

Investigation of Conservative Genes in 711 Prokaryotes

Dong-Geun Lee^{1,2} and Sang-Hyeon Lee^{1,2*}

¹Department of Pharmaceutical Engineering, College of Medical and Life Science, Silla University, Busan 617-736, Korea

²Department of Bioscience, Graduate School, Silla University, Busan 617-736, Korea

Received May 12, 2015 / Revised June 11, 2015 / Accepted June 16, 2015

A COG (Cluster of Orthologous Groups of proteins) algorithm was applied to detect conserved genes in 711 prokaryotes. Only COG0080 (ribosomal protein L11) was common among all the 711 prokaryotes analyzed and 58 COGs were common in more than 700 prokaryotes. Nine COGs among 58, including COG0197 (endonuclease III) and COG0088 (ribosomal protein L4), were conserved in a form of one gene per one organism. COG0008 represented 1356 genes in 709 of the prokaryotes and this was the highest number of genes among 58 COGs. Twenty-two COGs were conserved in more than 708 prokaryotes. Of these, two were transcription related, four were tRNA synthetases, eight were large ribosomal subunits, seven were small ribosomal subunits, and one was translation elongation factor. Among 58 conserved COGs in more than 700 prokaryotes, 50 (86.2%) were translation related, and four (6.9%) were transcription related, pointing to the importance of protein-synthesis in prokaryotes. Among these 58 COGs, the most conserved COG was COG0060 (isoleucyl tRNA synthetase), and the least conserved was COG0143 (methionyl tRNA synthetase). Archaea and eubacteria were discriminated in the genomic analysis by the average distance and variation in distance of common COGs. The identification of these conserved genes could be useful in basic and applied research, such as antibiotic development and cancer therapeutics.

Key words : Conservative gene, COG (cluster of orthologous groups of protein), ortholog, prokaryotic genome

서 론

염기서열분석법의 발달과 대규모 연구로 미생물 유전체(genome)의 염기서열에 대한 보고는 증가하고 있지만 유전자의 산물인 각 단백질의 기능을 실험적으로 모두 밝히지는 못하고 있으며[4] 유전자와 기능 파악을 위해 비교유전체학 등 생물정보학적 기법이 적용되고 있다. 각 생물종(species)이 나타내는 생명현상은 전체 생물체에서 공통적인 것과 각 생물종에 특이적인 것들의 조합으로 나타나는데, 유전자 측면에서는 공통유전자와 각 생물종 특이적 유전자에 의한 것이다. 따라서 현재 지구상에 존재하는 생명체의 이해를 위해서는 생물종 등의 각 분류단위별 특이적 유전자와 모든 생명체의 공통유전자에 대한 이해가 필요하다[13]. 이러한 유전자를 파악하려면 분석대상 유전자의 수가 많아 생물정보학이 필요하다.

종분화(speciation)와 복사(duplication)를 통하여 생물종들의 공통조상에 존재하던 공통조상 유전자(ancestral gene)가 각 생물종의 유전체에 분포한다고 알려져 있다. 공통조상에서

종분화를 통하여 다른 생물종들에 분포되어 있는 유전자들의 집합을 orthologs라고 하며, 동일 ortholog내의 구성원들은 서열이 유사하며 기능은 동일하다[13]. COG (Cluster of Orthologous Groups of protein)는 동일 ortholog들에서 발견되는 단백질의 집합으로 서로 유사한 구조와 기능을 갖는 것으로 알려져 있다. 각 COG는 하나의 공통조상유전자에서 유래되어 3가지 이상의 생물에 분포하는 단백질들의 집합으로 간주할 수 있다[4, 13]. Galperin 등에 의하면 COG 기법을 이용하면 첫째 파악된 게놈서열에서 ortholog와 paralog를 실험없이 탐색할 수 있고 둘째 생물체가 나타낼 수 있는 잠재적 기능을 파악할 수 있으며 셋째 생물체의 생물학적 기능을 상세히 파악할 수 있다는 장점이 있다[4]. Ortholog에 기반한 연구는 COGs (Clusters of Orthologous Groups of protein) 이외에도 진핵생물에서 구한 KOGs (euKaryotic orthologous groups of protein) [14], 고세균의 archaeal COGs [16], phage의 ortholog인 POGs (phage orthologous groups) [9] 등이 보고되고 있다.

COGs 자료는 1997년 7종의 미생물 유전체에서 720개의 COGs로 분류한 것을 시작으로[13] 2002년 43종의 미생물 유전체에 존재하는 총 77,069개의 유전자들을 3,852개의 COG로 분류하였으며[7] 2005년에는 66종의 미생물 유전체에 포함된 총 144,324개의 유전자들을 4,873개의 COG로 분류하였다. 2015년 현재 근 10년 만에 업데이트되면서 진핵미생물 관련 COG 등이 빠지면서 4,631개의 COG로 줄었지만 생물수는 대폭 늘어 총 711종의 원핵생물 유래의 유전체에서 분류하고

*Corresponding author

Tel : +82-51-999-5624, Fax : +82-51-999-5636

E-mail : slee@silla.ac.kr

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

있다[4].

보존적 유전자(conservative gene)를 강 등[7]과 이 등[10]이 보고하였지만 유전체 개수가 각각 43개와 66개로 현재의 711개와 큰 차이가 있으며, 효모 등의 진핵미생물이 본 연구에서는 빠졌으며, COG 개수도 현재와 차이를 보여 보존적 유전자에 대한 재정립이 필요하다 할 수 있다. 따라서 본 논문에서는 2015년 현재 711종의 원핵생물 유전체에서 유지되고 있는 보존적 유전자들의 종류와 기능[11] 그리고 보존성의 정도를 파악하고자 하였다.

재료 및 방법

재료

원핵생물 유전체의 유전자 유사성에 관한 자료는 COGs에서 정리된 자료를 이용하였다[3]. 각 원핵생물이 함유하고 있는 COG 자료를 확보하였고, 711개의 원핵생물 전체가 공통적

으로 보유하고 있는 COG 종류 등도 파악하였다. 이들은 2015년 1월 현재 711종의 원핵생물 유전체에 포함된 총 1,962,317개의 유전자들을 4,631개의 COG 그룹으로 분류해 놓았다[4]. Table 1은 실제로 분석한 자료인 711종의 원핵생물들의 분류학적 위치와 구성하는 생물종의 개수를 나타내고 있다.

아미노산 서열 분석

711종의 원핵생물 사이에 공통적으로 함유하는 보존적 COG에 속하는 단백질들의 서열을 NCBI 공개 데이터베이스에서 추출하고, 각 보존적 COG에 속하는 단백질들은 ClustalX (ver. 2.1) 프로그램을 이용한 다중서열비교를 수행한 후 bootstrap NJ method (n=200)를 통해 distance value를 담고 있는 '*.phb' 파일을 작성하였다[8]. Phylodraw 프로그램(ver 0.8)을 이용하여 각 단백질의 distance value를 구하였고, distance value를 포함한 자료의 분석과 정리에는 엑셀 프로그램을 이용하였다.

Table 1. Numbers of studied organisms and their phylogenetic groups derived from COGs database

Phylogenetic Group		Number of organisms	Abbreviations	
Archaea	Creanarchaeota	21	aC	
	Euryarchaeota	56	aE	
	Thaumarchaeota	4	aT	
	other Archaea	2	aZ	
Bacteria	Acidobacteria	6	A	
	Actinobacteria	74	B	
	Bacteroidetes	55	C	
	Chlorobi	5	D	
	Chlamydiae	6	E	
	Chloroflexi	9	F	
	Cyanobacteria	31	G	
	Deinococcus-Thermus	6	H	
	Firmicutes	Bacilli	33	I
		Clostridia	49	K
		other Firmicutes	6	L
		Mollicutes	10	M
	Fusobacteria	5	N	
	Aquificae	8	Q	
	Planktomycetes	6	P	
	Proteobacteria	Alpha	75	R
		Beta	52	S
		Delta	28	T
		Epsilon	11	U
		Gamma	103	V
Spirochaetes	7	W		
Synergistetes	5	X		
Thermotogae	7	Y		
other Bacteria	31	Z		

유전체의 보존성 분석

분석대상 711종 원핵생물 사이에서 공통적으로 발견되는 4,631개의 COG 중에서 700개 이상의 원핵생물이 공통적으로 보유하는 58개의 COG (Table 2)가 나타내는 distance value의 평균과 편차를 각 원핵생물에 대하여 구하였다. 각 원핵생물이 나타내는 평균과 분산을 Creanarchaeota, Acidobacteria 등의 문(phylum) 혹은 Bacilli, alpha-Proteobacteria 등 강(class) 수준으로(Table 1) 처리하여 표시하였다.

결과 및 고찰

보존적 유전자의 분포

Table 2는 711종 원핵생물 중 700종 이상의 원핵생물에서 보존적인 것으로 나타난 총 58개의 COG들을 보유 생물수, COG 번호, 구성하는 유전자 개수로 정리한 것이다. 전체 분석대상 711종의 원핵생물 모두에 보존적인 COG는 COG0080 (Ribosomal protein L11) 뿐이었고, 이는 리보솜의 구성성분일 뿐만 아니라 원핵생물의 스트레스 환경에서 발견되는 sigma factor의 활성화에 관여하는 등[17] 다양한 기능을 수행한다. 그리고 분석대상 원핵생물 모두에서 발견되어 이들의 생존에 필수적인 것을 알 수 있었다. 분석대상 모두에 보존적인 유전자의 수는 43종의 미생물에서 72개[7], 66종의 미생물에서 62개였는데[10] 본 연구에서는 711종의 원핵생물에서 1개로 나타났다. 이는 분석대상 생물수가 증가하면 보존적 유전자 수가 감소한다는 보고[10]와 일치하는 것이었다.

710종에 보존적인 COG는 COG0525, COG0197, COG0480, COG0085 였다. 711종 중에서 해당 COG를 보유하지 않은 원핵생물은 COG0525의 경우 beta-Proteobacteria 강(class)의 *Candidatus Nasuia deltocephalinicola* str. NAS-ALF, COG0197은 Acidobacteria 문(phylum)의 *Chloracidobacterium thermophilum* B, COG0480은 Synergistetes 문의 *Fretibacterium fastidiosum*, COG0085는 beta-Proteobacteria 강의 *Advenella kashmirensis* WT001 균주였다.

생물종의 수와 단백질의 수가 일치하는 즉 하나의 생물종에 하나의 유전자가 있는 COG는 COG0197, COG0088 등 9개였고 COG0008은 709개의 생물종에 1,356개의 유전자로 가장 많았다. 10,000개 이상의 유전자로 구성된 COG가 5개이고 3,000~9,000개 정도의 유전자를 갖는 COG가 55개 이상인 것 [4]과 비교하면 보존적 유전자들의 COG는 개수가 적은 편이었다(Table 2).

자연계에 존재하는 diterpenoid인 totarol을 처리하면 *Bacillus subtilis*의 ribosomal subunit들을 포함한 단백질의 발현 변화와 함께 항균작용이 나타나 COG 자료를 이용하여 새로운 항균제를 찾을 가능성이 있으며[12] 리보솜을 구성하는 단백질들은 리보솜 형태로 단백질 합성에 관여하는 외에 항암 작용이 있는 p53 경로의 활성화[2], 종양발달 및 면역신호와 분화에 관계하는 등 여러 생리작용과 병리학적인 과정에 참가하는 것으로 알려져 있다[18].

보존적 유전자의 기능

Table 3은 700종 이상의 원핵생물에서 발견되는 58개의 보존적 ortholog들을 기능별로 분류한 후 66종의 원핵생물에서 모두 발견된 62개의 보존적 ortholog 결과[10]와 비교한 것이다. Ortholog중 번역(translation)에 관여하는 유전자들이 총 50개(86.2%)로 나타났다. 분석대상 모두에 보존적인 COG는 COG0080 하나뿐이지만, 진핵미생물 3종 등 총 66종의 미생물 모두에서 공통적인 COG 62개 중 52개(82.5%)가 번역에 관련된 것과 본 연구의 700개 이상의 원핵생물에 보존적인 COG들을 비교하면 번역에 관련된 유전자의 비율이 높아진 것으로 나타났다. 이 등[10]은 물질대사와 자기복제 등 생명체의 가장 큰 특징을 유지하기 위한 유전자들이 미생물들 사이에 보존적인 것으로 보고하였고 특히 물질대사를 담당하는 단백질들의 합성관련 유전자들의 보존비율이 아주 높아 원시생명체부터 생명체들은 물질대사를 주로 수행하는 것으로 유추하였다. 본 연구와 이 등[10]의 결과를 비교하면 DNA 복제관련 유전자는 3개에서 COG0258 하나로 감소하였고 오히려 전사(transcrip-

Table 2. Number of organisms, COG and its containing organisms among 711 organisms. Prefix COG was omitted at each number

# of organisms	# of COG (# of proteins)										
711	0080 (716)										
710	0525 (738)	0197 (710)	0480 (1158)	0085 (1145)							
709	0086 (784)	0103 (712)	0087 (710)	0081 (709)	0088 (709)	0098 (709)	0200 (709)	0008 (1356)			
708	0013 (812)	0495 (764)	0522 (744)	0048 (712)	0049 (712)	0094 (710)	0052 (709)	0093 (709)	0096 (709)		
707	0124 (764)	0143 (760)	0016 (731)	0099 (717)	0184 (710)	0091 (709)	0090 (707)	0092 (707)	0097 (707)		
706	0441 (888)	0180 (861)	0361 (787)	0060 (741)	0102 (708)	0244 (707)	0459 (1076)				
705	0089 (706)										
704	0202 (745)	0186 (706)									
703	0024 (983)	0250 (919)	0533 (733)	0051 (723)	0100 (704)	0185 (704)	0256 (703)	0037 (1150)			
702	0030 (825)	0231 (785)	0172 (741)	0201 (716)	0072 (713)	0198 (703)					
701	0258 (835)										
700	0018 (763)	0532 (706)									

Table 3. Comparison of COGs between Lee *et al.* (2005) and this study. Prefix COG was omitted at each number

Functional category Function	Lee <i>et al.</i> (2005)	This study
Cell cycle control, mitosis and meiosis ATPase for cell cycle control	0037	
Translation tRNA synthetase	0008, 0012, 0013, 0016, 0018, 0060, 0072, 0124, 0143, 0162, 0172, 0180, 0441, 0442, 0495, 0525,	0008, 0013, 0016, 0018, 0037, 0060, 0072, 0124, 0143, 0172, 0180, 0441, 0495, 0525
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, 0256
Ribosomal small subunit	0048, 0049, 0052, 0092, 0096, 0098, 0099, 0100, 0103, 0184, 0185, 0199, 0522	0048, 0049, 0051, 0052, 0092, 0096, 0098, 0099, 0100, 0103, 0184, 0185, 0186, 0522
Translation initiation factor	0231, 0480	0231, 0361, 0532
Translation elongation factor	0361, 0532	0480
rRNA methylation	0030	0030
Polypeptide chain release factor	2890	
Methionine aminopeptidase		0024
GTPases - translation elongation factors tRNA A37 threonylcarbamoyltransferase		0533
Transcription RNA polymerase	0085, 0086	0085, 0086, 0202
Transcription antiterminator		0250
Replication, recombination and repair Exonuclease	0258	0258
Topoisomerase	0550	
DNA polymerase	0592	
Posttranslational modification, protein turnovers, chaperones Protease with possible chaperon activity	0533	
Protein folding related chaperonin		0459
General function prediction only EMAP domain	0073	
Intracellular trafficking and secretion Preprotein translocase	0201	0201
Signal recognition GTPase	0541, 0552	

tion) 관련하여 COG0202와 COG0250이 추가되어 4개로 증가하였다. 이것은 이 등[10]은 분석대상 미생물 모두에 보존적인 COG를 조사한 반면 본 연구에서는 711종의 분석대상 중 700개 이상의 원핵생물이 보유한 COG를 Table 3에 나타낸 결과로 사료된다. 실제로 이 등[10]이 사용한 COG 자료에서는 66종의 미생물 중에서 65개가 COG0202를 그리고 64개가 COG0250을 보유하고 본 연구에서는 711종의 분석대상 중 704개가 COG0202를 703개가 COG0250을 보유하고 있는 것으로 파악되었다(Table 2).

709종 이상의 원핵생물에서 발견되는 13개의 ortholog 중 COG0085와 COG0086은 RNA polymerase의 구성성분이며

COG0525와 COG0008은 tRNA synthetase이고 나머지 9개는 ribosomal large subunit들이었다. 708종에 보존적인 9종류의 ortholog들을 보면 COG0013과 COG0495는 tRNA synthetase 이고 COG0094와 COG0093은 ribosomal large subunit들이고 나머지 5개는 ribosomal small subunit들이었다. 즉 708종 이상의 원핵생물에 보존적인 22개의 ortholog 중에 전사에 관련하여 2개, tRNA synthetase 관련 4개, ribosomal large subunit가 8개, ribosomal small subunit가 7개로 단백질합성에 관련된 ortholog 들이 분석대상 대부분의 원핵생물에서 보존적인 것을 알 수 있었다.

66종의 미생물 모두에서 발견된 COG와 711종 중 700개 이

상의 원핵생물 중에서 파악한 본 연구결과를 비교하면 기존의 COG번호를 유지하면서 기능만 재분류된 것은 4개였다. COG0037이 cell cycle control 등의 기능에서 tRNA synthetase로, COG0361이 translation elongation factor에서 translation initiation factor로, COG0480이 translation initiation factor에서 translation elongation factor로, 그리고 COG0533이 chaperon 활성을 보이는 protease에서 번역(translation)관련하여 ANN 코돈의 인식을 돕는 tRNA A37 threonylcarbamoyltransferase TsaD로 재분류되었다. Ribosomal small subunit인 COG0051, COG0186 등은 강 등[7]의 보고에 존재하다가 이 등[10]의 보고에서 빠졌었고 본 연구에서 다시 추가되었다. Galperin 등[4]은 COG 자료를 업데이트하면서 이전 버전에서 분류에 오류가 있거나 기능을 알지 못하던 것은 실험이나 high-throughput 기법으로 재분류 하였는데 이것이 동일한 COG번호가 이 등[10]의 연구와 본 연구에서의 차이에 영향을 주었을 것이다. 그리고 이 등[10]이 translation elongation factor로 분류한 COG0533은 원래 translation initiation factor였는데 분류에 실수가 있었던 것으로 파악되었다.

각 COG의 구성원들의 변이로 나타나는 distance value의 합이 낮다는 것은 각 종들 간의 아미노산 서열 차이가 작다는 것으로 보존성이 높다는 것을 의미한다[7, 10]. 분석대상 58개의 COG 중 변이가 가장 적은 즉 보존성이 가장 높은 COG는 Isoleucyl-tRNA synthetase (COG0060) 였고 보존성이 가장 낮은 COG는 Methionyl-tRNA synthetase (COG0143)였다.

유전체의 보존성

Fig. 1은 711종 미생물의 개별 유전체를 분석에 사용된 58개의 보존적 유전자가 보이는 distance value를 평균과 편차 (variation)로 표시한 결과이다. 유전자 수가 많아지면 적은 유전자 수로 파악하지 못하는 유전체의 특성이 분석 가능하다[7, 8, 10]. 본 연구에서는 모든 원핵생물이 공유하는 COG들만을 분석하지 않아 오류의 가능성은 존재하지만, 분석대상 711종

중에서 700종 이상의 생물 즉 98.5% 이상의 분석대상에 보존적인 COG를 분석하므로 오류가 크지 않을 것이다.

유전체의 보존성을 나타내는 Fig. 1에서 보이는 특징은 첫째 고세균(Archaea)과 진정세균(Bacteria)이 변이의 평균에서 많은 차이를 보인다는 것이다. 즉 고세균은 0.48 이상의 변이를 보이지만 진정세균은 0.35 이하의 변이를 보였다. Creanarchaeota 등 각 문(phylum)을 구성하는 생물종들의 평균이 나타내는 표준편차를 감안하면 고세균 사이에서는 차이가 크지 않아 어느 문(phylum)의 변이가 적은 지 즉 보존적인지 확실하지 않은데 이 등[10]은 수평적 유전자 전달(horizontal gene transfer) 등 진화의 부정확한 bifurcating에 의한 것으로 토의하였다. 두 번째 특징은 편차의 범위가 고세균에 비해 진정세균이 크며 둘 사이에 편차의 평균은 큰 차이가 없었다. 수평적 유전자 전달에 의해 하나의 유전체에 존재하는 각 유전자들이 큰 편차를 보일 수 있다[1, 10]. 이런 관점에서는 Fig. 1에서 W로 표시된 Spirochaetes 문(phylum)이 편차가 가장 커 수평적 유전자 전달을 가장 많이 받은 것으로 사료되며, D로 표시된 Chlorobi 문(phylum)은 반대의 경우로 사료된다. Chlorobi 문(phylum)은 절대혐기성세균으로 광독립영양을 하는 반면 Spirochaeres 문(phylum)은 종속영양을 하며 대부분 혐기성세균으로 구성되어 있다[5]. 세 번째 특징은 Firmicutes와 Proteobacteria 문(phylum)에 속하는 여러 개의 강(class)이 존재하는데 분포에서 서로 인접하지 않고 차이를 보였다. 이러한 양상은 Proteobacteria에 속하는 alpha-Proteobacteria 등 5개의 강(class)이 나타내는 분포에 비해 Firmicutes가 더 크게 나타났다. 즉 Bacilli 강(Fig. 1의 I), Clostridia 강(K), Mollicutes 강(M), other Firmicutes (L)는 동일한 Firmicutes 문에 속하지만 Mollicutes 강은 평균과 편차에서 다른 강(class)들과 떨어져 있었다. 한편, Mollicutes 강은 COGs에서는 Firmicutes 문에 속하게 분류하였지만 NCBI의 taxonomy browser에서는 Tenericutes 문에 속하는 것으로 분류되어 있는데[6] 서로 다른 문이라면 평균과 편차의 차이는 특이적 사항이 아닐 수도

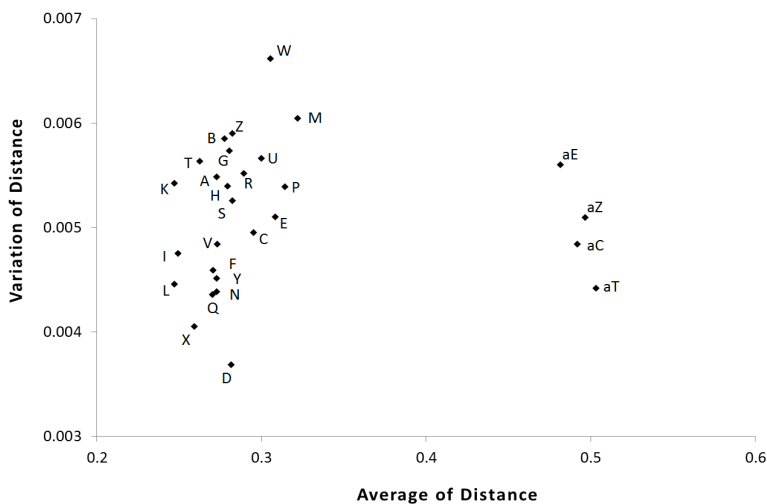


Fig. 1. Distribution pattern of prokaryotic genomes by distance value with conserved orthologs. X-axis represents distribution of distance averages and Y-axis shows the variance of distance averages for each phylum or class (Table 1).

있을 것이다. 마지막 특징은 각 분류 단위를 구성하는 유전체의 수와 유전체 변이의 정도는 비례하지 않는 것으로 나타났다. 즉 각각 7개와 10개의 유전체로 구성된 Spirochetes (Fig. 1의 W)와 epsilon-Proteobacteria (U)가 49개 유전체인 Clostridia (K), 33개 유전체인 Bacilli (I), 55개 유전체인 Bacteroidetes에 비해 변이의 평균과 편차가 모두 높게 나타났다.

문(phylum)이나 강(class) 수준이 아닌 개별 생물종(species) 수준에서는 고세균의 Thaumarchaeota에 속하는 *Candidatus Nitrosopumilus korensis* AR1이 최대의 변이 평균을 보였으며 Clostridia 강의 *Thermincola potens* JR가 최저의 변이 평균을 보였는데 개별 생물체에 대한 연구가 추가로 필요할 것으로 판단되었다.

Vishwanath 등[15]은 고세균, 진정세균, 진핵미생물에 존재하는 리보솜 단백질 분석으로 Crenarcheota와 진핵미생물이 유사하다고 보고하였고 이 등[10]은 66종의 미생물 유전체를 분석하여 Crenarchaeota에 비해 Euryarchaeota가 진핵미생물과 더 유사하다고 보고하였다. 이 보고들[10, 15] 역시 고세균과 진정세균은 상대적으로 유사하지 않은 것으로 나타났는데 이는 본 연구와 일치하는 것이었다.

본 연구에서 파악된 58개의 보존적 유전자는 원핵생물의 생명을 유지하는데 중요한 역할을 담당하는 것으로, 이들이 원시 생명체의 종분화(speciation) 이전부터 유지된 것인지 환경 변화에 따라 유전자 추가 등으로 적응한 것인지 혹은 ortholog가 아닌 유전자에 의해 기능이 대체되는 유전자의 기능 대체현상(gene displacement)에 의한 것인지 알 수 없지만 현재 지구의 환경에서는 필요한 유전자라고 할 수 있을 것이다. 추후 원핵생물 등 생명체의 진화 과정에서 이 유전자들이 보존된 이유와 기능적 연계에 대한 생물학적 연구에 기초 자료를 제공할 수 있을 것이며 항균제 개발[12] 및 항암치료제 개발과 면역조절 분야에서도 이용할 수 있을 것이다[2, 18].

References

- Baptiste, E., Boucher, Y., Leigh, J. and Doolittle, W. F. 2004. Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol.* **12**, 406-411.
- Bhat, K. P., Itahana, K., Jin, A. and Zhang, Y. 2004. Essential role of ribosomal protein L11 in mediating growth inhibition-induced p53 activation. *EMBO J.* **23**, 2402-2412.
- <ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data>
- Galperin, M. Y., Makarova, K. S., Wolf, Y. I. and Koonin, E. V. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, D261-D269.
- <http://en.wikipedia.org/wiki/>
- <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=31969>.
- Kang, H. Y., Shin, C. J., Kang, B. C., Park, J. H., Shin, D. H., Choi, J. H., Cho, H. G., Cha, J. H., Lee, D. G., Lee, J. H., Park, H. K. and Kim, C. M. 2002. Investigation of conserved gene in microbial genomes using *in silico* analysis. *J. Life Sci.* **5**, 610-621.
- Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press.
- Kristensen, D. M., Waller, A. S., Yamada, T., Bork, P., Mushegian, A. R. and Koonin, E. V. 2013. Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. *J. Bacteriol.* **195**, 941-950.
- Lee, D. G., Lee, J. H., Lee, S. H., Ha, B. J., Kim, C. M., Shim, D. H., Park, E. K., Kim, J. W., Li, H. Y., Nam, C. S., Kim, N. Y., Lee, E. J., Back, J. W. and Ha, J. M. 2005. Investigation of conserved genes in microorganism. *J. Life Sci.* **15**, 261-266.
- Lee, D. G., Kang, H. Y., Lee, J. H. and Kim, C. M. 2003. Detection of conserved genes in proteobacteria by using a COG algorithm. *Kor. J. Biotechnol. Bioeng.* **17**, 560-565.
- Reddy, P. J., Ray, S., Sathe, G. J., Gajbhiye, A., Prasad, T. S., Rapole, S., Panda, D. and Srivastava, S. 2015. A comprehensive proteomic analysis of totarol induced alterations in *Bacillus subtilis* by multipronged quantitative proteomics. *J. Proteomics.* **30**, 247-262.
- Tatusov, R. L., Koonin, E. V. and Lipman, D. L. 1997. A genomic perspective on protein families. *Science* **278**, 631-637.
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J. and Natale, D. A. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinf.* **4**, 41.
- Vishwanath, P., Favaretto, P., Hartman, H., Mohr, S. C. and Smith, T. F. 2004. Ribosomal protein-sequence block structure suggests complex prokaryotic evolution with implications for the origin of eukaryotes. *Mol. Phylogenet. Evol.* **33**, 615-625.
- Wolf, Y. I., Makarova, K. S., Yutin, N. and Koonin, E. V. 2012. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct.* **7**, 46.
- Zhang, S., Scott, J. M. and Haldenwang, W. G. 2001. Loss of ribosomal protein L11 blocks stress activation of the *Bacillus subtilis* transcription factor sigma(B). *J. Bacteriol.* **183**, 2316-2321.
- Zhou, X., Liao, W. J., Liao, J. M., Liao, P. and Lu, H. 2015. Ribosomal proteins: functions beyond the ribosome. *J. Mol. Cell Biol.* **7**, 92-104.

초록 : 원핵생물 711종의 보존적 유전자 탐색이동근^{1,2} · 이상현^{1,2*}(¹신라대학교 의생명과학대학 바이오산업학부 제약공학전공, ²신라대학교 일반대학원 바이오과학과)

원핵생물체의 생명유지에 중요한 역할을 담당하는 유전자들을 밝히기 위해 미생물 유전체들 사이의 공통적 유전자를 파악하는 COG 알고리즘을 이용하였다. 원핵생물 711종 모두에 보존적인 것은 COG0080 (Ribosomal protein L11) 1개였다. 708종 이상의 원핵생물에 보존적인 22개의 ortholog 중 전사관련 2개, tRNA synthetase 관련 4개, ribosomal large subunit 8개, ribosomal small subunit 7개였다. 700종 이상의 원핵생물에 보존적인 COG는 58개였다. 이중 리보솜을 구성하는 소단위체 등 번역 관련 COG가 50개(86.2%), 전사관련 COG가 4개(6.9%)로 나타나 생명현상에서의 단백질의 중요성을 알 수 있었다. 58개의 COG 중 보존성은 COG0060 (Isoleucyl tRNA synthetase)이 가장 높았고 COG0143 (Methionyl tRNA synthetase)이 가장 낮았다. 문(phylum)과 강(class) 수준에서 보존적 유전자들의 평균과 분산으로 유전체 분석을 수행한 결과 변이가 큰 고세균은 진정세균과 구분되었으며 편차는 일부 진정세균이 고세균보다 컸다. 보존적 유전자를 탐색하는 본 연구의 기법은 기초과학 연구와 함께 항균제 개발과 항암요법 개발 등에도 유용할 것이다.