

ARACNE를 이용한 미생물 Metabolic network의 기능적 연관성 분석

Nguyen Thuy Vu An · 홍순호*

울산대학교 생명화학공학과

Quantitative Relationship Analysis of Bacterial Metabolic Network using ARACNE

Thuy Vu An Nguyen and Soon Ho Hong*

School of Chemical Engineering & Bioengineering, University of Ulsan, 29 Mugeo-dong, Nam-gu, Ulsan 680-749, Korea.

Abstract Metabolic network is composed of more than thousands of metabolic reactions. Therefore, understanding of metabolic behavior of microorganisms is required to engineer metabolism of microorganisms. In this paper, we employed ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) to quantify relationships among metabolic subpathways. The results showed that ARACNE analysis can give new insight into the study of bacterial metabolism.

Keywords: metabolic network; ARACNE; relationship network

서 론

최근 DNA microarray를 필두로하여 다양한 high-throughput 기술들이 개발됨에 따라서 시스템생물학 (Systems Biology) 연구가 생물공학 연구의 새로운 패러다임으로 자리잡고 있다. 시스템생물학은 genome, transcriptome, proteome, metabolome 등 다양한 level에서 생명현상을 관찰하고, 이의 복합적인 분석을 통하여 생명현상을 글로벌수준에서 이해하는 학문이다. 그리고, 이러한 생명현상에의 이해를 바탕으로 하여 생명체의 특성을 원하는 방향으로 개량하는 일련의 연구를 시스템생물공학 (Systems Biotechnology) 이라 정의한다. 시스템생물공학연구의 효율을 높이기 위하여는 균주의 대사반응의 변화를 분석 및 예측하는 기술의 확보가 필수적이다. 따라서, 대사회로 분석 및 대사거동 예측은 시스템생물공학연구의 핵심적인 분야로서 활발한 연구가 이루어지고 있다(1-5).

대사network는 수천 개의 유전자 및 효소반응들이 서로 연관되어있는 복잡한 네트워크 시스템이다. 대사network를 유전공학적으로 방법으로 조작하여 원하는 기능을 나타내도록 개량하기 위하여 대사network간의 상관관계를 명확히 파악하는 것이 선결되어야 한다. 대사network사이의 상관관계를 파악하지 못하고 유전자 조작을 수행하였을 경우, 기대 하던 효과를 얻지 못할 수 있을 뿐 아니라 의도하지 않았던 부작용도 일어날 수 있다(6-7).

본 연구에서는 대사network간의 기능적 연관성을 분석하기 위하여 transcriptome 연구에 주로 활용되던 ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) 기법이 활용되었다. ARACNE는 새롭게 개발된 transcriptome 데이터 분석 프로그램으로서, microarray 데이터로부터 transcription 네트워크를 구축하기 위하여 활용되어지고 있다. 즉, 전체 유전자의 발현량 데이터를 분석하여 어떠한 유전자들의 발현이 서로 연관되어져 있는지 파악하여, 유전자들 사이의 기능적 연관성을 예측하게 하여주는 프로그램이다(8-11). 본 연구진은 64개 미생물 균주의 유전정보를 바탕으로 각 대사network간의 상관관계를 연구하였다.

*Corresponding author

Tel: +82-52-259-1293, Fax: +82-52-259-1689
e-mail: shhong@mail.ulsan.ac.kr

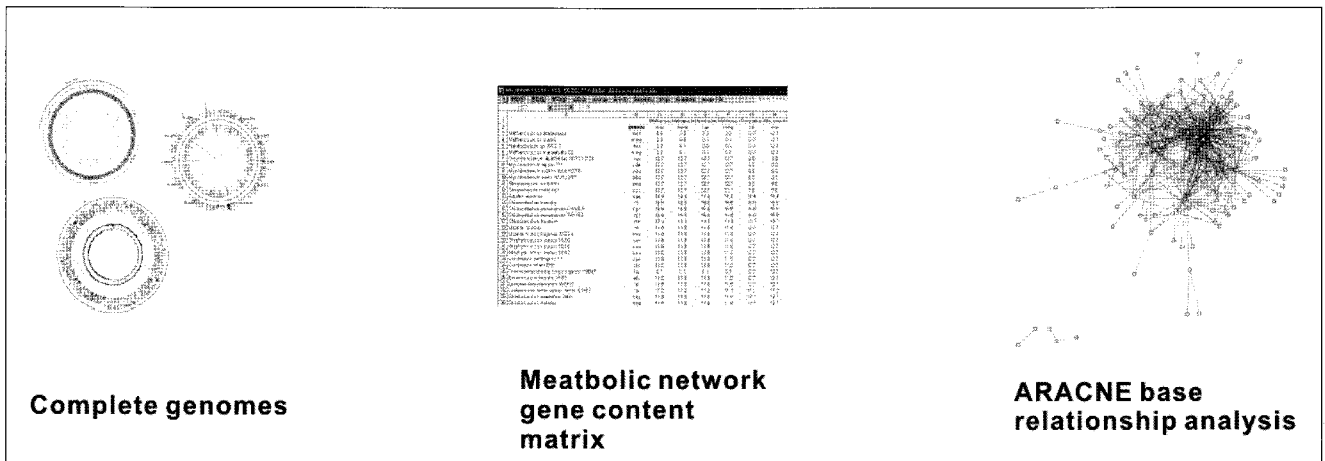


Fig. 1. Schematic diagram of the experimental procedure.

재료 및 방법

본 실험에서는 KEGG (Kyoto Encyclopedia of Genes and Genomes) 데이터베이스로부터 64개 박테리아의 유전정보를 확보하였으며, 이를 바탕으로 전체 대사network를 구축하였다(12). 전체 대사network를 기능별로 160개 소규모 대사subpathway로 분류하였으며, 박테리아별로 각 대사subpathway 관련 유전자의 포함여부를 조사하여 대사subpathway 매트릭스를 구축하였다(Fig. 1).

ARACNE는 동물세포의 복잡한 조절회로 (regulatory network)를 글로벌 수준에서 분석하기 위하여 개발된 프로그램이므로, 160개의 대사subpathway를 분석하기에는 충분한 수준의 분석능력을 제공 할 수 있다. ARACNE는 2단계의 과정을 통하여 데이터간의 상호 연관성을 분석한다. 첫 번째로 ARACNE는 Gaussian kernel density기법을 활용하여 각 데이터간의 상호연관성 정보 (mutual information, MI)를 파악한다. MI 계산을 위한 파라미터의 결정과 통계학적인 데이터분석을 위한 MI threshold의 결정이 매우 중요한 인자이다. 두 번째 단계는 통계학적으로 불충분한 중요성을 가지고 있는 연관성 데이터들을 제거한다. 따라서, ARACNE를 활용하여 구축한 연관성네트워크 (relationship network)는 높은 신뢰성을 지니고 있다.

대사subpathway 매트릭스를 input데이터로 사용하여 각 대사subpathway간의 기능적 연관성분석을 수행하였다. 최적의 네트워크를 구축하기 위하여 몇몇 파라미터의 값을 정하였다. 첫 번째 파라미터는 MI 계산을 위하여 필요한 Gaussian estimator의 kernel width이며, 이는 데이터의 양 및 통계적 특성에 연관되어져 있다. 두 번째 파라미터는 p-value이며, 이는 결과 연관성네트워크의 통계적 threshold와 연관되어져 있다. 본 연구에서는 kernel width와 p-value를 각각 0.3과 0.1으로 정하여 실험을 수행하였다. 구축된 연관성 네트워크는 Network Browser를 이용하여 시각화 하였다(9).

결과 및 고찰

본 연구진은 각각 다른 기능을 지니고 있는 대사subpathway들 사이의 진화 및 기능적 연관성을 분석하기 위하여 ARACNE 프로그램을 활용하였다. 다양한 조건에서 특정 유전자들의 발현 패턴이 유사하다면 그 유전자들 사이에 기능적 연관성이 있을 가능성이 높다 할 수 있다. 마찬가지로 특정 대사subpathway들이 미생물군주들 사이에 존재하는 패턴이 유사하다면 그 대사subpathway들이 서로 기능적 상관관계를 가지고 있다고 유추하여 볼 수 있다. 따라서 본 연구에서는 유전자들의 발현패턴을 분석하는 ARACNE라는 프로그램을 활용하여 미생물들의 subpathway들의 존재 패턴을 분석함으로써 각 subpathway사이의 기능적 상관관계를 분석하여 보았다.

ARACNE 분석결과 105개의 subpathway node사이 1,333개의 edge를 가지는 연관성 네트워크 지도가 구축되었다(Fig. 2). 즉, 163개의 전체 subpathway들 중에서 105개의 subpathway서로 상호적 연관성이 있었지만, 나머지 58개의 subpathway들은 다른 subpathway들과 의미 있는 연관성을 가지고 있지 않았다.

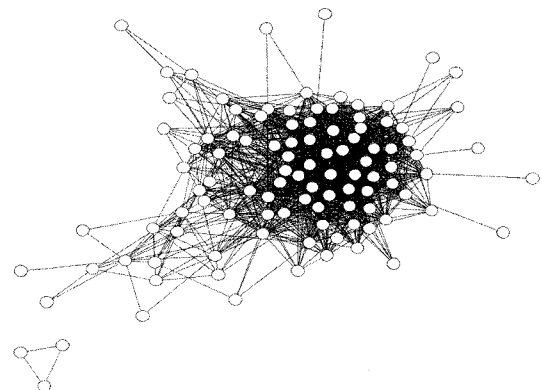


Fig. 2. The relationship network among metabolic subpathways.

연관성네트워크에 포함된 subpathway 혹은 node들의 기능별 분석을 수행하였다(Fig. 3). 중심 (central)대사회로에 포함되는 carbohydrate metabolism, nucleotide metabolism, amino acid metabolism 등에 연관된 subpathway들은 90% 이상의 subpathway들이 연관성네트워크에 포함됨으로서, 다른 subpathway들과 높은 기능적 연관성이 있음을 보여주었다. 그러나, secondary metabolite biosynthesis, signal transduction 등에 연관된 subpathway들은 상대적으로 매우 적은 수가 연관성네트워크에 포함됨으로서, 다른 subpathway들과 기능적 연관성이 낮다는 것을 보여주었다. 이러한 분석을 통하여 균주의 대사활동에서 중심적인 기능을 수행하는 소위 중심대사회로로 분류되는 대사회로들이 다른 대사subpathway들과의 연관성도 높다는 결과를 확인할 수 있었다.

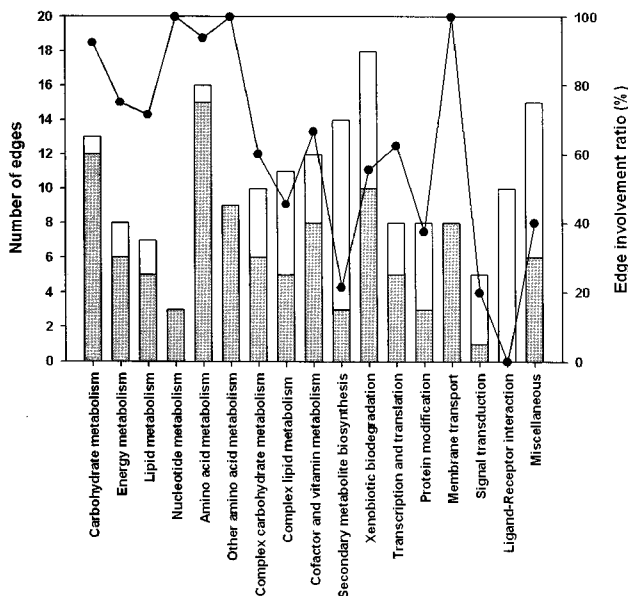


Fig. 3. Edge involvement graphs. White bars show the number of subpathway within the function groups, and grey bars show the number of relationship network involved subpathways. Dotted line shows the subpathway involvement ratio (number of relationship network involved subpathways/total number of subpathways).

연관성네트워크에 포함된 각 subpathway들은 평균적으로 25.4개의 다른 subpathway들과 연관성을 가지고 있었다. Subpathway들 중에 2개의 subpathway는 60개 이상의 subpathway들과 기능적 연관성이 있었으며, 10개의 subpathway는 50개 이상의 연관성 네트워크를 가지고 있었다(Table 1). 각 40, 30, 20개 이상의 연관성 네트워크를 가지고 있는 subpathway의 수는 17, 18, 9개 이었으며, 19개의 subpathway들은 10개 이상의 다른 subpathway들과 기능적 연관성이 있음이 예측되었다(Fig. 4). 105개의 subpathway중 30개의 subpathway는 10개 미만의 연관성 네트워크를 가지고 있었다.

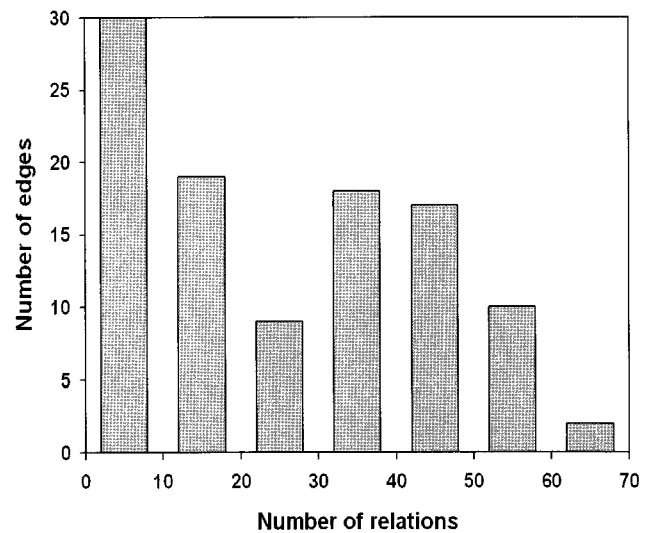


Fig. 4. Number of relationship profile. Grey bars show the number of edges which has specific number of relations (X-axis).

최다의 연관성을 가지고 있는 subpathway는 glycine metabolism으로서 67개의 다른 대사subpathway들과 기능적 연관성을 보여주었다. 또한 pyruvate metabolism (59개의 연관성), glyoxylate metabolism (58개), glutamate metabolism (58개), phenylalanine metabolism (56개) 등 소위 중심대사회로에 속하는 대사subpathway들이 높

Table 1. List of core edges in relationship network

Subpathway	No. of relationship	Subpathway	No. of relationship	Subpathway	No. of relationship
Glycine, serine and threonine metabolism	67	Lysine degradation	53	Other replication, recombination and repair factors	43
Nicotinate and nicotinamide metabolism	62	Sulfur metabolism	51	Benzoate degradation via hydroxylation	42
Pyruvate metabolism	59	Citrate cycle or TCA cycle	48	Peptidoglycan biosynthesis	41
Glyoxylate and dicarboxylate metabolism	58	Purine metabolism	48	Folate biosynthesis	41
Glutamate metabolism	58	Nitrogen metabolism	46	Oxidative phosphorylation	40
Phenylalanine metabolism	56	Valine, leucine and isoleucine biosynthesis	46	Phenylalanine, tyrosine and tryptophan biosynthesis	40
Lysine biosynthesis	55	DNA polymerase	46	Urea cycle and metabolism of amino groups	40
Benzoate degradation via CoA ligation	55	beta-Alanine metabolism	45	Glutathione metabolism	40
Butanoate metabolism	54	C5-Branched dibasic acid metabolism	44	Vitamin B6 metabolism	40
Tyrosine metabolism	53	Selenoamino acid metabolism	44		

은 연관성 수치를 보여주었다. 90% 이상의 중심대사회로 subpathway들이 연관성네트워크에 포함되었다는 사실과, 중심대사회로 subpathway들이 높은 연관성을 가지고 있다는 사실은 중심대사회로의 기능적 핵심성 및 중요성을 보여주고 있다. 본 실험결과는 transcriptome, proteome 발현패턴 분석을 통하여 중심대사회로에 포함된 유전자들의 중요성을 밝혀낸 기존 연구결과와 부합한다(13).

본 연구에서는 transcriptome연구에 활용되던 ARACNE 기법을 대사network의 기능적 연관성 분석에 활용하였으며, 소위 중심대사회로로 분류되던 대사subpathway들이 다른 대사subpathway들과 높은 기능적 연관성이 있음을 확인하였다. 본 연구에 활용된 새로운 대사network 분석 기법을 활용한다면 더욱 효율적인 대사network 분석연구가 수행될 수 있을 것이라 기대된다.

요 약

최근 미생물을 이용하여 기존에 화학산업을 통하여 생산되어지던 여러 화학물질 혹은 대사산물을 생산하려는 연구가 활발히 이루어지고 있다. 이러한 연구를 통하여 미생물의 대사특성을 개량하기 위하여는, 미생물의 대사특성을 분석하는 연구가 일차적으로 수행되어야 한다. 본 연구에서는 대사network간의 기능적 연관성을 분석하기 위하여 transcriptome 연구에 주로 활용되던 ARACNE 기법이 활용되었다. 특정 대사 subpathway들이 미생물 균주들 사이에 존재하는 패턴이 유사하다면 그 대사 subpathway들이 서로 기능적 상관관계를 가지고 있을 가능성이 높다는 가정 하에, ARACNE를 활용하여 미생물들의 subpathway들의 존재 패턴을 분석함으로써 각 subpathway사이의 기능적 상관관계를 분석하여 보았다. 본 연구에 활용된 새로운 대사network 분석기법을 활용한다면 더욱 효율적인 대사network 분석연구가 수행될 수 있을 것이라 기대된다.

감 사

본 연구는 2008년도 울산대학교 연구비의 지원으로 이루어졌습니다.

접수 : 2009년 3월 6일, 게재승인 : 2009년 6월 2일

REFERENCES

- Hong, S. H. (2007), Systems approaches to succinic acid-producing microorganisms. *Biotechnol. Bioprocess Eng.* **12**(2), 73-79.
- Ideker, T., T. Galitski, and L. Hood. (2001), A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343-372.
- Kitano, H. (2002), Computational systems biology. *Nature* **420**, 206-210.
- Kitano, H. (2002), Systems biology: a brief overview. *Science* **295**, 1662-1664.
- Lee, S. Y., D. Y. Lee, and T. Y. Kim (2005), Systems biotechnology for strain improvement. *Trends Biotechnol.* **23**, 349-358.
- Hong, S. H., T. Y. Kim, and S. Y. Lee. (2004), Phylogenetic analysis based on genome-scale metabolic pathway reaction content. *Appl. Microbiol. Biotechnol.* **65**, 203-210.
- Kim, J. S., H. Yun, H. U. Kim, H. S. Choi, T. Y. Kim, H. M. Woo, and S. Y. Lee. (2006), Resources for systems biology research. *J. Microbiol. Biotechnol.* **16**(6), 832-848.
- Butte, A. J. and I. S. Kohane (2000), Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.* **5**, 418-429.
- Duarte, N. C., S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Ø. Palsson (2007), Global reconstruction of the human metabolic network based on genomic and bibliomic data. *PNAS* **104**, 1777-1782.
- Franckea C, R. J. Siezena, and B. Teusink (2005), Reconstructing the metabolic network of a bacterium from its genome. *Trends in microbiology* **13**, 550-558.
- Hartemink, A. J., D. K. Gifford, T. S. Jaakkola, and R. A. Young (2001), Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. *Pac. Symp. Biocomput.* **2001**, 422-433.
- Margolin, A. A., K. Wang, W. K. Lim, M. Kustagi, I. Nemenman, and A. Califano (2006), Reverse engineering cellular networks. *Nature Protocol* **1**, 663-672.
- Kanehisa, M., S. Goto, S. Kawashima, and A. Nakaya (2002), The KEGG databases at GenomeNet. *Nucleic Acids Res.* **30**, 42-46.