

다종의 유전체로부터 탐지된 Ortholog 군집에 대한 분석

(An Analysis of Ortholog Clusters Detected from Multiple Genomes)

김 선 신 [†] 오 정 수 ^{**} 이 범 주 ^{***} 김 태 경 ^{****} 정 광 수 ^{****}
(Sun Shin Kim) (Jeongsu Oh) (Bum Ju Lee) (Tae Kyung Kim) (Kwang Su Jung)

이 충 세 ^{*****} 김 영 창 ^{*****} 조 완 섭 ^{*****} 류 근 호 ^{*****}
(Chung Sei Rhee) (Young Chang Kim) (Wan Sup Cho) (Keun Ho Ryu)

요 약 새로운 유전체 주석달기와 유전체 진화에 대한 연구를 위해서 올소로그(Ortholog)를 탐지하는 일은 매우 유용하다. 이전에 제안한 연구에서, 우리는 여러 종의 유전체로부터 올소로그 클러스터를 자동적으로 구축하는 방법을 제안하였다. 이 방법은 단지 두 종의 결과를 생성하는 InParanoid를 여러 종으로 확장하고 이와 동일한 질을 가진 결과를 산출한다. 한편, 새롭게 서열이 밝혀진 유전자의 기능을 보다 정확히 예측하기 위해, 패럴로그(Paralog)가 가급적 적게 포함되는 올소로그 클러스터를 구축하는 것이 중요한 문제가 될 수 있다. 이 논문에서, 우리는 임계값을 사용하여 보다 순수한 올소로그 클러스터를 구축하는 방법에 대하여 조사하였다. 우리는 20개의 원핵생물의 데이터셋으로부터 올소로그 클러스터를 구축하였다. 우리의 올소로그 클러스터를 COG(Clusters of Orthologous Group) 및 KO(Kegg Orthology)와 비교하였을 때, 약 90%의 유사도를 가지며 임계값의 증가와 더불어 증가하는 경향이 있다.

키워드 : 올소로그(ortholog), 유전체, InParanoid, 패럴로그(paralog), COG, 및 KO

Abstract It is very useful to predict orthologs for new genome annotation and research on genome evolution. We showed that the previous work can be extended to construct OCs(Ortholog Clusters) automatically from multiple complete-genomes. The proposed method also has the quality of production of InParanoid, which produces orthologs from just two genomes. On the other hand, in order to predict more exactly the function of a newly sequenced gene it can be an important issue to prevent unwanted inclusion of paralogs into the OCs. We have, here, investigated how well it is possible to construct a functionally purer OCs with score cut-offs. Our OCs were generated from the datasets of 20 prokaryotes. The similarity with both COG(Clusters of Orthologous Group) and KO(Kegg Orthology) against our OCs has about 90% and inclines to increase with the growth of score cut-offs.

Key words : ortholog, genome, InParanoid, paralog, COG, and KO

· 이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (지방연구중심대학육성사업/충북BIT연구중심대학육성사업단)

[†] 비 회 원 : 국립보건의연구원

sskim04@hotmail.com

^{**} 정 회 원 : 국가생물자원정보관리센터

ofang@kribb.re.kr

^{***} 비 회 원 : 충북대학교 전자계산학

bjlee@dblab.chungbuk.ac.kr

ksjung@dblab.chungbuk.ac.kr

^{****} 학생회원 : 충북대학교 정보산업공학

tkkim@chungbuk.ac.kr

^{*****} 종신회원 : 충북대학교 전기전자컴퓨터공학 교수

csrhee@chungbuk.ac.kr

khryu@dblab.chungbuk.ac.kr

^{*****} 비 회 원 : 충북대학교 생명과학 교수

youngkim@chungbuk.ac.kr

^{*****} 종신회원 : 충북대학교 경영정보학 교수

wscho@chungbuk.ac.kr

논문접수 : 2005년 4월 25일

심사완료 : 2007년 9월 27일

Copyright © 2008 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 데이터베이스 제35권 제2호(2008.4)

1. 서론

현재까지 아미노산서열의 수는 기하급수적으로 증가하여 전 세계적으로 다양한 데이터베이스에 축적되고 있지만, 그들 대부분의 기능은 알려져 있지 않다. 특히, GOLD(Genomes OnLine Database)에는 유전자 지도가 완전히 밝혀진 유전체가 400여종 이상 저장되어 있다. 이들 유전체로부터 단백질서열 상동성에 기반하여 올소로그(Ortholog)를 탐지하여 올소로그 클러스터(Ortholog Cluster)를 구축하는 일은 생물정보학에서 매우 중요하다.

단백질의 기능을 생물학실험을 통하여 알아내는 일은 많은 시간과 노력이 필요하기 때문에, 이미 기능이 알려진 단백질(또는 유전자)로부터 아직 기능이 알려지지 않은 단백질의 기능을 예측하는 일은 매우 유용하다. 이를 위해 이미 유전체서열이 완전히 밝혀진 유전자(또는 단백질)의 상동성에 기반하여, 동일한 기능을 할 것으로 예측되는 단백질(즉 올소로그)를 클러스터링하는 연구 [1-6]가 진행되어 왔다.

상동성(homology)을 가진 유전자를 상동체(homolog)라고 하는데 여기에는 두 가지 종류의 상동체가 있다. 하나는 올소로그[7]라고 하는데, 이는 같은 조상으로부터 본래의 기능을 간직한 채 종 분화를 통하여 진화한 유전자이다. 다른 하나는 패럴로그[7]라고 하는데, 이는 유전자 복제를 한 후 종 분화를 통하여 서로 다른 기능을 가지게 진화한 유전자이다. 그림 1에서 보여준 예를 고려하여 살펴보자. 초기단계에서 임의의 유전체에 속하는 임의의 유전자를 가정해보자. 이 유전자는 처음에 종 분화한 후 시간과 함께 진화하여, 중간단계에서 왼쪽에 있는 종은 그대로 진화하고 오른쪽에 있는 종은 유전자복제를 한 후 다시 종 분화와 함께 진화한다. 이제 현시점에서 A-B-D 및 C-E는 서로 올소로그 관계에 있게 된다. 한편, A-C, A-E, B-E, 및 C-D는 패럴로그 관계에 있게 된다.

위의 예에서 살펴본 것처럼 올소로그 클러스터를 만들 때, 가능하면 패럴로그를 배제 할수록 보다 순수한 그룹을 형성하게 된다. 여기에서 우리는 점수 임계값(score cut-off)의 변화에 따라 얼마나 순수한 올소로그 그룹을 만들 수 있는지를 조사한다.

절 2에서 관련연구에 관해 살펴보고 실험방법은 절 3에서 나타낸다. 실험결과 분석은 절 4에서 주어지고 절 5에서 우리의 일에 대해 토의 한다.

2. 관련 연구

COGs(Clusters of Orthologous Groups of proteins) [2]는 유전자 지도가 완성된 21개의 박테리아(bacteria), 아카이아(archaea), 및 진핵생물(eukaryotes)로부터 클러스터링하였다. 이들은 두 유전체 사이의 상호 최대 BLAST

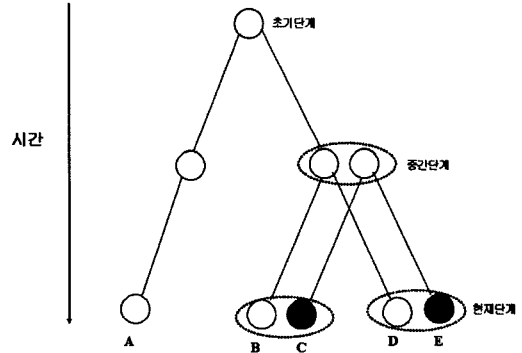


그림 1 종 분화와 유전자복제를 통하여 진화하는 과정

[8]히트(hit)를 가진 단백질쌍을 올소로그로 가정하였다. 이렇게 형성된 단백질쌍의 삼각형형태를 탐지하여 기본 올소로그 클러스터를 생성한다. 이제 삼각형의 한 변이 공통인, 즉 단백질쌍을 공유한 경우, 삼각형들을 통합하여 동일한 올소로그 클러스터를 구성한다. 또한 형성된 클러스터를 생물학적 분석을 통하여 재 분류한다. 이와 같은 알고리즘을 사용하여 현재까지 서열화가 완성된 66종의 유전체로부터 COG 데이터베이스[3]를 구축하였다. 하지만 몇 개의 형성된 클러스터는 한 개의 종으로부터 추출된 패럴로그가 많이 포함되어, 정확한 올소로그를 예측하는데 어려움이 있다. 또한 수작업으로 인하여 시간과 노력이 많이 소모된다.

이와 같이 형성된 올소로그 클러스터의 질은 BLAST 알고리즘에 대개 의존한다. 이 BLAST는 휴리스틱(heuristic)방법을 사용하여 매우 효율적이고 빠르지만[9-13], 진화적으로 거리가 먼 종들 사이에서 상동성을 탐지하는 것은 매우 어려운 것으로 알려져 있다[14-17].

또한, KO(Kegg Orthology)[18]는 대사(metabolic) 및 조절(regulatory) 회로 (pathway)로부터 기능이 확실히 밝혀진 유전자만을 올소로그 그룹으로 만들었기 때문에, 기능적으로 정확하게 분류되어 있다. 하지만 기능이 알려진 유전자의 수는 제한되어 있으므로 충분한 양의 올소로그를 함유하지 못하고 있다.

이와 같은 문제를 해결하기 위해 자동화된 InParanoid[4]프로그램이 제안되었다. 이 프로그램은 단지 두 개의 유전체에서 주된 올소로그쌍과 추가적인 올소로그(즉 inparalog)를 자동적으로 탐지한다. 또한 두 종류의 임계값을(즉 score cut-off와 overlap cut-off) 사용함으로써 패럴로그를 줄일 수 있어서, 보다 순수한 올소로그 클러스터를 탐지할 수 있었다. 그러나 이 방법은 단지 두 종 사이에서만 올소로그를 탐지한다.

또한, Li[5]등은 MCL알고리즘[19]을 사용하여 다종으로부터 올소로그를 탐지하는(즉 InParanoid를 확장한)

OrthoMCL 프로그램을 개발하였다. 이들도 올소로그 탐지의 정확성을 높이기 위해 임계값(즉 P-value cutoff) 설정하였다. 또한 이 프로그램의 결과는 InParanoid와 비슷한 결과를 산출함을 보여주었다. 그러나 이 결과를 생산하는데 InParanoid보다 약 두 배 더 많은 시간이 필요하였다[5].

이런 사실에 기반을 두고, 이전에 제안 했던 연구에서 [6] 우리는 다종으로부터 올소로그를 탐지하는 자동화 방법을 제안하였다. 이 방법은 InParanoid에 기반을 두고 이를 확장한 것으로 OrthoMCL보다 더 빠르게 올소로그를 탐지할 수 있는 알고리즘이다. 즉, 우선 임의의 n개의 유전체를 선정한다. 이 n개의 유전체가 가질 수 있는 모든 가능한 유전체쌍으로부터(즉, $n(n-1)/2$ 개), InParanoid 프로그램을 사용하여, 올소로그를 탐지하여 $n(n-1)/2$ 개의 올소로그 테이블에 저장한다. 이들 테이블로부터 공통의 올소로그를 가진 올소로그쌍을 찾아 통합한다. 따라서 우리의 방법이 산출한 결과는 InParanoid의 결과와 동일하다.

이 논문에서, 우리는 앞서 제안한 방법의 결과를, 임계값을 변화시키면서, COG 및 KO와 비교하여 분석하고 토의 한다.

3. 실험 방법

우리가 산출한 올소로그 클러스터를 조사하기 위해, COG와 KO 양쪽에 공통으로 속하는 서열화가 완성된 20개의 유전체를 임의로 선정하였다. 그림 2는 박테리아와 아키아로 이루어진 유전체를 나타낸다. 각 유전체는

실험에서 사용된 유전체

1. *Aquifex aeolicus* VF5
2. *Bacillus subtilis* subsp. *subtilis* str. 168
3. *Campylobacter jejuni* subsp. *jejuni* NCTC 11168
4. *Caulobacter crescentus* CB15
5. *Chlamydomonada pneumoniae* CWL029
6. *Deinococcus radiodurans* R1
7. *Escherichia coli* K12
8. *Haemophilus influenzae* Rd KW20
9. *Helicobacter pylori* 26695
10. *Lactococcus lactis* subsp. *lactis*
11. *Mesorhizobium loti* MAFF303099
12. *Neisseria meningitidis* MCS8
13. *Pasteurella multocida*
14. *Pseudomonas aeruginosa* PAO1
15. *Synechocystis* sp. PCC 6803
16. *Vibrio cholerae* O1 btovar *eltor* str. N16961
17. *Xylella fastidiosa* 9a5c
18. *Archaeoglobus fulgidus* DSM 4304
19. *Methanocaldococcus jannaschii* DSM 2661
20. *Thermoplasma volcanium*

그림 2 유전자 서열이 완전히 밝혀진 20개의 박테리아(bacteria)와 아키아(archaea)

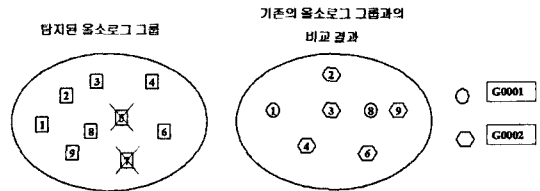


그림 3 탐지된 올소로그 그룹에 대한 기존의 올소로그 그룹과의 비교

FASTA 포맷으로 되어 있는 단백질서열의 데이터집합이다. 우리가 생성한 올소로그 클러스터에서 임의의 샘플을 무작위로 뽑아서 COG 및 KO와 비교하였다.

우리가 새롭게 만든 그룹과 기존의 임의의 그룹을 비교할 때, 우리의 그룹이 기존의 어느 그룹과 더 유사도가 큰지를 결정한다고 하자. 그림 3에서 보여주는 것처럼, 동일한 그룹의 각 요소는 같은 형태의 모양을 가진다고 하자. 왼쪽의 탐지된 우리의 그룹과 오른쪽의 기존의 그룹과 비교할 때, 탐지된 우리의 그룹 중에 요소 5와 7은 기존의 그룹에서 발견되지 않는다. 이 경우 기존의 그룹과 일치하지 않는 요소 5와 7은 배제된다. 육각형 형태의 그룹 G0002의 5개 요소 및 원형의 그룹 G0001의 2개 요소가 우리가 생성한 그룹 안에서 확인되었다. 이제 육각형의 그룹 G0002가 우리의 그룹과 가장 많은 요소가 일치 하고 있기 때문에, 이와 같이 우위를 차지하는 기존의 그룹을 우리의 그룹과 동일시하여 유사도를 측정하도록 정의한다. 즉, 기존의 그룹 요소 중에서 우리의 그룹 요소와 가장 많이 일치하는 요소의 수를 C라고, 기존에 존재하는 모든 그룹의 요소와 일치하는 모든 우리의 그룹 요소를 N이라고 할 때, 기존의 그룹과 우리의 그룹의 유사도를 $S=C/N$ 으로 정의한다. 이 경우 그림 3의 예에서 보여준 유사도는 5/7이 된다.

이제 우리가 생성한 클러스터의 실 예를 살펴보기로 하자. 그림 4는 제안된 방법을 사용하여 얻은 10개의 단백질로 구성된 OPC0001그룹이다. 1_15606054에서 1은 첫 번째 유전체이고 15606054는 유전자 신원이 된다. 1_2:116에서 116은 첫 번째 유전체와 두 번째 유전체 사이의 서열 유사도의 점수가 116임을 나타낸다. 그림 5는 그룹 OPC0001과 COG 및 KO와의 유사도를 임계값을 변화시키면서 조사한 결과를 나타내고 있다. 임계값이 증가함에 따라 그룹 내 원소의 수는 감소하지만 유사도는 증가하는 경향을 보이고 있다.

KO를 우리가 생산한 올소로그 클러스터의 일관성을 측정하는 하나의 기준으로 고려할 때, 임계값이 50bits 인 경우, OPC0001이 K02074와 가장 많은 수인 6개의 단백질을 공유하고 있다. 이때 참-긍정(True Positive)의 개수는 6이라고 하면 거짓-긍정(False Positive)의

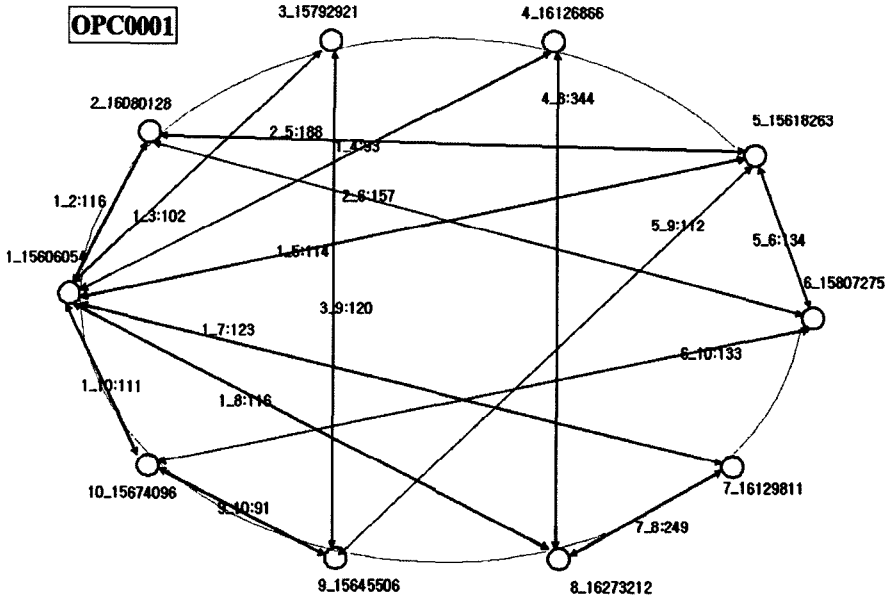


그림 4 우리가 제안한 방법에 따라 생성된 올소로그 클러스터의 실 예

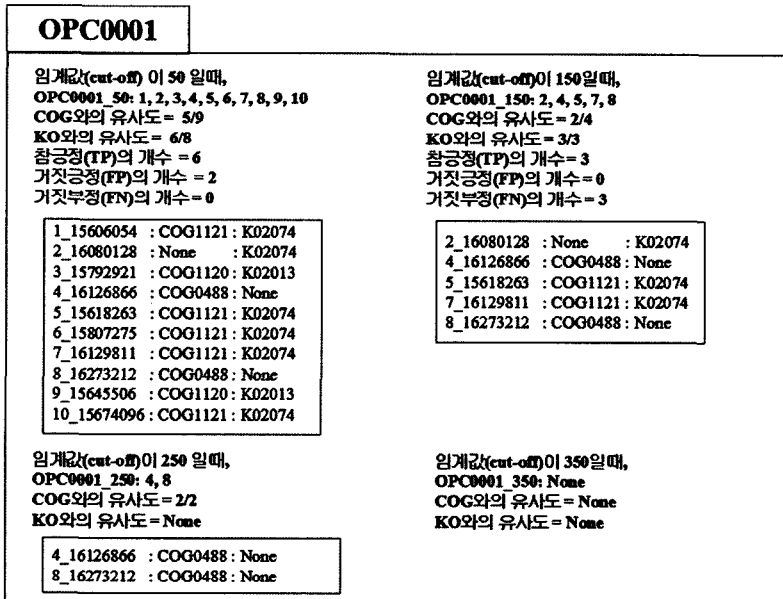


그림 5 그림 4의 올소로그 클러스터가 임계값의 변화에 따라 달라지는 유사도 및 올소로그의 개수

개수는 2이다. 그러나 거짓-부정(False Negative)와 참-부정(True Negative)의 개수는 이 정보만으로는 알 수 없다. 따라서 거짓-부정의 개수는 초기값을 0으로 두고 임계값이 증가함에 따라 변화과정을 측정하도록 한다. 임계값이 150bits인 경우, 3개의 참-긍정이 보여지고 거짓-긍정은 나타나지 않고 있다. 하지만 참-긍정의 개수는 6개에서 3개로 줄었으며 거짓-부정의 개수는

3개로 늘어났다.

4. 결과 분석

우리의 방법으로 산출한 올소로그 클러스터와 COG를 그림 6을 참조하여 비교하면, 임계값이 50bits인 경우 유사도는 약 85%이고 대체로 임계값의 증가에 따라 선형으로 증가하는 경향을 보여서 93%에까지 이르고 있

다. 이런 현상은 기본적으로 올소로그 클러스터를 형성하는 과정에서 BLAST 알고리즘을 사용한다는 점에서 서로 일치하기 때문에 기인한다고 생각할 수 있다.

한편, KO와의 비교결과를 살펴보면 유사도가 임계값이 50bits일 때 84%에서 임계값이 250bits에서 87%까지는 증가하는 경향이 보이고 있지만, 이후로는 거의 증가하지 않음을 알 수 있다. 이것은 아마도 서열 상동성에 기반하여 올소로그를 클러스터링하는 데 있어서 한계가 있음을 보여주는 것으로 생각된다.

그림 7에서는 참-긍정에 대한 거짓-긍정의 비율을 보이고 있다. 거짓-긍정의 비율이 2.7%에서 6.6%로 증가할 때 참-긍정의 비율은 17%에서 30%까지 증가한다. 참-긍정의 증가비율은 거짓-긍정의 증가비율의 3배에 해당한다.

그림 8은 거짓-긍정과 거짓-부정의 수의 비율의 변화를 보여준다. 임계값이 증가할 때 거짓-긍정의 수는 감소하지만 거짓-부정의 수는 상대적으로 증가한다. 임계

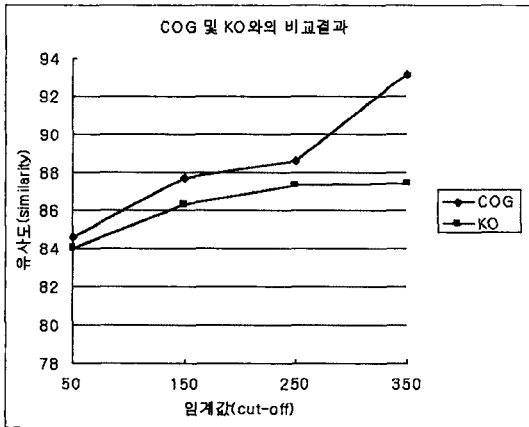


그림 6 우리의 올소로그 클러스터와 COG 및 KO와의 비교

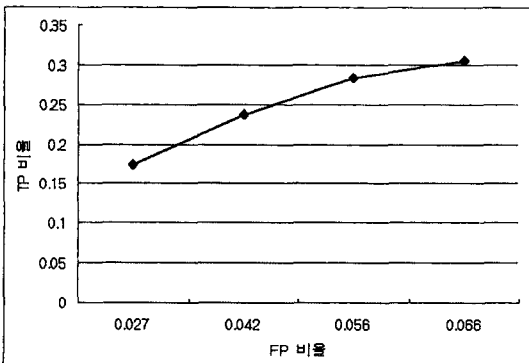


그림 7 KO를 기준으로 측정된 우리의 올소로그 클러스터의 임계값에 따른 거짓-긍정(FP)과 참-긍정(TP)의 비율 변화

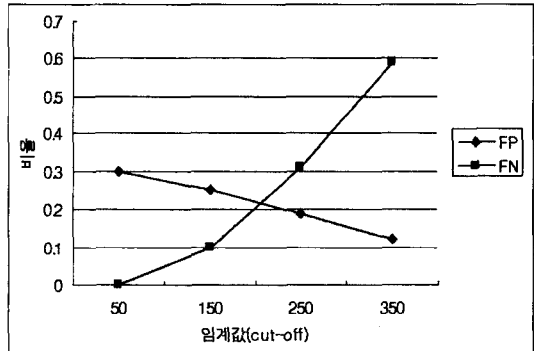


그림 8 KO를 기준으로 측정된 우리의 올소로그 클러스터의 임계값에 따른 거짓-긍정(FP)과 거짓-부정(FN)의 비율 변화

값이 50bits로부터 350bits로 증가할 때, 거짓-긍정의 비율은 약 30%에서 11%로 감소하지만 거짓-부정의 비율은 0%에서 60%로 증가한다. 즉, 거짓-부정의 증가비율은 거짓-긍정의 증가비율의 3배에 달한다.

5 결론

알려진 단백질 기능으로부터 서열 상동성에 기반하여 모르는 단백질 기능을 예측하기 위해, 유전체 서열이 완전히 알려진 다양한 종으로부터 올소로그를 탐지하는 것은 매우 유용하다. COG는 상호최대 BLAST 히트를 기반으로 분석을 통하여 올소로그 클러스터를 구축하였지만, 어떤 특정한 클러스터에는 한 종으로부터 유래한 패넛로그가 많이 포함되어있다. 이는 기본적으로 BLAST 알고리즘이 진화적 거리가 먼 종에서는 올소로그를 탐지하기가 상대적으로 매우 어렵기 때문이다.

이를 해결하기 위해 InParanoid와 OrthoMCL은 자동적으로 올소로그를 탐지할 뿐 아니라 임계값을 주어 올소로그 클러스터의 순도를 높였다. 이 논문에서, 우리는 이전연구에서 제안한 방법의 실험 결과인, 임계값을 변화시킬 때, 올소로그 클러스터의 순도가 어떻게 변화되는지를 조사하였다. 결과적으로 적절한 임계값이 주어지면, 기능에 따라 정확하게 분류된, KO 그룹과의 유사도가 약 90%까지 가능함을 알 수 있었다.

참고문헌

- [1] Tatusov, R. L., Koonin, E. V., Lipman, D. J.: A genomic perspective on protein families. *Science*, 278(5338) (1997) 631-637.
- [2] Tatusov, R. L., Galperin, M. Y., Natale, D. A., Koonin, E. V., et al.: The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28 (2000) 33-36.

- [3] Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Aviva R Jacobs, A. R. et al.: The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 4, (2003) 41.
- [4] Remm, M., Storm, C. E., Sonnhammer, E. L.: Automatic Clustering of Orthologs and in-paralogs from Pairwise Species Comparisons. *J. Mol. Biol.*, 314 (2001) 1041-1052.
- [5] Li, L. et al. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13(2003), 2178-89.
- [6] Kim, S., Jung, K. S., Ryu, K. H.: Automatic Orthologous-Protein-Clustering from Multiple Complete-Genomes by the Best Reciprocal BLAST Hits. In *Proc. of PAKDD 2006 Workshop, BioDM 2006*, 3916(2006) 60-70.
- [7] Fitch, W. M.: Distinguishing homologous from analogous proteins. *Syst. Zool.*, 19 (1970) 99-113.
- [8] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. et al.: Basic local alignment search tool. *J. Mol. Biol.*, 215 (1990) 403-410.
- [9] Chervitz, S. A., Aravind, L., Sherlock, G., Ball C. A. et al.: Comparison of the complete protein set of worm and yeast: orthology and divergence. *Science*, 282 (1998) 2022-2028.
- [10] Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L. et al.: Comparative genomics of the eukaryotes. *Science*, 287 (2000) 2204-2215.
- [11] Wheelan, S. J., Boguski, M. S., Duret, L., Makalowski, W.: Human and nematode orthologs - lessons from the analysis of 1800 human genes and the proteome of *Caenorhabditis elegans*. *Gene*, 238 (1999) 163-170.
- [12] Mushegian, A. R., Garey, J. R., Martin, J., Liu, L. X.: Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins enclosed by the human, fly, nematode, and yeast genomes. *Genome Res.*, 8 (1998) 590-598.
- [13] Kanehisa M., Peer B.: *Bioinformatics in the post-sequences era.* *nature genetics supplement*, 33 (2003) 305-310.
- [14] Bork P., Koonin E. V.: Predicting functions from protein sequence-where are the bottlenecks?. *Nat. Genet.*, 18 (1998) 313-318.
- [15] Eisen J. A.: Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.*, 8 (1998) 163-167.
- [16] Galperin M. Y., Koonin E. V.: Source of systematic error in functional annotation of genomes: domain rearrangement, nonorthologous gene displacement and operon disruption. *In Silico Biol.*, 1 (1998) 55-67.
- [17] Kimmen S.: Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics*, 20 (2004) 170-179.
- [18] Bono, H., Goto, S., Fujibuchi, W., Ogata, H. et al.:

Systematic Prediction of Orthologous Units of Genes in the Complete Genomes. *Genome Inform Ser Workshop Genome Inform.*, 9 (1998) 32-40.

[19] Dongen V. (2000), <http://micans.org/mcl/>



김 선 신

1989년 충북대학교 물리학과(이학사). 1995년 충북대학교 물리학과 물리학과전공(이학석사). 2002년 Syracuse University Computer Science(공학석사). 2007년 충북대학교 전자계산학과(이학박사). 2002년~2005년 충북대학교 자연과학대 및 전기전자컴퓨터공학부 시간강사. 2003년~2005년 북원대학교 컴퓨터공학부 시간강사. 2006년 North Carolina State University Computer Science 방문연구원. 2007년~현재 질병관리본부 국립보건연구원. 관심분야는 생물정보학, 데이터 마이닝, 데이터베이스, 병렬 처리



오 정 수

2003년 충북대학교 경영정보학과(학사) 2005년 충북대학교 경영정보학과(석사) 2005년~현재 국가생물자원정보관리센터 근무 중. 관심분야는 바이오인포매틱스, 클라우드컴퓨팅, SOA, 데이터베이스



이 범 주

2000년 서원대학교 전자계산학과졸업(공학사). 2003년 충북대학교 전자계산학과(이학석사). 2006년 충북대학교 전자계산학과(박사수료). 관심분야는 생물정보학, 엔자임 기능 예측, 데이터 마이닝, 시공간데이터베이스



김 태 경

2002년 충북대학교 경영정보학과(학사) 2005년 충북대학교 정보산업공학과(석사) 2005년~현재 충북대학교 정보산업공학과 박사과정 재학 중. 관심분야는 바이오인포매틱스, 그리드 컴퓨팅, XML, OLAP, SOA, SOA, BPM



정 광 수

2001년 충북대학교 화학공학부 졸업(공학사). 2004년 충북대학교 정보산업공학과 석사 졸업(공학석사). 2006년 충북대학교 대학원 전자계산학과(박사수료). 관심분야는 Bioinformatics, 단백질 서열 및 구조, 생명정보 데이터베이스



이 충 세

1979년 University of South Carolina 대학원 컴퓨터과학과(이학석사). 1990년 University of South Carolina 대학원 컴퓨터과학과(이학박사). 1979년~1988년 University of North Dakota Assistant Professor. 1989년~1991년 동아대학교 경영정보학과 부교수. 1991년~현재 충북대학교 전기전자컴퓨터공학 교수. 관심분야는 병렬컴퓨팅, 결합허용, 생명정보학, 정보보호



김 영 창

1971년 3월~1978년 2월 서울대학교 자연대학 미생물학과 미생물학 이학사. 1978년 3월~1980년 8월 서울대학교 대학원 미생물학 이학석사. 1981년 3월~1986년 8월 서울대학교 대학원 미생물 유전학 이학박사. 1989년 7월~1990년 7월 미국 Yale University 방문교수. 1991년 4월~1993년 2월 충북대학교 유전공학연구소 소장. 1994년 12월~1995년 2월 미국 Nebraska University 방문연구원. 1998년 3월~1999년 2월 미국 Kansas state University 방문연구원. 1985년 10월~현재 충북대학교 자연과학대학 생명과학부 교수. 2003년 3월~현재 충북대학교 바이오연구소 소장



조 완 섭

1985년 경북대학교 통계학과(학사). 1987년 한국과학기술원 전산학과(석사). 1996년 한국과학기술원 전산학과(박사). 2001년~2002년 미국 Univ. of Florida Post Doc. 연구원. 1987년~1990년 한국전자통신연구원 컴퓨터연구소 연구원. 1997년~현재 충북대학교 경영정보학과 교수. 관심분야는 데이터베이스, 데이터마이닝, XML, 바이오인포매틱스, BPM, SOA



류 근 호

1976년 숭실대학교 전산학과(이학사). 1980년 연세대학교 공학대학원 전산전공(공학석사). 1988년 연세대학교 대학원 전산전공(공학박사). 1976년~1986년 육군군수 지원사 전산실(ROTC 장교), 한국전자통신연구원(연구원), 한국방송통신대 전산학과(조교수) 근무. 1989년~1991년 Univ. of Arizona Research Staff (TempIS 연구원, Temporal DB). 1986년~현재 충북대학교 전기전자컴퓨터공학부 교수. 관심분야는 시간 데이터베이스, 시공간 데이터베이스, Temporal GIS, 지식기반 정보검색 시스템, 유비쿼터스컴퓨팅 및 스트림데이터처리, 데이터마이닝, 데이터베이스 보안, 바이오인포매틱스