

## 미생물의 보존적 유전자 탐색

이동근 · 이재화 · 이상현 · 하배진 · 심두희 · 박은정 · 김진욱 · 이화월 · 남천석 · 김남영 · 이어진 · 백진욱 · 하종명\*

마린바이오산업화 지원센터, 신라대학교 공과대학 생명공학과

Received March 14, 2005 / Accepted April 8, 2005

**Investigation of Conserved Genes in Microorganism.** Dong-Geun Lee, Jae-Hwa Lee, Sang-Hyeon Lee, Bae-jin Ha, Doo-Hee Shim, Eun-Kyung Park, Jin-Wook Kim, Hua-Yue Li, Chun-Suk Nam, Nam-Young Kim, Eo-Jin Lee, Jin-Wook Back and Jong-Myung Ha\*. *Marine-Biotechnology Center for Bio-functional material Industries, Department of Bioscience and Biotechnology, College of Engineering, Silla University* - To figure out conserved genes in 66 microbial species and measuring the degree of conservation, analyses based on COG (Clusters of Orthologous Groups of proteins) algorithm were applied. Sixty-six microbial genomes, including three eukaryotes, hold 63 conserved orthologs in common. The majority (82.5%) of the conserved genes was related to translation, meaning the importance of protein in living creatures. Ribosomal protein S12 (COG0048) and L14 (COG0093) were more conserved genes than others from the distance value analysis. Phylogenetically related microbes grouped in genome analysis by average and standard deviation of 63 conserved genes. The 63 conserved genes, found in this research, would be useful in basic research and applied ones such as antibiotic development.

**Key words** - conserved gene, microbial genome, ortholog, COG (Clusters of Orthologous Groups of proteins)

변화하는 지구 환경에 적응하기 위해 다세포 생물들이 다양한 진화적 산물들을 획득하고 소실하며 조상의 유전적 형태에서 떨어져 갔지만, 변화가 상대적으로 적은 미시적 환경 속에 서식하는 미생물들은 그들 조상과의 공통점을 오랜 시간 유지해 왔다[2].

진화를 통하여 공통조상 유전자(ancestral gene)는 종분화(speciation)와 복사(duplication)로 각 생물의 유전체에 분포하였으며, 각 생물체의 생명현상은 전체 생물체의 공통현상과 각 분류단위별로 특이적 현상을 함께 띄게 되었다. 생명을 이해하기 위해서는 각 분류단위에 독특한 생명현상과 함께 모든 생명체가 공통적으로 나타내는 필수기능(housekeeping function)에 대한 이해가 필요하다[12].

Orthologs는 공통조상에서 종분화 되어 서로 다른 종(species)에 있는 유전자들의 집합이며, 동일 ortholog내의 구성원들은 서열의 유사성과 기능의 동일성을 갖는다. Paralogs는 한 유전체내에서 특정 유전자의 복사로 생성된 유전자들의 집합으로, 복사된 유전자에 새로운 기능이 진화되어 구성원 상호간의 공통 기능은 거의 없다[12]. COG (Clusters of Orthologous Groups of protein)는 ortholog들에서 유래된 단백질의 집합으로 유사한 구조와 기능을 갖는 것으로 알려져 있다.

각 COG는 하나의 유전체에서 유래된 각 단백질들과 다른 유전체들에 속하는 단백질들 사이의 아미노산 서열 비교를 통하여 얻어지는데, 각 COG는 하나의 공통조상유전자(ancient conserved domain)에서 유래되어 3가지 이상의 생물에 분포하는 단백질들의 집합으로 간주할 수 있다[12].

COG 알고리즘은 비교유전체학의 방법 중 하나이다. COG는 아미노산 서열이 유사한 각 단백질 군(protein family)에 대한 진화적 관점에서의 분석을 용이하게 하며, COG 알고리즘을 이용하여 미지의 단백질 기능 추측, 한 유전체에서 검출하지 못한 보존적 그룹에 대한 체계적 연구, 광범위 항생제의 연구 등이 가능하게 된다[12].

미생물 유전체 43종을 COG 알고리즘으로 분석하여 비교 대상 모두에서 발견되는 보존적 유전자(conservative gene) 72개를 밝힌 보고가 있었는데[8], 본 연구에서는 66종의 미생물 유전체를 대상으로 미생물 유전자집합(gene pool)중 공통적으로 유지되고 있는 유전자들의 종류와 기능[10] 그리고 보존성의 정도를 파악하고자 하였다.

## 재료 및 방법

### 재료

분석에 이용된 미생물 유전체(microbial genome)는 NCBI의 공개 서버에서 추출하였다[3]. 유전체들은 2005년 1월을 기준으로 고세균(archaea) 21종, 진정세균(eubacteria) 197종 등 총 218종이었다. 미생물 유전자의 유사성에 관한 자료는 COGs에서 정리된 자료를 이용하였는데, 이들은 2005년 1월 현재 66종의 미생물 유전체에 포함된 총 144,324개의 유전자들을 4,873개의 COG 그룹으로 분류해 놓았다[6]. Table 1은 실제로 분석한 자료인 66종의 미생물들이다. 단세포 미생물에 대한 COG 자료를 정리한 Unicellular cluster 공개서버에서[6] 각 미생물이 함유하고 있는 COG 자료를 확보하였고, 66개의 미생물 전체가 공통적으로 보유하고 있는 COG 종류 정도 파악하였다.

### \*Corresponding author

Tel : +82-51-999-5467, Fax : +82-51-999-5636

E-mail : jmha@silla.ac.kr

Table 1. Studied 66 genomes derived from COGs database

Phylogenetic Group		Organism	
Archaea	Creanarchaeota	<i>Aeropyrum pernix</i> , <i>Pyrobaculum aerophilum</i> , <i>Sulfolobus tokodaii</i>	
	Euryarchaeota	<i>Archaeoglobus fulgidus</i> , <i>Methanococcus jannaschii</i> , <i>Methanopyrus kandleri</i> , <i>Pyrococcus abyssi</i> , <i>Pyrococcus horikoshii</i> , <i>Halobacterium</i> sp. NRC-1, <i>Methanobacterium thermoautotrophicum</i> , <i>Thermoplasma acidophilum</i> , <i>Thermoplasma volcanium</i> , <i>Methanosarcina acetivorans</i> C2A	
Bacteria	Aquificales-Thermotogales	<i>Aquifex aeolicus</i> , <i>Thermotoga maritima</i>	
	Cyanobacteria	<i>Synechocystis</i> sp., <i>Nostoc</i> sp.	
	Firmicutes	Gram (positive)	<i>Bacillus halodurans</i> , <i>Bacillus subtilis</i> , <i>Lactococcus lactis</i> , <i>Streptococcus pyogenes</i> , <i>Clostridium acetobutylicum</i> , <i>Listeria innocua</i> , <i>Staphylococcus aureus</i> , <i>Streptococcus pneumoniae</i>
		Mollicutes	<i>Ureaplasma urealyticum</i> , <i>Mycoplasma genitalium</i> , <i>Mycoplasma pneumoniae</i> , <i>Mycoplasma pulmonis</i>
	Chlamydia & Spirochaetales group		<i>Chlamydia pneumoniae</i> , <i>Chlamydia trachomatis</i> , <i>Borrelia burgdorferi</i> , <i>Treponema pallidum</i>
	Proteobacteria	Alpha	<i>Mesorhizobium loti</i> , <i>Caulobacter crescentus</i> , <i>Rickettsia prowazekii</i> , <i>Sinorhizobium meliloti</i> , <i>Agrobacterium tumefaciens</i> , <i>Brucella melitensis</i>
		Beta	<i>Neisseria meningitidis</i> MC58, <i>Ralstonia solanacearum</i> , <i>Neisseria meningitidis</i> Z2491
		Gamma	<i>Escherichia coli</i> K12, <i>Escherichia coli</i> O157:H7 EDL933, <i>Escherichia coli</i> O157:H7, <i>Salmonella typhimurium</i> , <i>Vibrio cholerae</i> , <i>Yersinia pestis</i> , <i>Pasteurella multocida</i> , <i>Pseudomonas aeruginosa</i> , <i>Xylella fastidiosa</i> , <i>Haemophilus influenzae</i> , <i>Buchnera</i> sp. APS
		Epsilon	<i>Campylobacter jejuni</i> , <i>Helicobacter pylori</i> 26695, <i>Helicobacter pylori</i> J99
	Actinobacteria		<i>Mycobacterium leprae</i> , <i>Mycobacterium tuberculosis</i> H37Rv, <i>Corynebacterium glutamicum</i> , <i>Mycobacterium tuberculosis</i> CDC1551
Deinococcus-Thermus group		<i>Deinococcus radiodurans</i>	
Fusobacteria		<i>Fusobacterium nucleatum</i>	
Eucarya	Fungi	<i>Saccharomyces cerevisiae</i> , <i>Schizosaccharomyces pombe</i> , <i>Encephalitozoon cuniculi</i>	

**아미노산 서열 분석**

자료의 무결성을 검증하기 위하여 COGs 데이터베이스로부터[6] 확보된 유전자 항목과 NCBI로부터[3] 획득한 자료를 비교한 결과, 둘 간의 차이가 심해 COGs 데이터베이스로부터 확보된 자료를 기준으로 하여 분석하였다. 66종의 미생물들이 공통적으로 함유하는 보존적 COG에 속하는 단백질들의 서열을 NCBI 공개 데이터베이스에서 추출하고, 각 보존적 COG에 속하는 단백질들은 CLUSTALW를 이용한 다중서열비교를 통해 distance value를 담고 있는 \*.dnd' 파일을 작성하였다[9,11]. PhyloDraw 프로그램(ver 0.8)을 이용하여 각 단백질의 distance value를 구하였고, distance value를 포함한 자료의 분석과 정리에는 엑셀 프로그램을 이용하였다.

**유전체의 보존성 분석**

분석대상 66종 미생물 모두에서 발견되는 63개의 보존적 유전자가 나타내는 distance value의 평균과 분산을 각 미생

물에 대하여 구하였다. 각 미생물이 나타내는 평균과 분산을 각각 X축과 Y축에 Creanarchaeota, alpha-Proteobacteria 등 각 미생물 그룹별로 묶어서 도시하였다.

**결과 및 고찰**

**보존적 유전자**

Table 2는 66종의 미생물 모두에서 발견되는 63개의 보존적 ortholog들을 기능별로 분류하고 강 등의[8] 결과와 비교한 것이다. Ortholog중 단백질 합성에 관여하는 유전자들이 총 52개(82.5%)로 아주 높은 비율을 차지하였다. 그리고 세포분열, 전사, 복제에 관여하는 COG 7개(11.1%)가 보존적이었다. 물질대사와 자기복제 등 생명체의 가장 큰 특징을 유지하기 위한 유전자들이 미생물들 간에 보존적인 것으로 판단할 수 있었다. 특히 물질대사를 담당하는 단백질 합성관련 유전자들의 보존비율이 아주 높은 것으로 나타나, 원시생명체부터

Table 2. Comparison between results of this study and Kang *et al.* (2002)

Functional category	Function	Kang <i>et al.</i>	This study
Cell cycle control, mitosis and meiosis	ATPase for cell cycle control	0037	0037
Translation	tRNA synthetase	0008, 0016, 0018, 0060, 0072, 0124, 0143, 0162, 0172, 0180, 0441, 0442, 0495, 0525,	0008, 0016, 0018, 0060, 0072, 0124, 0143, 0162, 0172, 0180, 0441, 0442, 0495, 0525, 0012, 0013,
	Ribosomal large subunit	0080, 0081, 0087, 0088, 0089, 0090, 0091, 0093, 0094, 0097, 0102, 0197, 0198, 0200, 0244, 0255, 0256,	0080, 0081, 0087, 0088, 0089, 0090, 0091, 0093, 0094, 0097, 0102, 0197, 0198, 0200, 0244, 0255, 0256
	Ribosomal small subunit	0048, 0049, 0052, 0092, 0096, 0098, 0099, 0100, 0103, 0184, 0185, 0199, 0522, 0051, 0186	0048, 0049, 0052, 0092, 0096, 0098, 0099, 0100, 0103, 0184, 0185, 0199, 0522
	Translation initiation factor	0231, 0480	0231, 0480
	Translation elongation factor	0361, 0532	0361, 0532
	rRNA methylation	0030	0030
	Polypeptide chain release factor		2890
	Methionine aminopeptidase	0024	
	GTPases - translation elongation factors	0050	
	Transcription	RNA polymerase	0085, 0086, 0202
Transcription antiterminator		0250	
Replication, recombination and repair	Exonuclease	0258	0258
	Topoisomerase	0550	0550
	DNA polymerase	0592	0592
Posttranslational modification, protein turnovers	Protease with possible chaperon activity	0533	0533
	Thioredoxin reductase	0492	
General function prediction only	EMAP domain	0073	0073
	Predicted GTPase	0012	to tRNA synthetase
Intracellular trafficking and secretion	Preprotein translocase	0201	0201
	Signal recognition GTPase	0541, 0552	0541, 0552
Others	Energy production and conversion	0636	
	Carbohydrate transport and metabolism	1109	
	Nucleotide transport and metabolism	0125	
	Thiol-disulfide isomerase and thioredoxins	0526	

생명체들은 물질대사를 주로 수행하는 것으로 추측할 수 있었다. 미생물 43종에서 72개의 ortholog들이 보존적이었는데 [8] 본 연구의 66종 미생물들은 63개가 보존적이었다. 비교대상 유전체가 늘어나면 공통 COG의 수가 줄어드는 것은 당연하다고 할 수 있을 것이다. DNA replication, repair 그리

고 recombination에 관여하는 COG들과 Intracellular trafficking 및 secretion에 관계되는 COG들은 두 결과가 동일하여, 새로이 추가된 20종의 미생물들도 모두 관련 유전자를 보존하고 있는 것을 알 수 있었다.

단백질 합성관련 유전자들의 개수는 두 연구 모두 52개로

동일하지만, 본 연구에서는 COG0024, COG0050, COG0051, COG0186 등이 빠지고 COG0012, COG0013, COG2890 등이 첨가되었다(Table 2). 첨가된 COG0012는 2002년 현재까지 기능을 알 수 없었지만, 2005년 현재 translation factor로 추정되어 translation group에 편입되었고, COG0013은(<http://www.ncbi.nlm.nih.gov/COG/old/palox.cgi?seq=COG0013>) 2002년 현재에도 보존적이었지만 강 등의[8] 결과에서 누락되어 있었다. 한편 COG2890은 *Aeropyrum pernix*의 APE0554 단백질이 2002년까지 기능을 알 수 없는 상태였는데, 기능이 추정되어 COG2890에 편입되어 분석대상 모든 균주가 보유한 것으로 나타났다. 전사(transcription) 관련 COG0202와 COG0250은 43종의 미생물에서는 보존적이었는데[8], 본 연구에서는 DNA-directed RNA polymerase alpha subunit/40kD subunit (COG0202)은 *Methanosarcina acetivorans*에 존재하지 않았고, Transcription antiterminator (COG0250)은 *Saccharomyces cerevisiae*와 *Encephalitozoon cuniculi*에 존재하지 않았다. 따라서 이들은 본 연구의 비교대상인 66종의 미생물 전체가 보유하지 않는 COG 이었다. *S. cerevisiae*에서 유래된 YML010w\_2단백질은 2002년 현재 COG 데이터베이스에서는(<http://www.ncbi.nlm.nih.gov/COG/old>) COG0202로 분류되었으나, 2005년 현재의 COG site에서는 COG0202에 속하지 않는 것으로 나타났다. YML010w\_2의 아미노산 서열을 blast로 탐색한 결과 일치하는 단백질이 없어, 자료의 업데이트 과정 등에서 변화된 것으로 추정되었다. 따라서 이들을 66종의 미생물 모두에 보존적인 COG로 표시할 수 없었다.

**유전자의 보존 정도**

Table 3은 66종 미생물 모두에 대하여 보존적인 것으로 나타난 총 63개의 유전자들 중 일부를 보존정도에 따라 정리한 것이다. 보존정도를 파악하기 위하여 distance value를 모두 합산하였다[8]. Distance value가 낮다는 것은 각 종들 간의 유전자 염기서열 차이가 작다는 것으로 보존성이 높다는 것을 의미한다. Ortholog and paralog 칼럼은 특정 종 내에서 유전자 복사에 의한 paralog가 존재할 때 paralog들까지 모두 합산한 결과이고, ortholog 칼럼은 특정 종 내의 여러 paralog들 중에서 가장 보존성이 높은 유전자 하나만을 기준으로 처리한 결과이다. 각각 Ribosomal protein S12 (COG0048)와 Ribosomal protein L14 (COG0093)의 보존성이 가장 높았으며, cell cycle control에 관여하는 ATPase인 COG0037의 보존성이 가장 낮았다. 전반적으로 리보솜 단백질 등 번역(translation)에 연관된 유전자들의 보존성이 높았다. 총 43종의 미생물 비교에서[8] 보존성이 높았던 COG0050과 COG0051은 66종의 미생물에 보존적이지 않아 보존정도를 파악할 때 빠졌다. 강 등은[8] 보존성을 파악할 때, 본 연구의 ortholog 칼럼 방법을 행하였는데 COG0050과 COG0051을 제외한 보존성이 높은 상위 10 개는 본 연구와 일치하였다. 이는 43종

Table 3. Ranking of conserved genes by conservation score (sum of distance value in each genome). The lesser value of conservation score represents that their orthologs were more conserved among microorganisms studied.

Rank	Ortholog and Paralog		Ortholog	
	Gene name	Conservation score <sup>a</sup>	Gene name	Conservation score <sup>b</sup>
1	COG0093	18.88740	COG0048	16.81466
2	COG0048	19.59167	COG0093	17.02286
3	COG0185	19.76947	COG0185	18.16033
4	COG0541 <sup>†</sup>	20.36047	COG0100	18.25909
5	COG0100	20.88018	COG0480*	18.82781
6	COG0099	21.18599	COG0099	19.11755
7	COG0090	21.32669	COG0090	19.19326
8	COG0092 <sup>†</sup>	21.63797	COG0080	19.21205
9	COG0552 <sup>†</sup>	21.77386	COG0094	19.56590
10	COG0094	21.77728	COG0049	20.19332
2				
13	COG0080	22.10116	COG0541 <sup>†</sup>	20.36048
14	COG0013 <sup>†</sup>	22.40891	COG0103	20.72743
2				
22	COG0081	23.15026	COG0086*	21.53684
23	COG0197	23.39066	COG0092 <sup>†</sup>	21.63798
24	COG0533	23.46116	COG0096	21.68347
25	COG0102	23.69625	COG0552 <sup>†</sup>	21.77385
2				
32	COG0200	24.64735	COG0013 <sup>†</sup>	22.40890
2				
37	COG0060	25.35302	COG0550*	22.71654
2				
47	COG0180	28.03673	COG0008*	24.07544
2				
55	COG0592	31.46243	COG0258*	25.76525
56	COG0480*	33.97377	COG2890	25.99539
57	COG0550*	34.30197	COG0073*	26.03856
58	COG0086*	35.92802	COG0244	26.04530
59	COG2890	36.29447	COG0072	26.34868
60	COG0258*	37.13544	COG0162	26.88977
61	COG0008*	44.37281	COG0592	27.54710
62	COG0073*	54.72456	COG0018	27.78686
63	COG0037*	56.86200	COG0037*	29.58232

<sup>a</sup>: Sum of distance of all proteins belong to a specific COG.  
<sup>b</sup>: Sum of distance of most conserved protein in each organism for a specific COG.  
<sup>†</sup>: Only one ortholog in each species, 66 ortholog in 66 species, for a specific COG.  
<sup>\*</sup>: The sum of orthologs and paralogs was more than 90 for a specific COG

이외에 본 연구에서 추가된 23종의 미생물들에서도 보존성이 높다는 것을 나타내는 결과일 것이다.  
 단일 유전체에 paralog 없이 ortholog 하나만 있는 경우

ortholog의 보존성은 낮고, paralog가 있는 경우 ortholog의 보존성이 높다는 보고도 있었지만[7], 본 연구에서는 일치하지 않는 경우도 있다는 것을 알 수 있었다. 즉 COG0480의 경우는 위의 보고와 일치하였지만, COG0541, COG0092, COG0552, COG0013의 경우를 보면 paralog들을 고려한 ortholog and paralog 칼럼에 비해 ortholog 칼럼에서 보존성 순위가 낮은 것으로 나타났지만, COG0037, COG0073, COG0258 등의 경우는 보면 paralog들이 많이 있어도 ortholog의 보존성 순서의 변화는 크지 않았다.

**유전체의 보존성**

Fig. 1은 66종 미생물의 개별 유전체를 63개의 보존적 유전자가 나타내는 distance value를 평균과 분산으로 나타낸 결과이다. 유전자 단위로 구해진 distance value들은 하나의 유전자 관점에서는 유전체의 특성을 대변할 수 없지만, 유전자 표본의 수가 많아질수록 유전체의 특성을 가깝게 표현한다고 말할 수 있을 것이다[8,9]. 서로 공유하지 못하는 유전자까지 분석하는 것은 논쟁이 될 수 있어, 본 연구에서는 분석대상 미생물 66종 모두에 보존적인 유전자 63개가 나타내는 단백질에 대한 distance value의 합을 이용하여 미생물 유전체를 분석하였다. 대부분의 종들이 기존의 미생물 중 분류상의 근연종과 그룹화 되어 있는 것을 알 수 있었다. 고세균 그룹 모두가 distance value가 높은 곳에 위치하였으며, 진핵생물 3종은 고세균과 distance value가 비슷한 것을 알 수 있었다. 고세균, 진정세균, 진핵미생물의 리보솜 단백질을 분석하여 Creanarcheota와 진핵생물이 유사함을 밝힌 보고가 있었는데[14], Fig. 1에서는 Creanarchaeota에 비해 Euryarchaeota가 진핵생물에 더 근접한 것으로 나타났다. 이는 63개의 보존적 유전자 중 리보솜 단백질 30개를 제외한 33개의 유전자들에 의한 결과로 생각되었다(Table 2). 강 등은[8] *Thermoplasma volcanium*의 평균값이 가장 높았는데 본 연구에서는 진핵생물인 *Encephalitozoon cuniculi*의 평균치가 가장 높았고(0.5291, 표준편차 0.0057), *T. volcanium*은 평균 0.4969(표준편차 0.0048) 이었다. 표준편차를 감안하면 둘 간의 차이가 크지 않아 어느 종이 더 보존적인지 확실하지 않다. 또한 표준편차와 평균을 고려할 때 Eucarya와 Euryarchaeota가 일부 겹치는 것으로 나타났는데, 이는 horizontal gene transfer 등[14] 진화의 부정확한 bifurcating과 일치하는 것으로 생각되었다.

근연종 내에서 각 유전체 평균의 범위는 Creanarchaeota, beta-Proteobacteria, Actinobacteria, Chlamydia-Spirochaetales, Cyanobacteria에서 적었다. 표준편차의 범위는 Mollicutes와 gamma-Proteobacteria에서 적었고 Chlamydia-Spirochaetales, Gram positive bacteria, Euryarchaeota에서는 넓어 동일한 유전체내에서 개별 유전자의 변이정도가 서로 다른 것을 알 수 있었다. 한 유전체에 포함된 각 유전자들의 편차가 큰 것

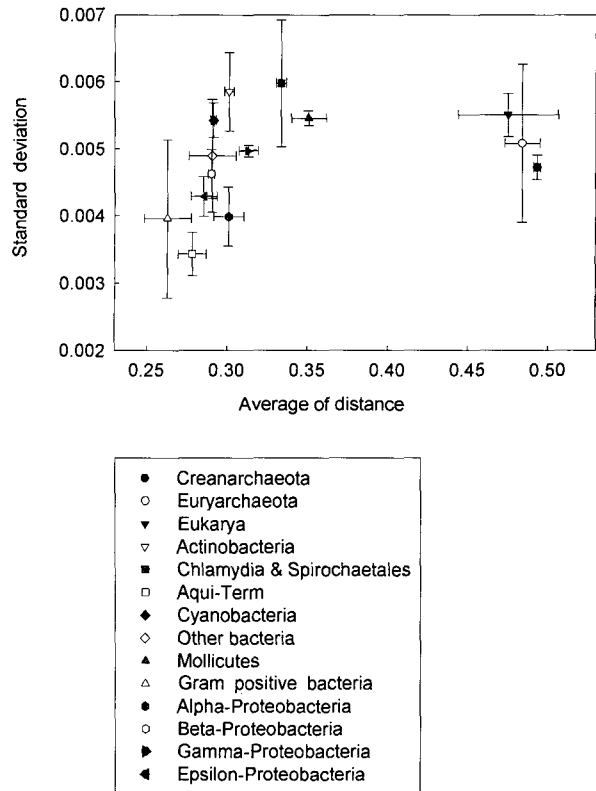


Fig. 1. Distribution pattern of microbial genomes by distance value. X-axis represents distribution of distance average (sum of distance value for each conserved COG was divided by 63) and Y-axis shows the variance of distance value among 63 COGs within each bacterial group.

은 수평적 유전자 전달에 의한 결과일 수 있다[14]. 그리고 gamma-Proteobacteria의 경우 9속 (genus) 11 유전체가 비교 되었는데도, 2속 3유전체인 epsilon-Proteobacteria그룹과 3속 3유전체인 Eukarya 그룹에 비해 평균과 표준편차의 범위가 작아, 속과 유전체 수에 따라 특정 미생물그룹 내 유전체들의 변이 정도가 결정되는 것은 아니라는 것을 알 수 있었다.

본 연구 결과에서 도출한 63개의 보존적 유전자는 생명체의 본질적 기능에 중요한 역할을 담당하는 것으로, 이들이 원시 생명체의 출발에서부터 보존적이었는지, 지구 환경의 변화에 따라 추가 등의 변화를 겪었는지 혹은 유전자의 기능 대체현상(gene displacement)에 의한 것인지[11] 알 수 없지만, 생명체의 진화 과정에서 이 유전자들이 보존된 이유와 기능적 연계에 대한 생물학적 연구에 기초 자료를 제공할 수 있을 것이다.

**요 약**

생명체의 본질적 기능에 중요한 역할을 담당하는 유전자들을 밝히기 위해 미생물 유전체들 사이의 공통적 유전자를 밝히는 COG알고리즘을 이용하였다. 진핵생물 3종을 포함한

66종의 미생물에서 63개의 유전자가 보존적이었으며, 단백질 합성에 관여하는 유전자들이 총 52개로 생명현상에서의 단백질의 중요성을 알 수 있었다. 각 보존적 유전자들의 distance value를 이용하여 종간의 유전자 변이의 정도를 보면, ribosomal protein S12 (COG0048)와 ribosomal protein L14 (COG0093)의 보존성이 가장 높았다. 보존적 유전자들의 평균과 분산으로 유전체 분석을 수행한 결과, 고세균과 진정세균의 각 그룹 등 근연종들이 독자적 그룹을 형성하였다. 하지만 각 그룹내의 속 및 유전체의 수와 유전체 변이의 정도는 비례하지 않는 것을 알 수 있었다.

### 참 고 문 헌

1. Baptiste, E., Y. Boucher, J. Leigh, and W. F. Doolittle. 2004. Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol.* **12**, 406-411
2. Fraser, C. M., J. A. Eisen, and S. L. Salzberg. 2000. Microbial genome sequencing. *Nature* **406**, 799-803
3. <ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria>
4. Harper, J. T. and P. J. Keeling. 2004. Lateral gene transfer and the complex distribution of insertions in eukaryotic enolase. *Gene* **340**, 227-235
5. Henikoff, S., E. A. Greene, B. S. Pietrokovski, T. K. Attwood, and L. Hood. 1997. Gene Families: The Taxonomy of Protein Paralogs and Chimeras. *Science* **278**, 609-614
6. <http://www.ncbi.nlm.nih.gov/COG/new/release/phylox.cgi>
7. Jordan, I. K., Y. I. Wolf, and E. V. Koonin. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evolutionary Biology* **4**, 22
8. Kang, H.-Y., C.-J. Shin, B.-C. Kang, J.-H. Park, D.-H. Shin, J.-H. Choi, H.-G. Cho, J.-H. Cha, D.-G. Lee, J.-H. Lee, H.-K. Park, and C.-M. Kim. 2002. Investigation of Conserved Gene in Microbial Genomes using *in silico* Analysis, *Korean Journal of Life Sci.* **5**, 610-621
9. Kimura, M. 1983. *The neutral theory of molecular evolution.* Cambridge University Press
10. Lee, D.-G., H.-Y. Kang, J.-H. Lee, and C.-M. Kim. 2003. Detection of Conserved Genes in *Proteobacteria* by using a COG Algorithm. *Korean J. Biotechnol. Bioeng.* **17**, 560-565
11. Mushegian, A. 1999. The minimal genome concept. *Curr. Opin. Genet.* **9**, 709-714
12. Tatusov, R. L., E. V. Koonin, and D. L. Lipman. 1997. A genomic perspective on protein families. *Science* **278**, 631-637
13. Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin. 2000. The COG database, a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33-36
14. Vishwanath, P., P. Favaretto, H. Hartman, S. C. Mohr, and T. F. Smith. 2004. Ribosomal protein-sequence block structure suggests complex prokaryotic evolution with implications for the origin of eukaryotes. *Mol. Phylogenet. Evol.* **33**, 615-625