통합형 미생물 유전자 예측 시스템의 구축에 관한 연구

A Study on Construction of Integrated Prokaryotes Gene Prediction System

장 종 원*, 류 윤 규**, 구 자 효*, 윤 영 우*
Jong-won Chang*, Yoon-kyu Ryoo**, Ja-hyo Ku*, Young-woo Yoon*

요 약

유전자 서열 분석기의 발달로 유전체 서열 데이터는 급속도로 증가하여 자동적으로 유전체에 주석을 첨부하는 과정이 필요하다. 유전체에 주석을 다는 작업 중 가장 어려운 과정이 유전체내에 존재하는 단백질을 코드화하고 있는 유전자의 탐색이다. 진핵생물과 원핵생물은 유전자 구조에서 현격한 차이를 보이고 있으므로 유전자를 예측하는 방법도 각각 달라야 한다. 지금까지 전체 유전체 서열이 밝혀진 231종의 생물에서 200종이 원핵생물이다. 그러므로 비교 유전체학을 통한 생물공학 연구에서 진핵생물보다 원핵생물이 더 적합하다 할 것이다. 게다가 원핵생물의 경우 intron이라는 구조를 가지고 있지 않아 유전자 예측이 더 간단하다. 이전에 연구된 원핵생물의 유전자 예측 정확성은 80%~90%에 이르고 있고 최근의 연구에서는 유전자 예측 정확도 100%를 목표로 하고 있고, 본 논문에서는 *E. coli* K-12와 *S. typhi* 유전체의 경우, 유전자 예측 정확도가 각각 98.5%와 98.7%를 보여 기존의 GLIMMER보다 더 우수한 결과를 나타내었다.

Abstract

As a large quantity of Genome sequencing has happened to be done a very much a surprising speed in short period, an automatic genome annotation process has become prerequisite. The most difficult process among with this kind of genome annotation works is to finding out the protein-coding genes within a genome. The main 2 subjects of gene prediction are Eukaryotes and Prokaryotes ; their genes have different structures, therefore, their gene prediction methods will also obviously varies. Until now, it is found that among of the 231 genome sequenced species, 200 have been found to be prokaryotes, therefore, for study of biotechnology studies, through comparative genomics, prokaryotes, rather than eukaryotes could may be more appropriate than eukaryotes. Even more, prokaryotes does not have the gene structure called an intron, so it makes the gene prediction easier. Former prokaryotes gene predictions have been shown to be 80%~ to 90% of accuracy. A recent study is aiming at 100% of gene prediction accuracy. In this paper, especially in the case of the E. coli K-12 and S. typhi genomes, gene prediction accuracy which showed 98.5% and 98.7% was more efficient than previous GLIMMER.

Key words : Gene Prediction, Prokaryotes, GLIMMER, BLAST, HMMER,

I. 서 론

2001년 2월, 10여년을 끌어온 인간 유전체 사업(Human Genome Project, http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml)의 인간 유전체 지도 초안이 발표된 이후 전 세계적으로 생명공학에 대한 연구와 관심이 어느 때보다 활발하다. 이에 따라 유전체 서열의 분석 연구가 일반화됨으로써 생명현상의 기반이 되는 유전자의 대량 발굴과 기능 분석이 주요한 과제로 부각되고 이러한 생명현상의 연구로 도출되는 유전자나 단백질의 서열, 구조 등의 데이터의 정보를 만들고 이를 이용하는 생물정보학이라는 학문의 필요성이 강조되고 있다. 2004년 11월 현재 전체 유전체 서열의 분석이 완료된 생물은 231종이며, 976종의 생물종들에 대한 유전체 서열 분석이 세계적으로 이루어지고 있다(http://www.genomesonline.org/). 유전체 서열 분석이 완료된 이후, Genome annotation의 단계를 밟아야 한다. 유전체 서열 분석에서 얻어지는 서열 데이

*영남대학교 컴퓨터공학부 **대구보건대학 행정전산과
접수 일자 : 2004. 11. 18 수정 완료 : 2005. 1. 27
논문 번호 : 2004-3-8

터의 량이 많기 때문에 자동화된 annotation 과정이 필요하게 된다(그림 1). Genome annotation 작업 중에서 가장 중요하지만 어려운 작업이 genome 내에 존재하는 protein-coding genes을 알아내는 것이다. protein-coding gene은 전사(transcription)와 번역(translation)이라는 과정을 거쳐 단백질로 발현됨으로써 생체의 모든 생명현상의 근간을 이루게 된다. 이렇듯 유전체 내에서 protein-coding gene을 예측하는 작업을 유전자 예측(Gene prediction)이라 한다. 일반적인 유전자 예측 기법은 기존의 알려진 유전자 정보를 가진 데이터베이스(GenBank)와 비교를 통해 유전자를 판별하는 방법을 사용하였다[1]. BLAST 등의 서열정렬 프로그램을 이용하여 기존 데이터베이스의 서열과 비교하여 유전자를 찾는 상동성 기반(homology based)의 방법은 false positive가 낮고 정확한 예측이 가능하다는 장점을 가지는 반면에 새로운 유전자의 탐색이라는 목표는 이루기 힘들다. 유전자 예측의 다른 한 방법은 유전자라고 정의될 수 있는 특성들을 이용하여 유전자 부위를 판별하는 방법으로 *ab-initio* 방법이라 불리운다[2]~[4]. 유전자의 개시코드(AUG)와 종료코드(UAG, UAA, UGA), 유전자의 프로모터, 라이보좀 바인딩 사이트(RBS, Ribosome binding sites, SD sequence) 등의 특성을 고려하여 유전자를 결정하는 이 방법은 알려지지 않은 새로운 유전자를 찾을 가능성을 열어두고 있다. 그러나 유전자가 아님에도 유전자로 판별하는 false positive가 높다는 단점이 있다.

유전자 예측을 하고자 할 때 대상 생물을 크게 진핵생물(Eukaryotes)과 원핵생물(Prokaryotes)로 구분하게 되며 두 생물의 유전자 구조가 현격한 차이를 가지기 때문에 하나의 방법을 모든 생물종에 대하여 사용할 수 없다. 두 대상 생물의 가장 큰 차이는 intron 부위의 존재 유무이다. 진핵생물의 경우에는 실제 유전자를 코딩하지 않은 intron 부위를 제외시키는 알고리즘이 선행되어야 하며 원핵생물의 경우 intron 이라는 부위가 없어 유전자 예측에 용이한 면을 보여준다. 실제로 현재까지 개발된 여러 방법들을 보면 진핵생물의 유전자 예측 신뢰도는 40%에 미치지 못하지만 원핵생물의 경우 최고 약 80~90%의 정확도를 보이고 있다. 현재까지 전체 유전체 서열이 분석된 231종 중 200종이 세균류, 아케아 등의 미생물로써 기존 데이터베이스와의 비교를 통해 새로운 단백질의 기능 등을 유추하는 비교 유전체학을 통한 생명현상의 연구에 미생물이 진핵생물보다 더 적절하다. 그리고 미생물 유전자 정보의 활용은 신약개발뿐만 아니라 환경, 화학, 생물소재 분야의 새로운 공정 및 제품 개발, 개선에 큰 파급효과를 미치고 있다.

이에 본 논문에서는 미생물 DNA 서열에 대한 유전자를 탐색하기 위하여 *ab-initio* 방법을 적용하여 현재까지 보고된 연구 중 가장 높은 정확도를 보이는 GLIMMER 프로그램[2][3]의 진행단계 중 ORF(Open Reading Frame)를 이용하여 학습하는 모델과 함께 각 생물 종에 따라 미리 최적으로 학습한 유전자 모델을 생성하는 방법, 진화학적으로 근연 관계에 있는 다른 종과의 합병 모델 을 생성하는 방법 등을 통해 기존 GLIMMER 보다 더 우수한 유전자 특성에 기반한 *ab-initio* 방법을 수행하였으며 이와 동시에 기존에 알려진 유전자를 BLAST를 이용하여 상동성에 기반한 유전자 예측 방법도 수행하였다. 이로써 미생물을 대상으로 하는 유전자 예측 시스템을 구축하고 예측된 유전자의 기능을 기존 데이터베이스와 비교하여 그 기능까지 예측할 수 있는 시스템을 개발하였다.

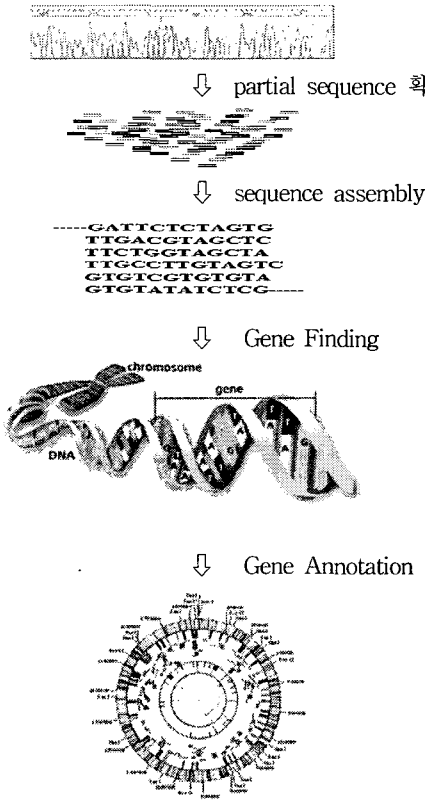

그림 1. 서열 분석의 단계
Fig 1. Step of sequence analysis

II. 유전자 예측 시스템 구축 방법

2.1 학습모델의 생성

Markov model을 이용하는 GLIMMER의 경우 유전자 예측 방법을 살펴보면 첫 단계로 질의 염기 서열에서 500 base 이상의 ORF 위치를 찾는 long ORF라는 프로그램을 실행하고 두 번째 단계인 질의 서열에서 찾은 ORF를 파싱하는 extract, 세 번째 단계에서는 extracted ORF sequence를 이용하여 학습모델을 만드는 build-icm, 마지막 네 번째 단계로 질의 서열과 학습모델을 비교하여 유전자 부위를 찾아내는 glimmer2라는 과정을 거쳐 결과 값을 내보낸다[2].

본 연구에서는 기존의 GLIMMER를 다음의 방법을 이용하여 학습모델을 개선하고자 하였다(그림 2).

개선 1 : 기존의 long ORF를 이용한 학습 모델은 질의 서

열이 크지 않은 경우 잘못된 학습 모델을 생성할 가능성이 매우 크다. 질의 서열에서 long ORF를 찾아 학습모델을 생성하는 기존의 방법 이외에 각 질의 서열이 알려진 생물종이라면 그 생물종의 알려진 유전자의 서열 리스트로 학습하는 방법과 전체 유전체를 대상으로 long ORF를 찾아 미리 학습모델을 구축하여 제공하므로 정확성이 향상된 학습모델을 만들어 제공한다.

개선 2 : 개선 1의 방법을 계통 분류학적으로 다른 종의 유전자 리스트나 long ORF를 병합하여 학습모델을 제공함으로써 생물 유전자의 다양성을 충족시키므로 기존의 방법으로 찾지 못했던 새로운 유전자모델을 찾는다.

개선 3 : 최초로 만들어진 학습 모델을 glimmer2 프로그램을 이용하여 유전자 부위를 예측하여 그 결과를 상동성 비교 프로그램인 BLASTX로 기존의 알려진 유전자와 비교하여 false positive를 제거한 유전자 리스트를 다시 학습모델로 이용하는 방법을 반복하여 정확성을 향상시키고자 하였다.

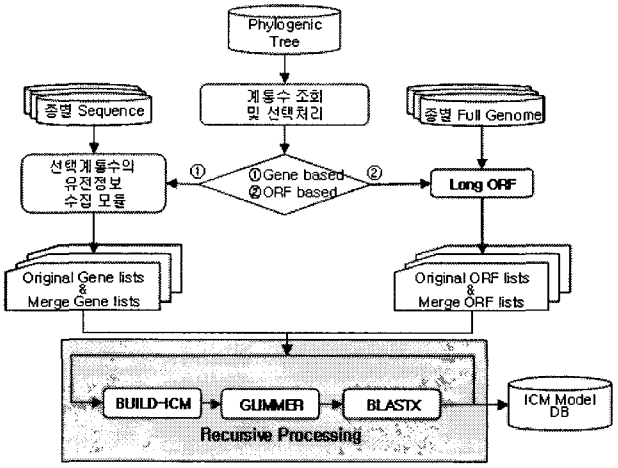

그림 2. 학습모델 생성 흐름도

Fig 2. Build training model flowchart

2.2 *Ab-initio* 처리 프로세서

기존의 GLIMMER의 경우 학습모델을 만들어 glimmer2 프로그램을 이용하여 질의 서열에서 유전자 부위를 예측한 결과를 나타낸다. 본 연구에서는 후보 유전자의 서열 정보 리스트를 BLASTX를 이용하여 스코어(bit)와 기댓값(expectation value)을 사용자가 기준을 정해 스코어 값이 높고 e-value가 낮은 경우 결과를 파싱하여 후보유전자 데이터베이스에 기록하고 스코어가 낮고 e-value가 높은 경우 후보 유전자에서 제외하는 것이 아니라 HMMER라는 프로그램을 이용하여 Protein family 데이터베이스인 PFam A 데이터베이스와 정렬과정을 거쳐 스코아가 낮더라도 지역적으로 높은 정렬값을 가진다면 단백질 도메인을 찾아 후보 유전자 데이터베이스에 기록하여 새로운 유전자 부위의 true negative를 줄이고자 하였다(그림 3).

2.3 상동성(homology) 기반 처리 프로세서

이미 알려져 있는 유전자와의 비교는 BLAST 프로그램 중 질의 DNA 서열과 DB내에 있는 DNA 서열을 비교하여 탐색하는 BLASTN과 질의 DNA 서열을 아미노산 서열로 바꾸어 아미노산 DB에 있는 서열을 비교하여 탐색하는 BLASTX를 이용하여 비교하였다. BLASTN은 GenBank에 있는 NT DB, BLASTX는 swiss-prot DB를 이용하여 상동성 탐색을 실시하였다(그림 4).

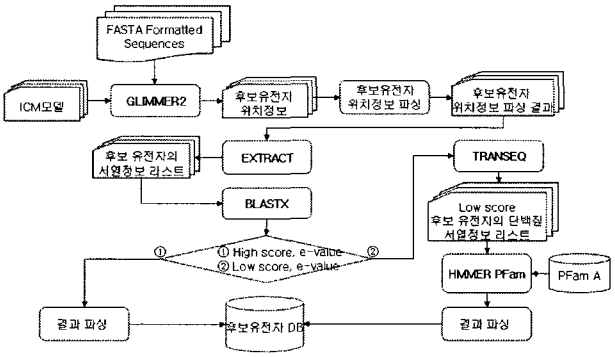

그림 3. *Ab-initio* 처리 흐름도

Fig 3. *Ab-initio* processing flowchart

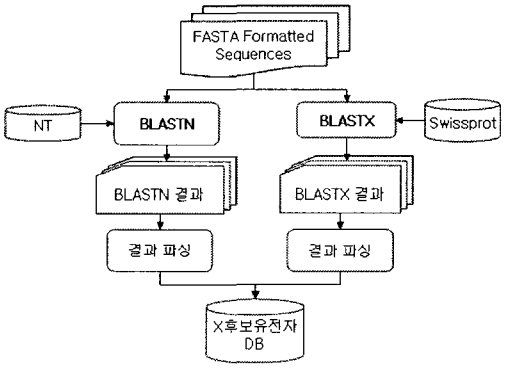

그림 4. Homology 처리 흐름도

Fig 4. Homology processing flowchart

2.4 시스템의 구축

본 연구에 사용된 GLIMMER2.02와 BLAST 2.2.6 프로그램은 Unix 기반의 프로그램이고 BLAST의 경우 windows 용이 배포되고 있다. 그러나 모두 text 기반의 명령어로 실행하는 도구로서 컴퓨터에 익숙하지 않은 생물학자들이 사용하기 쉽도록 각

각의 프로그램을 GUI 환경의 windows application용 프로그램으로 포팅하여 사용하였다. 사용된 컴퓨터는 운영체제 Window XP professional, CPU는 Intel Pentium Ⅳ 2.6GHz, 512RAM의 일반 PC에서 시험하였다.

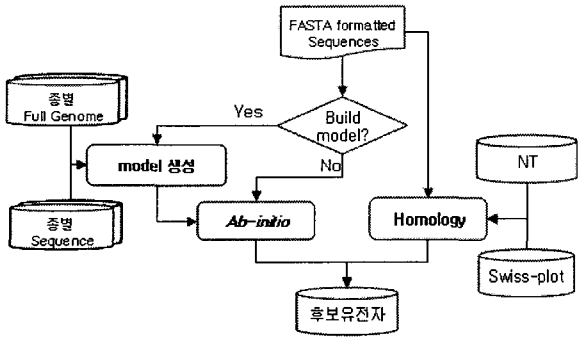

그림 5. 유전자 예측 시스템 전체 흐름도

Fig 5. Overall flowchart of gene prediction system

Ⅲ. 유전자 예측 시스템 평가 방법

3.1 Material

GenBank(ftp://ftp.ncbi.nih.gov/genbank/)에서 시스템을 평가하기 위한 전체 유전체 서열 정보 데이터와 정보를 얻었다. 평가에 사용된 생물종은 *Escherichia coli* K-12*(E.coli* K-12*), Salmonella enterica serovar typhi(S.typhi)* 등으로서 기존의 GLIMMER와의 성능 비교를 위하여 GLIMMER의 평가에 사용된 종을 이용하여 테스트하였다. 그리고 참고로 사용된 유전체는 NCBI(National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/)의 분류 기준(http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi)에 의하여 15종을 선정(그림 6) 하였으며 GenBank database에서 획득하였고 시험에 사용된 참고 유전체의 정보를 표 1.에 나타내었다.

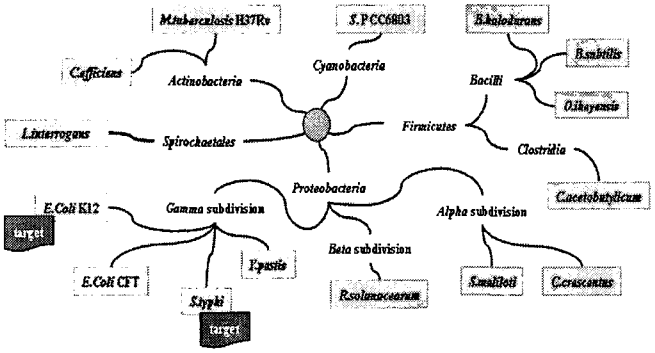

그림 6. 원핵생물 유전체의 진화학적 계통수

Fig 6. Phylogenetic tree of Prokaryotes genomes

3.2 Gene Prediction Accuracy 측정

개선된 유전자 예측 시스템의 평가는 기존의 GLIMMER와 개선된 유전자 예측 시스템에 각각 *E. coli* K-12와 *S. typhi*의 전체 유전체 서열을 질의 서열로 넣어 나타난 결과 값을 Gen Bank에 등록되어 있는 유전자들과 비교하여 평가하였다. 정확도(Accuracy)와 Additional predicted gene rate(%)의 측정은 다음과 같은 방법으로 평가한다.

$$\text{Accuracy}(\%) = 100 \times \frac{\text{Number of genes predicted correctly}}{\text{Number of genes annotated in GenBank}}$$

$$\text{Additional predicted gene rate}(\%) = 100 \times \frac{\text{Number of predicted genes that don't appear in the annotation}}{\text{Number of genes annotated in GenBank}}$$

표 1. 평가에 사용된 생물종의 정보 요약

Table 1. Summary of information about the reference genomes

Reference species	Genome size (Mb)	GC contents(%)	No. of ORFs(>500bp)
Escherichia coli K-12	4.70	50.8	2,416
Salmonella enterica serovar typhi	4.87	52.1	2,581
Corynebacterium efficiens YS-314	3.19	63.1	397
Mycobacterium tuberculosis H37Rv	4.47	65.6	602
Synechocystis PCC6803	3.62	47.7	2,188
Bacillus halodurans	4.06	43.7	2,803
Bacillus subtilis	4.07	43.5	2,710
Oceanobacilius iheyensis	3.68	35.7	2,486
Clostridium acetobutylicum	3.99	30.9	2,658
Caulobacter crescentus	3.83	67.2	317
Sinorhizobium meliloti	3.70	62.7	275
Ralstonia solanacearum GMI1000	3.76	67.0	151
Escherichia coli CFT073	5.30	50.5	2,719
Yersinia pestis KIM	4.66	47.6	2,754
Leptospira interrogans serovar 56601 Chromosome I	4.39	35.0	2,454

표 2. GLIMMER2.02에 의해 예측된 두 원핵생물 유전체의 true gene과 false gene의 수

Table 2. The number of true and false genes found by GLIMMER2.02 for 2 prokaryotic genomes

Interesting Prokaryotic Genomes	Annotated Genes	No. of true gene predicted by GLIMMER	Accuracy (%)	No. of false genes predicted by GLIMMER	Additional gene prediction rate(%)
E.coli K-12	4,279	4,167	97.3	597	13.95
S.typhi	4,395	4,300	97.8	1,123	25.55

Ⅳ. 결과 및 평가

4.1 기존의 GLIMMER 결과

GLIMMER(ver 2.02)를 이용하여 기본 설정으로 *E.coli* K-12 와 *S.typhi*의 전체 유전체 서열을 분석하였다. 그 결과 *E.coli* K-12의 경우, GenBank에 등록되어 있는 유전자 4,279개 중 4,167개를 예측하여 약 97.3%의 정확도를 보이고 알려지지 않은 유전자 예측한 false gene이 597개 이였다. 또한 *S.typhi*의 경우에는 4,395개의 등록된 유전자 중 4,300개의 유전자를 예측하여 정확도 97.8%를 보이고 false gene은 1,123개로서 25.55%로 나타나 유전자의 예측율이 기존 논문[2][3]의 결과와 같음을 확인하였다. 이를 표 2에 나타내었다.

4.2 개선된 유전자 예측 시스템 결과

E.coli K-12와 *S.typhi*의 유전체를 표 1에 표시한 15종의 균주들과 병합하여 모델을 생성하는 개선된 유전자 예측 시스템을 이용(그림 4)하여 시험하였다. 다시 말하면 각 reference species의 유전체에서 genelist를 추출한 후 시험 균주 *E.coli* K-12와 *S.typhi*의 genelist와 병합하여 학습모델로 이용하여 유전자 예측 시스템을 시험하였다. 그 결과 표 3에서 보는 바와 마찬가지로 평균 예측 정확도가 *E.coli* K-12와 *S.typhi*에서 각각 97.6%, 97.8%로 기존 GLIMMER 보다 증가한 것을 확인할 수 있었다. 또한 GenBank에 등록되지 않은 새로운 유전자라 할 수 있는 false positive prediction rate도 각각 16.3%, 27.6%로 기존의 방법(13.95%, 25.55%)보다 더 증가되었다. 특히 target 균주와 병합한 reference 균주 중 *Bacilli* 그룹에서 98.5%, 98.7%의 우수한 예측 정확성을 보였다. 또한 기존의 방법으로는 찾지 못한 GenBank에 등록된 유전자도 표 4에서 나타난 바와 같이 더 찾을 수 있어 개선된 시스템이 유효함을 증명하였다.

그림 7에서는 *E.coli* K-12의 유전체를 입력 sequence로 하여 유전자 예측 시스템을 거친 후 결과 화면을 나타내었다. 각 화면은 질의 서열과 *ab-initio* 결과, BLASTN을 수행한 결과, BLASTX를 수행한 결과를 나타낸다.

Ⅴ. 향후 계획

본 논문에서는 미생물에서 유전자 부위를 예측하는 시스템을 제안하였다. 유전자의 특성을 이용하여 유전자 부위를 예측하는 *ab-initio* 방법과 알려져 있는 유전자의 서열과 상동성 비교를 통하여 유전자 부위를 예측하는 방법을 함께 사용하여 예측 정확성을 높이고 새로운 유전자 부위도 더 많이 예측 할 수 있음을 실험을 통하여 입증하였다. 이에 선택적인 스플라이싱 등의 문제를 해결하여 진핵생물로 유전자 예측의 범위를 확장시키는 연구가 계속 되어져야 할 것이다. 또한 생물정보학은 생물학과 컴퓨터과학의 경계에 있는 학문으로 컴퓨터 기술을 이용하여 생물학 데이터를 저장, 분석 및 해석하는 분야라고 할 수 있다. 지금까지 개발된 생물정보학 도구들을 살펴보면 서열정렬, 유전자 분석, 단백질 구조 분석, 단백질의 기능 예측 등 생물학의 전 분야에 대해서 개발되고 있다. 그러나 DNA에서부터 단백질에 이르기까지 하나로 통합된 도구가 없는 실정이다. 생물학의 전 분야에 걸쳐 자동화되고 사용자의 편의를 고려한 도구가 제공된다면 생물학자들의 연구 기간 단축은 물론 새로운 연구 분야의 개척도 제공할 수 있을 것이다. 이에 유전체학과 단백질체학을 통합하여 서열, 구조, 기능 등을 함께 비교 분석할 수 있는 시스템의 구축이 시급하다 할 것이다.

표 3. 개선된 방법에 의한 *E.coli* K-12 와 *S.typhi* 유전체의 유전자 예측 정확도 및 false 유전자 예측비(%)

Table 3. Gene prediction performance of the evolutionary model for the *E.coli* K-12 and *S.typhi* genomes

Reference species	No. of Genes predicted by evolutionary methods		No. of false Genes predicted by evolutionary methods	
	E.coli K-12	*S.typhi*	*E.coli K-12*	*S.typhi*
Average	4177(97.6)	4300(97.8)	697(16.3)	1214(27.6)
S.typhi	4165(97.8)		604(14.1)	
E.coli K-12		4298(97.8)		1090(24.8)
C.efficiens YS-314	4171(97.5)	4302(97.9)	619(14.5)	1102(25.1)
M.tuberculosis H37Rv	4165(97.3)	4290(97.6)	559(13.1)	1030(23.4)
Synechocystis PCC6803	4168(97.4)	4229(96.2)	611(14.3)	1127(25.6)
B.halodurans	4216(98.5)	4337(98.7)	886(20.8)	1379(31.4)
B.subtilis	4213(98.5)	4340(98.7)	919(21.5)	1462(33.3)
O.iheyensis	4196(98.1)	4330(98.5)	899(21.0)	1361(31.0)
C.acetobutylicum	4167(97.3)	4275(97.3)	823(19.2)	1264(28.8)
C.crescentus	4172(97.5)	4297(97.8)	585(13.7)	1049(23.8)
S.meliloti	4181(97.7)	4300(97.8)	642(15.0)	1119(25.5)
R.solanacearum GMI1000	4175(97.6)	4295(97.7)	635(14.9)	1107(25.2)
E.coli CFT073	4177(97.6)	4303(97.9)	588(13.7)	1171(26.6)
Y.pestis KIM	4139(96.7)	4305(98.0)	604(14.1)	1499(34.1)
L.interrogans serovar 56601 Chromosome I	4169(97.4)	4295(97.7)	781(18.3)	1242(28.3)

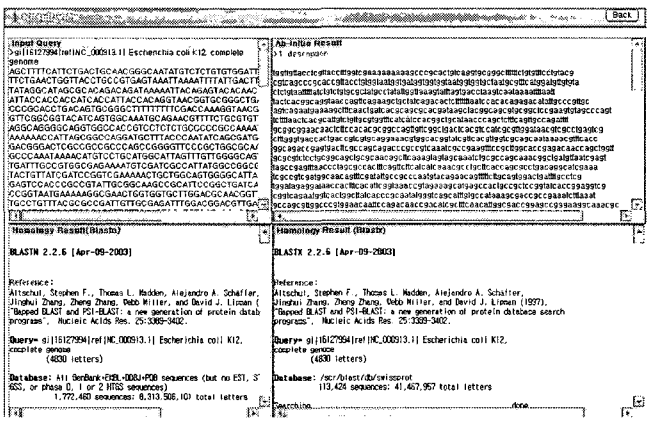

그림 7. 유전자 예측 시스템 결과 예

Fig 7. A result of gene prediction system

표 4. 개선된 방법에 의한 true 유전자 및 missing 유전자 탐색 개수의 비교
Table 4. Comparison of the number of genes found only by the evolutionary method and the current GLIMMER

Reference species	No. of New True Genes identified by evolutionary methods		No. of missing Genes in evolutionary methods	
	E.coli K-12	*S.typhi*	*E.coli* K-12	*S.typhi*
S.typhi	14		16	
E.coli K-12		20		22
C.efficiens YS-314	12	12	8	9
M.tuberculosis H37Rv	10	7	12	16
S.ynechocystis PCC6803	32	25	33	25
B.halodurans	55	50	6	12
B.subtilis	48	52	2	12
O.iheyensis	45	47	19	16
C.acetobutylicum	32	38	32	62
C.crescentus	9	8	4	11
S.meliloti	18	8	4	8
R.solanacearum GMI1000	11	5	3	10
E.coli CFT073	16	18	6	15
Y.pestis KIM	21	25	9	20
L.interrogans serovar 56601 Chromosome I	35	36	30	40

참 고 문 헌

[1] Altschul SF, Madden TL, Schaffer AA, Zhang Z, Miller W, Lipman DJ. "Gapped BLAST and PSI-BLAST : a new generation of protein database search programs." *Nucleic Acids Research,* Vol. 25, pp. 3389~3402, 1997

[2] Salzberg SL, Delcher AL, Kasif S, White O. "Microbial gene identification using interpolated Markov models," *Nucleic Acids Research.* Vol. 26, pp. 544~548, 1998

[3] Delcher AL, Kasif S, White O. Salzberg SL. "Improved Microbial gene identification with GLIMMER," *Nucleic Acids Research.* Vol. 27, pp. 4636~4641, 1999

[4] Besemer J, Borodovsky M. "Heuristic approach to deriving models for gene finding." *Nucleic Acids Research.* Vol. 27 pp. 3911-3920, 1999

장 종 원(Jongwon Chang)
1992년 2월 영남대 응용미생물(이학사)
1994년 2월 영남대 미생물학과(이학석사)
2002년 8월 영남대 컴퓨터공학(공학석사)
2004년 8월 영남대 컴퓨터공학(박사과정 수료)
주관심분야 : 생물정보학, 영상 처리

류 윤 규(Yoonkyu Ryoo)
1995년 2월 영남대 경영학과(경영학사)
1997년 8월 경일대 컴퓨터공학(공학석사)
2003년 8월 영남대 컴퓨터공학(박사수료)
1998년 3월~현재 대구보건대학 행정전산과 교수
주관심분야 : 영상처리, 전자상거래, 생물정보학

구 자 효(Jahyo Ku)
2000년 2월 계명대 컴퓨터공학과(공학사)
2002년 8월 영남대 컴퓨터공학(공학석사)
2005년 2월 영남대 컴퓨터공학(박사수료예정)
주관심분야 : 생물정보학, 영상처리

윤 영 우(Youngwoo Yoon)
1972년 2월 영남대 전자공학과(공학사)
1983년 2월 영남대 전자공학과(공학석사)
1988년 2월 영남대 전자공학과(공학박사)
1988년 9월~현재 영남대 컴퓨터공학과 교수
주관심분야 : 영상처리, 생물정보학, 컴퓨터 설계