과 분석 결과의 전체 흐름과 데이터 관리를 위해 데이터베이스 시스템을 구축하였다. 주로 gene annotation의 전체 절차와 각 단계별 데이터의 종류와 연결관계를 분석하여 데이터베이스를 설계하고 구현하였다. Prokaryotes과 eukaryotes은 별개의 스키마를 가지도록 구성하였는데, 이는 분석도구의 차이와 유전자 구조의 차이 및 gene ontology 분석 방법 등의 다양한 차이가 있어, 각각에 적합한 스키마를 구성하기 위해서이다(그림 2). DBMS로서 MySQL을 사용하였다.

2.2 시스템 인터페이스 및 모듈별 구현 결과

현재 개발 완료된 WeGAS 시스템에 대한 구현 내역은 표 1과 같다.

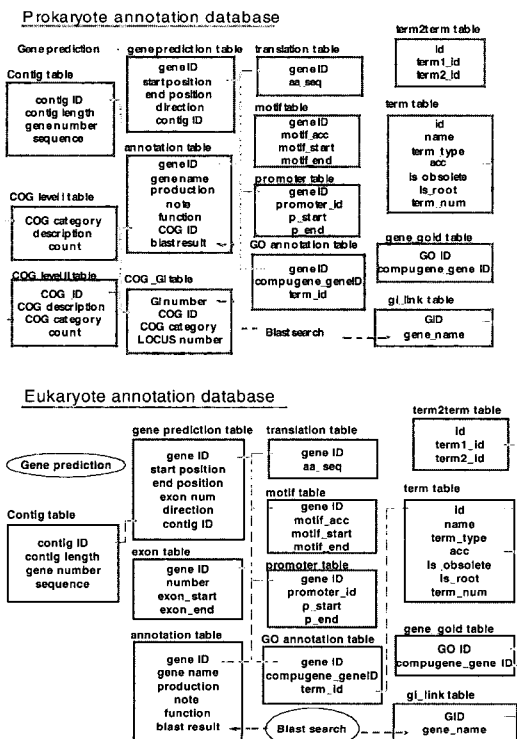


그림 2 Prokaryotes과 eukaryotes 유전체에 대한 데이터베이스 스키마

표 1 WeGAS 시스템 개발 결과 요약

분야	서브분야	진행 내용	차별성 및 의의
annotation 프로그램	Hmblay search 프로그램	*BLAST 기반 번역 환경 구축	사실성 향상
	Gene prediction 모듈	*Prokaryote gene prediction 프로그램 구축 *Eukaryote gene prediction 프로그램 구축 *Training set 개발을 위한 decomposition 알고리즘 개발	Selectivity 향상에 (유전자종 분류)
	Promoter search 모듈	*Prokaryote promoter search 프로그램 개발 *Eukaryote promoter search 프로그램 개발	Genome scan - annotation 가능
유전자기능정보관리	Motif search 프로그램	*Prosite 기반 motif search 프로그램 개발	Genome scan - annotation 가능
	Gene ontology 분석 프로그램	*Prokaryote 기반 COG 기반 gene ontology 분석 프로그램 *Eukaryote 기반 UniGene 기반 gene ontology 분석 프로그램 *Eukaryote 기반 UniGene 기반 gene ontology 분석 프로그램	유전자기능정보관리 유전자기능정보관리
기타 annotation 시스템 구축	annotation GO database 시스템 구축	*annotation GO database 구축	신규
	annotation 연관 데이터베이스 구축	*기타 annotation 분석 데이터 및 GO database 연관 데이터 구축 *annotation 연관 데이터베이스 구축	신규
기타 GO	annotation 연관 분석 프로그램	*annotation 연관 분석 프로그램 개발	신규
	annotation 연관 분석 프로그램	*annotation 연관 분석 프로그램 개발	신규

전체 시스템의 인터페이스는 그림 3과 같이 단계별 처리에 대한 메뉴로 구성되어 있다. 프로젝트별로 데이터베이스를 구축하게 되므로, 프로젝트별 데이터베이스를 호출할 수 있는 기능과 각 분석 모듈별로 분석을 호출하는 기능 및 최종 분석 결과에 대한 브라우저의 호출 등이 포함되어 있다.

한편, 각 분석 모듈의 실행 인터페이스는 그림 4와 같이, 실행을 시작하는 인터페이스와 진행 상황을 표시할 수 있는 인터페이스 및 완료된 상황을 볼 수 있

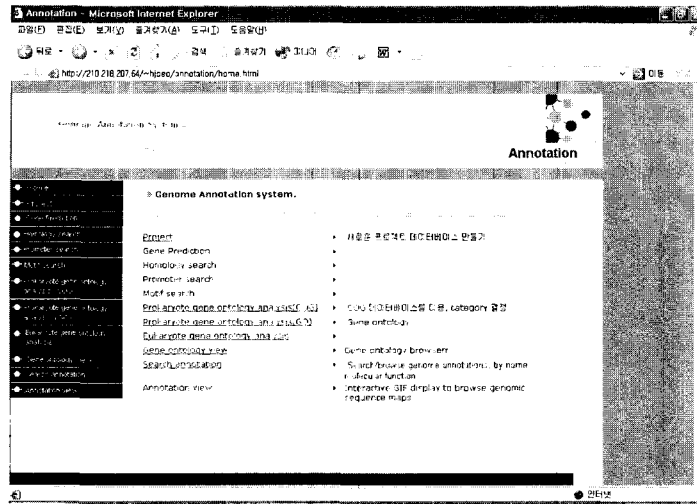


그림 3 WeGAS 시스템의 시작 페이지



그림 4 WeGAS 시스템의 프로모터 분석 인터페이스

는 인터페이스 등으로 구성되어 있다. 그림 4는 프로모터 분석의 인터페이스로 eukaryotes과 prokaryotes에 대해서는 내부 분석 모듈은 다르지만 유사한 인터페이스로 구성되어 있다.

Gene ontology [22] 분석을 위한 프로그램은, eukaryotes과 prokaryotes에 대한 분석 방법의 차이로 인해 전혀 다르게 개발하였다. Prokaryotes에 대해서는 COG를 기반으로, Eukaryotes에 대해서는 GO를 기반으로 하여 해당 데이터베이스에 대한 상동성 검색과 해당 클러스터에 대한 mapping 및 정리

등이 이루어지도록 하였다(그림 5).

즉, COG와 GenBank를 통해 상동성 분석된 prokaryote 유전자 각각의 기능을 mapping 할 수 있는 프로그램을 구현하고, UniGene과 LocusLink, GenBank를 통해 상동성 분석된 eukaryote 유전자 각각의 기능을 mapping 할 수 있는 프로그램을 구현하였다. 그 분석 결과는 그림 6과 같이 출력된다.

한편, 최근에 prokaryotes에 대한 GO 데이터베이스를 활용할 수 있도록 시스템을 보강하였으며, 그 결과는 데이터베이스 스키마에 포함되어 있다.

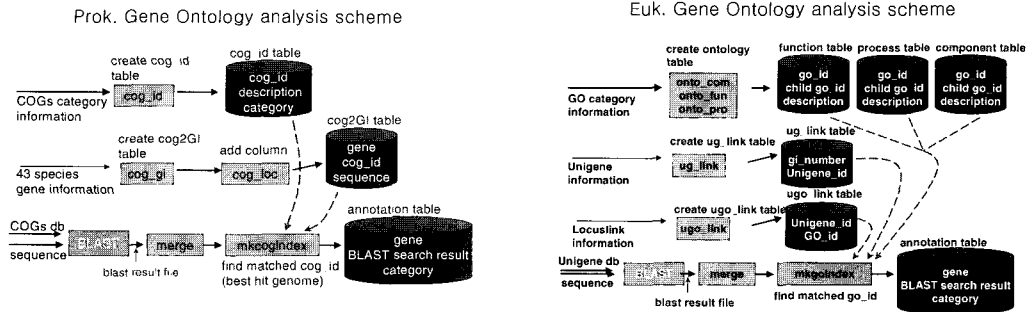


그림 5 Prokaryotes과 eukaryotes에 대한 gene ontology 분석 프로그램 구성



그림 6 Gene ontology 분석 결과 출력

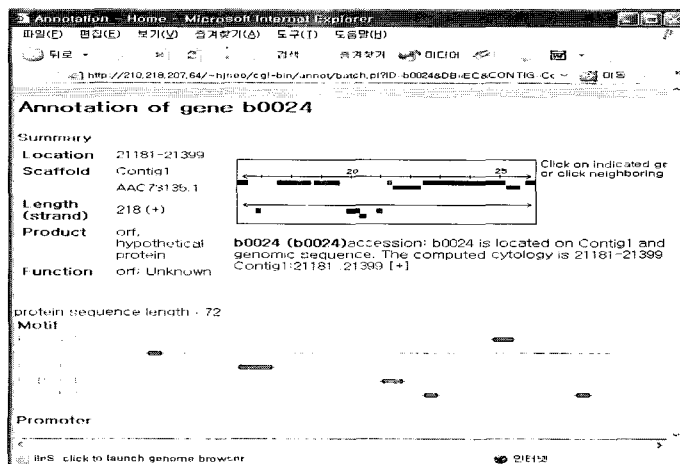


그림 7 유전자별 기능 분석 정보

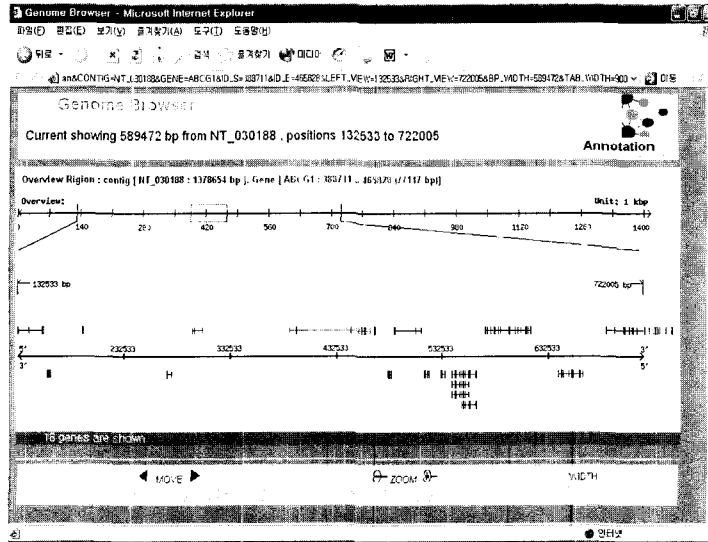


그림 8 genome 브라우저

유전체 분석 결과에 대해 통합적으로 각 유전자에 대해 데이터베이스에 저장되어 있으며, 이는 WeGAS의 자체 데이터베이스 분석 프로그램을 통해 검색할 수 있도록 데이터베이스 검색 프로그램을 개발하였다. 검색 프로그램의 결과는 해당 조건에 맞는 유전자에 대한 리스트로 출력되며, 각 유전자에 대한 상세 내용은 그림 7과 같이 출력된다.

그림 7에서와 같이 분석 결과는, 각 유전자의 기본적인 기능 분석 결과 외에도 포함되어 있는 motif 분석 결과와 관련된 프로모터의 분석 결과 등이 포함된다.

유전자에 대한 개별적인 검색 외에, 분석된 유전체의 유전자 map을 통해 전체 유전체를 조감할 수 있는 기능으로 간단한 genome 브라우저를 개발하였다. 그림 8과 같이 유전체를 위치별로 이동하면서 유전자의 분포와 구조를 파악할 수 있다.

줌인 및 줌아웃 등으로 전체 화면이 보여주는 범위를 스케일링 하는 기능과 화면 상에 나타나는 넓이를 조절하는 기능 등을 제공하므로, 전체적인 조감이나, 세부적인 검색 등의 기능을 모두 수행할 수 있다. 각 유전자에 대한 세부 정보는 클릭해서 그림 7과 같이 볼 수 있다.

3. 결론

WeGAS의 최종 분석 결과는 유전체 분석의 요약 자료를 생성하는데, 이러한 자료는 전문적인 생물정보 분석을 체계적으로 수행하지 않으면 생성할 수 없

는 복잡한 단계를 거치고 전문적인 생물정보 분석 인력과의 공조를 통해 생성할 수 있는 것이다. 이러한 결과를 본 시스템은 정해진 절차에 따라 단시간 내에 효과적으로 생성해 낼 수 있어, 유전체 연구자들이 개인의 유전체 분석연구를 통한 고급 연구 결과를 생성할 수 있도록 직접적으로 지원한다. 유전체 분석을 위한 분석 모듈은 유전체 연구자 뿐만 아니라 대부분의 생물학자에게 필요한 도구이다. 따라서 적절한 사용자 인터페이스를 갖춘 이러한 통합분석 도구는 생물학 연구자의 분석 기술을 향상시키는데 직접적으로 기여할 것이다.

본 연구를 통해, 다양한 분석을 위한 모듈을 자체 개발하고 이들 도구를 통합하여 완성된 독자시스템을 개발한다는 면에서, 유전체 자동화 분석을 위한 요소기술을 해결하여 확보하고 전체 시스템화 하는 기술을 완성하여 국내 기술화 한다는데 기술적으로 의의가 있다. 유전체 분석을 위한 분석 모듈인 상동성 분석, 단백질 motif 분석, promoter 분석 등은 기능유전체학의 주요 분석방법들로서 다양한 응용분야를 가지고 있다. 따라서, 이러한 고급분석 도구나 모듈의 기반기술의 향상에 기여할 수 있으며, 이들 도구를 활용한 유전정보 분석 소프트웨어들의 개발에 직접 활용될 수 있다. 유전체 기능 분석시스템은 여러 분석 모듈과 사용자 인터페이스에 의한 이들의 유기적인 연결로 구성된 고급 분석 소프트웨어로, 유전체 연구의 핵심적인 경쟁력을 좌우하므로, 효과적인

시스템의 개발로 농생물 연구를 위한 다양한 유전체 분석 연구의 연구효율을 급속히 향상시키고 경쟁력 제고에 직접적인 기여를 하게 될 것이다[23].

본 연구는 3차년도에 걸쳐 연구개발이 진행될 예정이며, 현재 1차년도의 연구개발이 완료되었고, 2차년도의 개발이 진행 중이다. 다음과 같은 면에서 추후 연구 개발이 진행될 예정이다.

- 농생물 유전체 정보 데이터베이스 시스템 구축
- 농생물 관련 식물, 동물, 미생물에 대한 공개 유전체 데이터베이스를 수집하여 자체 데이터베이스 시스템을 구축하고, 검색시스템을 개발한다. 본 연구의 직접적인 1차 응용분야는 농생물과 유전체 분석이므로 이와 관련된 모델 생물에 대한 annotation을 1차 목표로 하여 개발을 진행한다.
- 농생물 유전체 정보 데이터베이스 시스템과의 연계시스템 구현 - 전체 기능 분석시스템과 농생물 유전체 정보 자체 데이터베이스 시스템과 외부 서버 시스템과의 연계와 자료 검색 및 검출을 위한 모듈을 개발한다.
- Gene prediction 알고리즘 개량 - gene prediction 알고리즘의 최신 방식을 반영하고, training 시험 결과를 통해 optimization 된 파라미터를 개발하고, 파라미터간의 weight를 반영하도록 알고리즘을 개량한다.
- Annotation 결과 feedback 시스템 개발 - 농생물 전문가에 의해 annotation 결과를 검증하고 검증 결과에 대한 report를 생성하고, 수동으로 annotation 된 결과를 분석에 반영할 수 있는 시스템을 개발한다.
- 농생물 유전체 정보 시스템 GUI 구축 - 통합시스템을 위한 전체 모듈의 연결을 위한 GUI를 개발하고, 여러 포맷기와 report 생성기의 GUI를 보완하고 통일화한다.
- 전체 분석시스템 통합 - 전체 시스템 GUI와 데이터베이스 및 분석 모듈을 연결하여 하나의 시스템으로 통합한다.

참고문헌

[1] 이연수(2001) 국내의 유전체 연구 동향 및 발전 방향에 대한 연구(국가과학기술위원회 제출 자료).
 [2] <http://genome.wustl.edu/est/esthmpg.html>
 [3] Christine Gemund, Chenna Ramu(2001). Gene2 EST. Nucleic Acid Research 29:1272~1277.

[4] Boguski MS, Lowe TM, Tolstoshev CM (1993). dbEST-database for "expressed sequence tags." Nat. Genet 4:332~3.
 [5] Kathy Liszewski(2000) Mining Data in Digital World via Bioinformatics:Companies Pursue New Approaches for Processing and Transforming Genomic Data. Genetic Engineering News Vol. 21, No. 1, January 1, 2000.
 [6] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller and David J. Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389~3402.
 [7] Steven L. Salzberg, Arthur L. Delcher, Simon Kasif and Owen White(1998) Microbial gene identification using interpolated Markov models. Nucleic Acids Research 26:544~548.
 [8] Alexander V. Lukashin and Mark Borodovsky (1998) GeneMark.hmm: new solutions for gene finding. Nucleic Acids Research 26:1107~1115.
 [9] Chris Burge and Samuel Karlin(1997) Prediction of Complete Gene Structures in Human Genomic DNA. J. Mol. Biol. 268:78~94.
 [10] John Burke, Hui Wang, Winston Hide, and Daniel B. Davison, Alternative Gene Form Discovery and Candidate Gene Selection from Gene Indexing Projects. GENOME RESEARCH 8:276~290.
 [11] L. Charles Bailey, Jr., David B. Searls, and G. Christian Overton(1998) Analysis of EST-Driven Gene Annotation in Human Genomic Sequence. GENOME RESEARCH 8:362.376.
 [12] Microbial Genome Program DOE/SC-0016.
 [13] Kyle Durick, John Mendlein, and Kleantlis G. Xanthopoulos(1999) Hunting with Traps: Genome-Wide Strategies for Gene Discovery and Functional Analysis. Genome Research 9:1019~1025.
 [14] L. Charles Bailey, Jr., David B. Searls, and G.

Christian Overton(1998) Analysis of EST-Driven Gene Annotation in Human Genomic Sequence. *GENOME RESEARCH* 8:362.376.

[15] Kim, K.-B., Seo, H., Nam, H., Tae, H., Kim, P.-G., Lee, D., Jeong, H. and Park, K., "An Integrated Sequence Data Management and Annotation System For Microbial Genome Projects." *Intelligence system for Molecular Biology*, 2001.

[16] Seo, H., Kim, K.-B., Tae, H., Park, W. and Park, K., "Development of gene ontology analysis and classification tools for microbial genome annotation." *Research in Computational Molecular Biology*, 2002.

[17] Lee, D., Park, C. and Park, K., "ScanPromW: A windows program searching for promoter patterns against a genome sequence." *Intelligence system for Molecular Biology*, 2002.

[18] Kim, K.-B. and Park, K., "Prediction of Transcription regulatory sites with Dependency-Reflecting Decomposition Model." *Intelligence system for Molecular Biology*, 2002.

[19] Stephen J. O'Brien, Marilyn Menotti-Raymond, William J. Murphy, William G. Nash, Johannes Wienberg, Roscoe Stanyon, Neal G. Copeland, Nancy A. Jenkins, James E. Womack, Jennifer A. Marshall Graves(1999) The Promise of Comparative Genomics in Mammals. *SCIENCE* 286:458~481.

[20] <http://www.ncbi.nlm.nih.gov/Genbank/index.html>

[21] Benson DA.(2000) GenBank. *Nucleic Acid Research* 28(1):15~8.

[22] Jeffrey Skolnick and Jacquelyn S. Fetrow(2000) From genes to protein structure and function: novel applications of computational approaches in the genomic era. *TIBTECH* 18: 34~39.

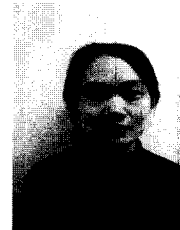
[23] Sophia Tsoka, Christos A.(2000) Recent developments and future directions in computational genomics Ouzounis. *FEBS Letters* 480:42~48.

박 기 정



1982. 3~1986. 2 서울대학교 공과대학 컴퓨터공학과 공학학사
 1986. 3~1987. 2 한국과학기술원 전자계산학과 석사과정
 1987. 3~1989. 2 한국과학기술원 생물과학과 이학석사
 1989. 2~1998. 6 한국과학기술연구원 생명공학연구소 (선임)연구원
 1991. 3~2002. 2 한국과학기술원 생물과학과 이학박사
 1998. 6~2000. 2 한국과학기술원 의과학센터 연구원
 2000. 3~현재 (주)스몰소프트 부설 정보기술연구소 소장/책임연구원
 E-mail : kjpark@bioinfo.smallsoft.co.kr

서 화 정



1991. 3~1995. 2 경북대학교 미생물학과 (이학사)
 1999. 9~현재 (주)스몰소프트 부설 정보기술연구소 연구원
 2000. 3~2002. 2 경북대학교 미생물학과(이학석사)
 2002. 3~현재 충북대학교 컴퓨터공학과 (박사과정 재학 중)
 E-mail : hjseo@bioinfo.smallsoft.co.kr

태 흥 석



1994. 3~2001. 2 경북대학교 미생물학과 졸업
 2001. 3~2003. 2 경북대학교 미생물학과 이학석사
 2001. 1~현재 (주)스몰소프트 부설 정보기술연구소 연구원
 E-mail : hstae@bioinfo.smallsoft.co.kr

김 기 봉



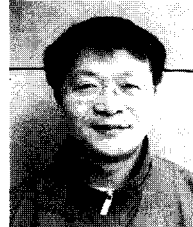
1985. 3~1992. 2 경북대학교 미생물학과 졸업
 1994. 3~1999. 2 한국과학기술연구원 생명공학연구소 연구원
 1995. 3~1997. 2 경북대학교 미생물학과 석사
 1998. 3~2001. 2 충남대학교 컴퓨터공학과 박사수료
 1999. 5~현재 (주)스몰소프트 대표이사 (실장/기술이사 역임)
 관심분야 : 생물정보학, 기계학습
 E-mail : kkbkim@bioinfo.smallsoft.co.kr

남 혜 원



1982. 3~1986. 2 이화여자대학교 전산
학과(이학사)
1986. 3~1990. 7 Computer Graphics
Korea(프로그래머/주임)
1997. 2~1998. 12 생명공학연구원 연
구원
1998. 3~2000. 2 충남대학교 정보과학
과(석사수료)
1999. 1~현재 (주)스몰소프트 부설 정
보기술연구소 선임연구원
E-mail : hwnam@bioinfo.smallsoft.co.kr

이 대 상



1989. 3~1993. 2 경북대학교 유전공학
과(이학사)
1993. 3~1995. 2 한국과학기술원 생물
과학과(이학석사)
1995. 3~1999. 7 한효과학기술원(연구원)
1999. 10~2000. 12 인바이오넷(연구원)
2002. 3~2003. 2 한국과학기술원 생물
과학과(박사수료)
2001. 1~현재 (주)스몰소프트 부설 정
보기술연구소 선임연구원
E-mail : dslee@bioinfo.smallsoft.co.kr

● 제17회 호남·제주지부 임시총회 및 하계학술대회 ●

- 일 자 : 2003년 8월 22일
- 장 소 : 광주대학교
- 주 최 : 호남·제주지부
- 논문접수마감 : 2003년 8월 1일
- 접 수 처 : 원광대 정성태 교수
E-mail : stjung@wonkwang.ac.kr
광주대 강문설 교수
E-mail : mskang@kwangju.ac.kr
- 문 의 처 : 호남·제주지부 사무국(Tel. 062-670-2387)