

ORIGINAL ARTICLE

Orthologs로 파악한 원핵생물 계(kingdom)와 문(phylum)의 특징적 단백질

이동근 · 김안드레 · 이상현^{1)*}

신라대학교 의생명과학대학 제약화장품학과, ¹⁾신라대학교 생활과학대학 외식조리학부

Characteristic Proteins of Prokaryotic Kingdoms and Phyla Identified by Orthologs

Dong-Geun Lee, Andre Kim, Sang-Hyeon Lee^{1)*}

Department of Pharmaceutical and Cosmetics, College of Medical and Life Science, Silla University, Busan 46958, Korea

¹⁾Division of Culinary Arts, College of Life Science, Silla University, Busan 46958, Korea

Abstract

Based on the COGs (Clusters of Orthologous Groups of proteins) of 2,296 prokaryotes, we searched for COGs that were common or absent from archaea and bacteria at the kingdom and phylum level. There was a negative correlation between the number of members of a phylum and the number of COGs that all members had in common ($r=-0.508$) or did not have ($r=-0.636$). Of the 212 COGs found only in 193 archaea, more than 190 archaea possessed 18 COGs, including 10 ribosomal proteins (r -proteins). Of the 1,572 COGs found only in 2,103 bacteria, more than 2,060 bacteria possessed 19 COGs, including 10 r -proteins. The distribution of the averages and standard deviations obtained from the phylogenetic trees of r -proteins and other COGs in archaea and bacteria was not statistically significant ($p > 0.05$). There were 10 COGs that were found in all members of each phylum and not in other phyla including COG4920 (Uncharacterized membrane protein, DUF2208 domain) in the phylum Crenarchaeota, and 8 photosynthesis-related COGs and COG5784 (Uncharacterized cyanobacterial/plant protein Pro1025) in the phylum Cyanobacteriota. There were 170 COGs in 18 phyla and 1 class that were found in other taxa but not in their respective phyla or class, but there was no COG that was found in all members of other taxa but not in all members of its respective taxa. The results of this study will provide basic data necessary for the development of antibiotics or selective media specific to each prokaryote, as well as for understanding the structure of microbial communities.

Key words : Archaea, Bacteria, COG (Cluster of Orthologous Groups of proteins), Ortholog

1. 서 론

지구의 첫 생명체로 여겨지는 원핵생물은 변화하는 지구환경에 적응하여 살아왔고, 환경 변화에 따라 원핵생물 각 종은 새로 생기거나 사라졌을 것이며(Lee, 2025)

현재의 다양한 환경에서 발견되는 원핵생물은 초기 원핵생물의 유전자 변화가 서식 환경에 적합한 경우였을 것이다. 하나의 원핵생물이 가진 유전자 조합은 다른 원핵생물이 가지지 않은 유전자들과 다른 원핵생물과 공

Received 16 April, 2026; Revised 11 May, 2026;

Accepted 28 May, 2026

*Corresponding author : Sang-Hyeon Lee, Department of Pharmaceutical and Cosmetics, College of Medical and Life Science, Silla University, Busan 46958, Korea

Phone : +82-51-999-5624

E-mail : slee@silla.ac.kr

© The Korean Environmental Sciences Society. All rights reserved.

© 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.

통적인 유전자들의 조합이다(Segerman, 2012).

공통 조상 유전자 유래로 다른 종들에서 발견되는 유전자들의 집합을 ortholog라 정의하며, 동일한 ortholog에서 생성된 단백질들은 아미노산 서열이 유사하고 기능이 동일하다(Galperin et al., 2025). COG (cluster of orthologous groups of proteins)는 동일한 ortholog에서 발견되는 단백질들의 집합이며, 원핵생물은 여러 개의 COG를 갖고 있다(Galperin et al., 2025). COGs 자료는 1997년 7개 게놈에서 처음 보고되었고 여러 차례 업데이트 되었다(Galperin et al., 2025). 이 등(Lee et al., 2025)은 2,296개의 원핵생물에서 마련된 2024년의 COGs 자료로 원핵생물 각각이 가진 COG의 개수가 다르고 2,260개 이상의 원핵생물에 공통인 COG들의 보존성 등을 보고하였다.

한편 원핵생물 각 분류 그룹에 보존된 특징적 단백질 (conserved specific protein, CSP)을 이용한 분류 가능성이 보고되었다(Naushad et al., 2014). COG가 공통 유전자를 바탕으로 하므로 특정 COG가 CSP가 될 수 있을 것이다. 각 원핵생물이 가진 COG의 종류가 달라 특정 COG를 선택적으로 제어하는 항생제 등 약물 개발이 가능할 것이다(Lee et al., 2025). 고균과 세균의 계(kingdom)나 문(phylum) 등 분류단위에 공통적으로 존재하거나 공통적으로 없는 COG를 확인하면 의약품(Naushad et al., 2014) 혹은 선택배지의 개발과 함께 분류단위의 확인(Wang and Huang, 2018)에도 사용할 수 있을 것이다.

환경에 존재하는 미생물의 동정과 환경시료의 미생물 군집 구조 파악에 16S rRNA 유전자가 널리 사용되는데, rRNA 오페론의 개수가 최소 1개에서 최대 12개로 rRNA 유전자 수가 많은 미생물이 실제 미생물 개수에 비해 PCR (polymerase chain reaction) 산물이 많이 증폭되어 미생물 군집 구조 파악에 왜곡을 줄 수 있다(Fogel et al., 1999). Wang and Huang(2018)은 711개의 원핵생물에서 마련된 r-RNA와 2014년도 COGs 자료를 바탕으로 선정한 단백질을 이용하여 23개 문에 속하는 56개의 세균을 재료로 문(phylum) 수준 분류를 수행하였으나 세균수가 적고 고균이 없는 한계를 가졌다.

이에 이 연구에서는 2025년도에 소폭 업데이트된 COGs 자료를 바탕으로 2,296개의 원핵생물 각각이 가진 COG를 구하고 고균과 세균의 계나 문 단위별로 공통적으로 갖거나 갖지 않는 COG를 탐색하여 원핵생물

의 분류와 약물 및 선택배지 개발 그리고 환경시료의 미생물 군집 구조의 정확한 파악 등에 활용할 수 있는 기초 자료를 제공하고자 하였다.

2. 재료 및 방법

2.1. 재료

2025년 8월 기준으로 NCBI의 COGs 자료를 이용하여 2,296개의 원핵생물의 5,061개의 COGs를 분석하였다(Galperin et al., 2025). 각 원핵생물에 존재하는 COG들을 구한 후에, Excel 프로그램으로 원핵생물을 문이나 강(class)의 분류 단위까지 나눈 COGs 자료(Galperin et al., 2025)에서 정한 기준으로 정렬하였다. 이후 각 분류군 모두가 공통적으로 갖거나 갖지 않는 혹은 일부 구성원만 갖는 COG들을 확인하였다.

2.2. 계통수 작성과 분석

고균에서만 발견되는 212개의 COG 중에서 193개의 전체 고균 중 190개 이상의 고균에 존재하는 18개 COG와 세균에서만 발견되는 1,572개의 COG 중에서 전체 2,103개의 세균 중 2,060개 이상에 존재하는 19개 COG를 선별하였다. 이 연구에서는 COG는 단백질 그룹으로 COG의 구성원 각각은 단백질로 구분하였다. 선별된 각 그룹에 속하는 단백질들의 서열 번호를 정리한 후 NCBI에서 각 서열 번호에 해당하는 아미노산 서열을 다운받았다. Mega 12 프로그램(Kumar et al., 2024)으로 각 그룹에 속하는 단백질들의 아미노산 서열을 정렬하고, 계통수(phylogenetic tree)를 Maximum-likelihood 방법과 bootstrap (n=100)으로 구하였다. Dendroscope 프로그램(Huson and Scornavacca, 2012)으로 계통수에서 각 단백질 그룹에 속하는 각 단백질의 distance to root를 구하고 Excel 프로그램을 이용하여 결과를 정리하였다.

3. 결과 및 고찰

3.1. 각 분류군별 공통 COG 개수

원핵생물은 고균계(Archaea)와 세균(Bacteria)계로 구성된다. Table 1에 원핵생물을 구성하는 고균계와 세균계의 2 계와 개별 문(phylum)의 구성원 모두가 공통으로 갖는 COG의 개수와 공통으로 없는 COG의

Table 1. Taxonomic levels, number of members, and number of COGs having all or none at the taxonomic levels. Parentheses indicate the kingdoms, classes or other taxonomic levels

Kingdom Phylum Class	# of members	# of COGs all members	
		Having	Not having
Archaea	(193)	(3)	(1,572)
Asgardarchaeota	8	330	3,090
Crenarchaeota	29	329	3,149
Euryarchaeota	103	84	1,912
Nitrososphaerota	13	411	3,278
Thermoplasmatota	17	320	3,009
Other Archaea	(23)	(78)	(3,065)
Bacteria	(2,103)	(0)	(212)
Acidobacteriota	16	573	1,771
Actinomycetota	253	37	1,374
Aquificota	9	628	3,109
Bacillota	355	43	1,021
Bacilli	(152)	(170)	(1,462)
Clostridia	(159)	(96)	(1,297)
Erysipelotrichia	(14)	(432)	(2,781)
Negativicutes	(12)	(630)	(2,342)
Tissierellia	(16)	(392)	(2,468)
Other Bacillota	(2)	(1,100)	(2,970)
Bacteroidota	192	42	1,520
Bdellovibrionata	7	607	2,654
Campylobacterota	20	505	2,438
Chlamydiota	12	442	3,399
Chlorobiota	5	917	3,156
Chloroflexota	19	303	1,885
Cyanobacteriota	75	394	1,858
Deferribacterota	7	862	2,997
Deinococcota	7	821	2,739
Fusobacteriota	9	451	2,678
Mycoplasmota	24	159	3,346
Myxococcota	17	789	1,688
Planctomycetota	44	397	1,553
Pseudomonadota	808	0	698
α -Proteobacteria	(300)	(4)	(1,091)
β -Proteobacteria	(166)	(13)	(1,101)
γ -Proteobacteria	(338)	(1)	(852)
Other Proteobacteria	(4)	(680)	(2,596)
Spirochaetota	22	224	1,997
Synergistota	10	541	2,832
Thermodesulfobacteriota	56	352	1,449
Thermotogota	12	600	2,925
Verrucomicrobiota	24	81	1,978
Other Bacteria	100	15	1,148

개수를 표시하였다. 전체 COG의 개수가 5,061개이고, Table 1에서 Asgardarchaeota는 구성원 8개 모두가 330개의 COG를 공통으로 갖고 구성원 모두에 3,090개의 COG가 공통으로 없으므로, 1,641개의 COG는 8개의 구성원 일부가 갖는 COG이다. 고균과 세균 전체에서 문의 구성원 수와 구성원 모두가 공통으로 갖는 ($r=-0.508$) 혹은 공통으로 갖지 않는($r=-0.636$) COG의 개수 사이는 음의 상관관계를 보였다. Other Archaea와 other Bacteria를 하나의 문으로 간주하였을 때만 고균과 세균 전체가 30개 문이라서 통계분석의 신뢰도를 갖지만, 고균의 6개 문과 세균의 24개 문 그리고 Bacillota와 Pseudomonadota 문의 10개 강(class)으로 상관관계를 구했을 때 통계분석의 신뢰도가 낮아도 더 높은 상관관계를 보였다(자료 미제시). 공통으로 갖거나 갖지 않는 COG는 각 그룹에 보존된 특징적 단백질(CSP)이 존재한다는 것으로 진화적으로 공통조상에서 유래되었다는 것을 강력하게 나타내며 해당 그룹이 특정 환경에서 생존하는데 필수적인 대사 경로에 필요할 가능성이 높고 선택적 독성을 갖는 약물 개발의 대상이 될 수 있어(Naushad et al., 2014) 선택배지 개발에도 이용될 수 있을 것이다.

3.2. 세균계 및 고균계 특이 COG

원핵생물은 세균계와 고균계로 나누어지고, COGs 자료는 세균 2,103개와 고균 193개에서 구한 것으로 두 계의 구성원 수에서 차이를 보이지만 세균계에는 없는 즉 고균계에서만 발견되는 COG는 212개였다(Table 1).

Table 2에 고균계와 세균계에서만 발견되는 COG들을 보유 원핵생물 개수의 내림차순으로 정렬하면서 COG ID (기능)도 나타내었다. 193개의 고균 모두에 존재하는 COG는 3개였고, 2,103개의 세균 모두에 존재하는 COG는 없었다. 많은 수의 고균에 존재하는 상위 18개의 COG는 전사(COG1095, 1644, 1761), 번역(10개 리보솜 단백질과 COG1500, 1093, 2092, 5257)에 관련된 기능이 대부분이었고 tRNA의 이어맞추기(splicing)도 있었다(COG1676). 여기에는 이(Lee, 2025)가 보고한 리보솜 단백질(r-단백질)들 중에서 고균 193개 모두에 존재하는 r-단백질 S4e와 S8e COG는 고균에만 있어 포함되고, r-단백질 L12e와 L44e COG는 192개의 고균과 1개의 세균에 존재하여 제외되었다(Table 2). 많은 수의 세균에 존재

Table 2. High-ranking archaea and bacteria-specific COG IDs (functions) and number of prokaryotes among 193 archaea and 2,103 bacteria, respectively

Kingdom Archaea		Kingdom Bacteria	
# of prokaryotes	COG ID (Function)	# of prokaryotes	COG ID (Function)
193	1471 (Ribosomal protein S4e)	2,089	0268 (Ribosomal protein S20)
	1500 (Ribosome maturation protein)	2,088	0216 (Protein chain release factor RF1)
	2007 (Ribosomal protein S8e)	2,086	0222 (Ribosomal protein L7)
		2,083	0305 (Replicative DNA helicase)
192	1093 (Translation initiation factor 2 α)	2,082	0233 (Ribosome recycling factor)
	1095 (DNA-directed RNA polymerase E')	2,081	0211 (Ribosomal protein L27)
	1676 (tRNA splicing endonuclease)		0254 (Ribosomal protein L31)
	1727 (Ribosomal protein L18e)		0360 (Ribosomal protein S6)
	2125 (Ribosomal protein S6e)	2,080	0264 (Translation elongation factor EF-Ts)
	2238 (Ribosomal protein S19e)		0292 (Ribosomal protein L20)
	5257 (Translation initiation factor 2 γ)		0691 (tmRNA-binding protein)
191	1644 (DNA-directed RNA polymerase N)	2,077	0629 (Single-strand DNA binding protein)
	2092 (Translation elongation factor)	2,072	0706 (Membrane protein insertase)
190	1632 (Ribosomal protein L15e)	2,066	0291 (Ribosomal protein L35)
	1717 (Ribosomal protein L32e)		0335 (Ribosomal protein L19)
	1761 (DNA-directed RNA polymerase L)	2,063	0275 (16S rRNA C1402 N4-methylase)
	1890 (Ribosomal protein S3ae)	2,061	0193 (Peptidyl-tRNA hydrolase)
	2139 (Ribosomal protein L21e)		0261 (Ribosomal protein L21)
	2147 (Ribosomal protein L19e)	2,060	0359 (Ribosomal protein L9)

The prefix COG is omitted from the COG ID

하는 상위 19개의 COG는 복제(COG0305, 0629), 번역(10개 r-단백질과 COG0216, 0233, 0264, 0691, 0275, 0193)에 관련된 기능이 대부분이었고 세포막 단백질의 접힘과 배치 기능(COG0706)도 있었다. 복제, 전사, 번역에 관련된 세균 및 고균 특이적 COG가 대부분의 구성원에 분포된 것을 알 수 있었다. 세균과 고균 특이적 단백질 중 r-단백질 그룹이 많았다. 711개의 원핵생물에서 마련된 2014년도 COGs 자료의 r-단백질 S10 (COG0051)을 이용한 계통도와 16S 및 23S rRNA 유전자의 계통도가 세균의 문 수준에서 거의 동일하고 r-단백질 L2 (COG0090), S3 (COG0092), S7 (COG0049), L14 (COG0093), S12 (COG0048)도 상관성이 높다고 보고되었지만 (Wang and Huang, 2018) 이들 r-단백질 그룹은 모두 고균과 세균에 함께 분포하여 본 연구에서는 제외되었다. Wang and Huang(2018)은 r-단백질 유전자의 수가 원핵생물들 사이에 큰 차이가 없어 원핵생물에 따라 유전자 개수 차이가 많은 rRNA를 이용한 동정을 보완할 수 있다고 하였다. 이는 환경시료

의 미생물 군집구조 파악에 도움이 될 것이고, 이 연구결과도 군집구조 파악의 기초자료로 활용될 수 있을 것이다. 고균의 상위 18개와 세균의 상위 19개 COG를 비교하면 각각 r-단백질 그룹이 10개씩이었다. 계통수에서 구한 root distance의 평균과 표준편차를 보면(Fig. 1A) 고균 특이 r-단백질 그룹에 비해 세균 특이 r-단백질 그룹들이 root distance의 평균과 표준편차의 범위가 넓었지만, t-test에서 통계적 유의성은 높지 않았다($p > 0.05$).

구성원 수는 고균은 190개 이상이고 세균은 2,060개 이상으로 많은 차이를 보였지만 모두 30개 이상으로 통계적 분석은 충분하였다. 원핵생물 2,296개에 분포하는 r-단백질의 보존성의 보고(Lee, 2026)에서 88개의 r-단백질 그룹이 6개에서 2,285개의 원핵생물에 분포하였고, 2,270개 이상의 원핵생물에 총 18개의 r-단백질 그룹이 존재하였고 이들은 모두 고균과 세균에 분포하였다.

Fig. 1B는 r-단백질 그룹을 제외한 고균 및 세균 특이 COG들의(Table 2) 계통수에서 구한 root distance의

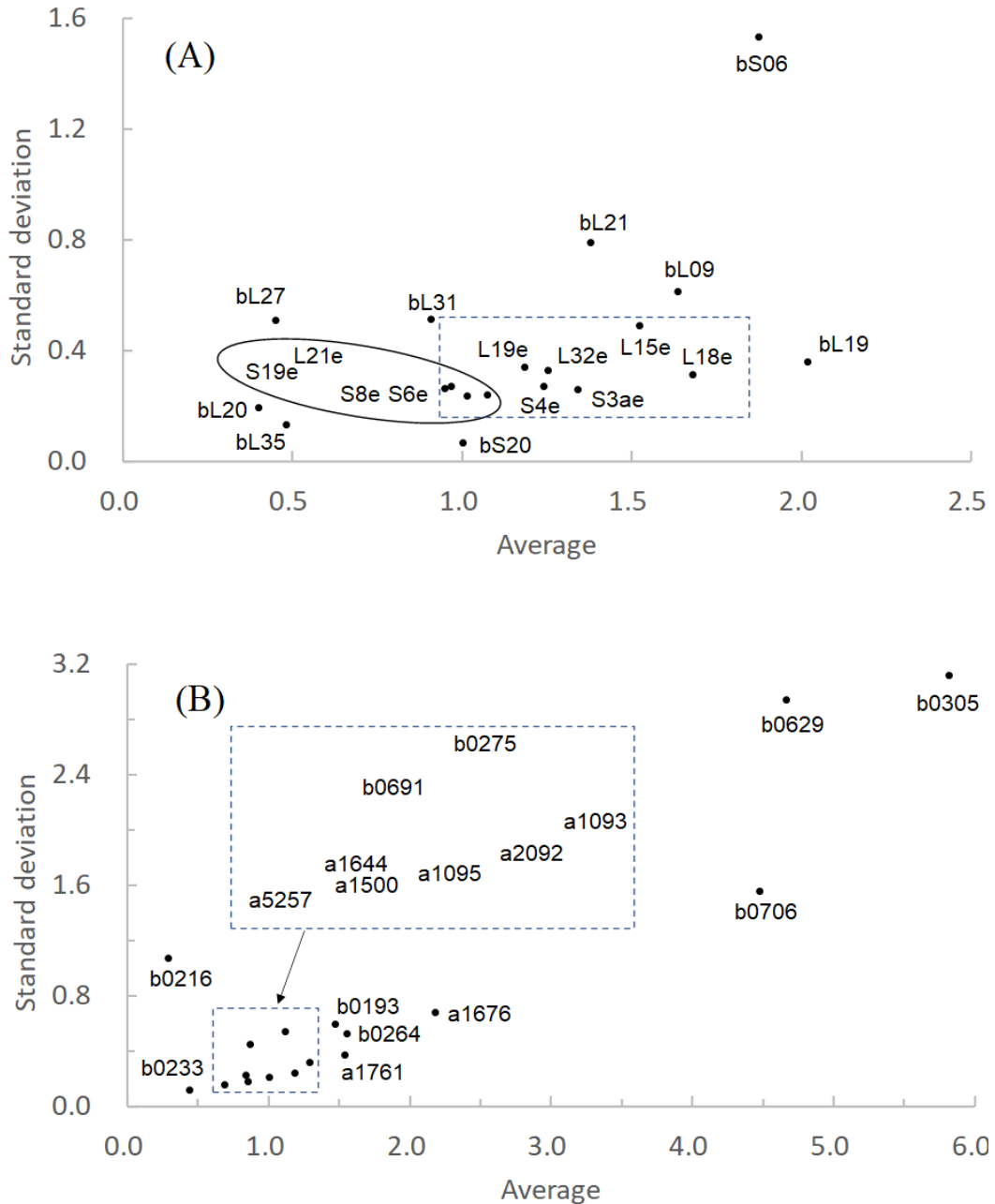


Fig. 1. Average and standard deviation of root distances for COGs of archaea- and bacteria-specific. (A) r-protein groups except r-protein group bL7 (average 8.681, standard deviation 0.410). L and S represent large and small subunit, respectively, and used instead of COG ID (Table 2). The prefix b indicates bacteria and suffix e indicate archaea, respectively. The dashed box represents the range of archaea-specific r-protein groups and the order of the legends inside the ellipse corresponds to the order of the dots. (B) non r-protein groups. The legends inside the large dashed box corresponds to the order of dots in the small dashed box. The prefix a and b indicates archaea and bacteria, respectively, and COG was omitted prior to numbers.

평균과 표준편차를 나타낸다. 평균 2.5를 넘는 것이 r-단백질 그룹(Fig. 1A)에서는 bL7 (COG0222) 1개였고, r-단백질 이외 그룹(Fig. 1B)에서는 3개였다. 표준편차 0.8을 넘는 것도 r-단백질 그룹(Fig. 1A)에서는 1개였지만 r-단백질 이외 그룹(Fig. 1B)에서는 4개였다. 고균과 세균, r-단백질과 r-단백질 제외 그룹의 4가지 조합에서 평균과 표준편차는 모두 통계적 유의성은 낮았다($p > 0.05$).

3.3. 각 문(phylum) 특이 COG

다른 문에서는 발견되지 않고 해당 문에서만 발견되는 COG는 8개 문과 5개 강에서 확인되었다(자료 미제시). 하지만 해당 분류군의 구성원 모두가 갖고 다른 원핵생물에 없는 COG는 총 10개로 Crenarchaeota 문의 COG4920 (Uncharacterized membrane protein, DUF2208 domain)가 1개이고, Cyanobacteriota 문이 9개로 COG5674 (NAD(P)H dehydrogenase cyanobacteria/chloroplast subunit O), COG5678 (Cytochrome b6f complex cytochrome f subunit), COG5743 (Photosystem I light harvesting complex protein, LHCA5/LHCA6/HLI5/Ycf17 family), COG5758(RNA-binding protein Tab2, required for psaB translation), COG5769 (Cyanobacterial/chloroplast membrane protein Tic20/Ycf60, involved in chloroplast protein import), COG5775 (Thylakoid membrane protein DAC/CCR2), COG5781 (Cytochrome b6 maturation protein CCB2), COG5783 (Chloroplast senescence-associated protein CSAP, DUF1997 (PF09366) domain), COG5784 (Uncharacterized cyanobacterial/plant protein Pro1025, DUF3531 (PF12049) domain), COG5791 (NAD(P)H dehydrogenase cyanobacteria/chloroplast-specific auxiliary subunit CRR7)이었다. Cyanobacteriota 문의 모든 구성원만 갖는 COG중 COG5784를 제외하고는 모두 광합성과 연관이 있는데(자료 미제시) Cyanobacteria는 광합성에 광계(photosystem) I과 II를 모두 사용하고 산소를 발생시키지만, 다른 광합성 원핵생물은 광계 I이나 II와 유사한 Reaction Center (RC) I이나 II중 하나만 사용하고 산소를 발생시키지 않으므로(Fischer et al., 2016) Cyanobacteriota 문의 구성원 전체만이 갖는 COG를 이해할 수 있었다. Crenarchaeota 문의 COG

4920인 uncharacterized membrane protein, DUF2208 domain은 운동에 관여하는 archaellum의 일부일 가능성이 제기되었지만(Albers and Jarrell, 2015) 아직 기능을 파악하지 못한 상태이다. 해당 분류그룹의 구성원 모두가 아닌 일부 구성원에만 존재하고 다른 분류그룹에는 없는 COG는 221개였다(자료 미제시).

COGs 자료(Galperin et al., 2025)는 Bacillota와 Pseudomonadota 문을 강까지 분류하였다(Table 1). 355개 Bacillota 문 구성원 전부에 43개의 COG가 존재하였지만(Table 1) 이들은 Bacillota 문을 제외한 807-1,930개의 원핵생물에도 있었다(자료미제시). 808개 Pseudomonadota 문 구성원 전부에 있는 COG는 없었고(Table 1) 806개의 구성원에 3개 (COG0564, 0568, 0539) 등 800개 이상의 구성원에 23개의 COG가 있었지만 이들은 Pseudomonadota 문을 제외한 847-1,483의 원핵생물에도 있었다(자료 미제시).

3.4. 각 분류군 특이 부재 COG

Table 3에 다른 분류군에서는 발견되는데 해당 분류군에서만 발견되지 않는 170개의 COG를 각 분류군에 맞추어 '2,296개의 원핵생물 중 해당 COG를 갖는 구성원수 (COG ID)'로 표시하였고 18개 문과 1개 강 내에서는 확인된 COG를 가진 원핵생물 개수의 내림차순으로 정렬하였다. 2,296개 원핵생물 중 다른 분류군의 구성원들은 모두 갖고 해당 분류군의 모든 구성원에만 없는 COG는 없었다. Crenarchaeota 문의 구성원은 29개인데(Table 1) 구성원 모두에 없는 24개의 COG는 2,231개(COG0484)에서 1,217개(COG3177)의 원핵생물에 존재하여, 연구대상 전체 2,296개의 원핵생물에서 이들 COG가 없는 다른 문의 구성원도 있다는 것을 알 수 있다. 해당 분류군의 구성원 모두에 없는 COG는 다른 원핵생물 666개(COG1373, Predicted ATPase)에서 2,231개(COG0484, DnaJ molecular chaperone)에 있었다. 해당 분류군만 갖지 않는 단백질의 저해제를 탐색/개발하고 사용하면 체내 유익균에는 작용하지 않고 다른 분류군의 성장 억제를 하는 등의 특이성 높은 항생제 혹은 선택배지 개발이 가능할 것인데, 당장은 쉽지 않음을 알 수 있었다.

Table 3. COGs that are not present in all members of each phylum and other taxonomic levels. Prefix COG was omitted

Taxon	COGs only absent in each taxonomic level: # of prokaryotes having (COG ID)
Archaea	
Asgardarchaeota	1,926 (2265), 1,915 (0108), 1,860 (0039), 1,719 (1642), 1,651 (0168), 1,631 (0591)
Crenarchaeota	2,231 (0484), 2,222 (0443), 2,216 (0576), 2,170 (0210), 2,134 (0322), 2,125 (0556), 2,122 (0745), 2,098 (0036), 2,086 (0042), 2,040 (0192), 2,027 (0236), 1,992 (0317), 1,984 (0652), 1,924 (0394), 1,895 (0133), 1,871 (1278), 1,834 (0229), 1,787 (5979), 1,674 (2204), 1,646 (0784), 1,531 (2203), 1,284 (1321), 1,256 (5975), 1,217 (3177),
Nitrososphaerota	2,188 (0563), 2,162 (0101), 2,159 (1132), 1,957 (0345), 1,927 (0332), 1,834 (1045), 1,677 (1902), 1,575 (1473), 1,559 (0554), 1,488 (1894), 1,437 (1905), 1,276 (2956), 1,098 (1943)
Thermoplasmata	1,993 (1968), 1,932 (0212), 1,861 (2097), 1,839 (0486), 1,763 (0511), 1,587 (1409), 1,354 (0381), 1,312 (0058), 1,894 (0318), 1,835 (0824), 1,092 (4106)
Other Archaea	2,054 (0735), 1,833 (1087), 1,575 (0602), 1,020 (3547)
Bacteria	
Aquificota	1,902 (0389), 1,679 (2,267), 1,672 (0110), 1,654 (2271), 1,470 (3153), 1,426 (1961), 1,370 (0599), 1,283 (1704)
Bdellovibrionata	1,785 (0493), 1,251 (1254)
Chlamydiota	2,070 (0152), 2,066 (0041), 2,060 (0034), 2,058 (0150), 2,054 (0330), 2,053 (0151), 2,051 (0078), 2,049 (0047), 2,043 (0046), 2,028 (0079), 2,017 (0498), 1,996 (0460), 1,935 (0299), 1,921 (0107), 1,912 (0226), 1,903 (0118), 1,848 (1117), 1,828 (0581), 1,806 (0704), 1,800 (2608), 1,787 (1522), 1,726 (0573), 1,625 (0286), 1,624 (0005), 1,584 (0235), 1,529 (2025), 1,523 (2086), 1,507 (1014), 1,501 (0846), 1,431 (0392), 1,396 (0655), 1,388 (0038), 1,349 (0543), 1,330 (0370), 1,305 (0075), 1,193 (1193), 1,160 (0863), 1,155 (3620), 1,128 (1180), 1,125 (4231), 1,121 (0798)
Chlorobiota	2,041 (0624), 1,817 (1502), 1,716 (0183), 1,716 (1960), 1,671 (2220), 1,636 (1488), 1,625 (0402), 1,575 (0515), 1,505 (0010), 1,494 (1199), 1,263 (3448), 1,082 (0785)
Deferribacterota	1,837 (0508), 1,772 (0665), 1,748 (0719), 1,743 (0396), 1,715 (0695), 1,498 (0367), 1,476 (1240), 1,446 (1071), 1,445 (0022)
Deinococcota	1,443 (1454)
Erysipelotrichia	1,676 (0156), 1,659 (0616), 1,456 (3063), 1,365 (1230), 1,213 (0651), 850 (1814)
Mycoplasmota	2,133 (0472), 2,021 (0505), 1,959 (0473), 1,952 (0147), 1,937 (0512), 1,922 (1053), 1,893 (0159), 1,881 (0134), 1,876 (0547), 1,863 (0372), 1,801 (0685), 1,715 (0001), 1,708 (0135), 1,580 (0379), 1,494 (1234), 1,432 (1994), 1,416 (0794), 871 (1586), 764 (1203)
Myxococcota	1,024 (1328)
Negativicutes	2,074 (0558), 1,335 (0025)
Pseudomonadota	
Other Proteobacteria	1,590 (2211), 666 (1373)
Synergistota	1,620 (0644)
Tissierellia	1,399 (3264)
Other Bacteria	1,965 (1253), 1,871 (1605), 1,868 (0708), 1,744 (0531), 1,474 (0419), 1,449 (3210), 800 (0338)

4. 결 론

2,296개의 원핵생물 각각이 가진 COG (cluster of orthologous group of proteins)를 바탕으로 고균과 세균의 계(kingdom)와 문(phylum) 단위별로 공통적으로 갖거나 갖지 않는 COG를 탐색하였다. 문

의 구성원 수와 구성원 모두가 공통으로 갖는 ($r=-0.508$) 혹은 공통으로 갖지 않는($r=-0.636$) COG 개수 사이에는 음의 상관관계가 있었다. 193개의 고균에서만 발견되는 212개 COG중 190개 이상의 고균들은 리보솜 구성 단백질(r -단백질) 10개 등 18개 COG를 가졌다. 2,103개의 세균에서만 발견되

는 1,572개의 COG중 2,060개 이상의 세균은 r-단백질 10개 등 19개의 COG를 가졌다. 고균과 세균의 r-단백질과 다른 단백질의 계통수에서 구한 평균과 표준편차의 분포는 통계적 유의성이 높지 않았다($p > 0.05$). 원핵생물의 다른 문에는 없고 각 문의 구성원 모두가 갖는 COG는 10개로 Crenarchaeota 문의 COG4920 (uncharacterized membrane protein, DUF2208 domain)와 Cyanobacteriota 문의 광합성 관련 8개의 COG와 COG5784 (uncharacterized cyanobacterial/plant protein Pro1025) 등이었다. 다른 분류군에서는 발견되는데 해당 문이나 강(class)에서만 발견되지 않는 170개의 COG가 18개 문과 1개 강에 있었지만, 다른 분류군의 구성원들은 모두 갖고 해당 분류군의 모든 구성원에만 없는 COG는 없었다. 본 연구결과는 각 원핵생물 특이적 항생제나 선택배지 개발 및 미생물 군집구조 파악에 필요한 기초자료 등을 제공할 수 있을 것이다.

감사의 글

이 논문은 2026년도 BB21+사업에 의하여 지원되었음

REFERENCES

- Albers, S. V., Jarrell, K. F., 2015, The archaeallum: How Archaea swim, *Front Microbiol.*, 6(1), 23.
- Fischer, W. W., Hemp, J., Johnson, J. E., 2016, Evolution of oxygenic photosynthesis, *Annu. Rev. Earth Planet. Sci.*, 44, 647-683.
- Fogel, G. B., Collins, C. R., Li, J., Brunk, C. F., 1999, Prokaryotic genome size and SSU rDNA copy number: Estimation of microbial relative abundance from a mixed population, *Microb. Ecol.*, 38(2), 93-113.
- Galperin, M. Y., Vera Alvarez, R., Karamycheva, S., Makarova, K. S., Wolf, Y. I., Landsman, D., Koonin, E. V., 2025, COG database update 2024, *Nucleic Acids Res.*, 53(D1), D356-D363.
- Huson, D. H., Scornavacca, C., 2012, Dendroscope 3: An Interactive tool for rooted phylogenetic trees and networks, *Syst. Biol.*, 61(6), 1061-1067.
- Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., Tamura, K., 2024, MEGA12: Molecular evolutionary genetic analysis version 12 for adaptive and green computing, *Mol. Biol. Evol.*, 41(12), 1-9.
- Lee, D. G., 2025, Conserved genes distributed among 2,296 prokaryotes, *J. Life Sci.*, 35(2), 202-207, <https://doi.org/10.5352/JLS.2025.35.2.202>
- Lee, D. G., 2026, Comparison of conservation of ribosomal proteins distributed among 2,296 prokaryotes, *J. Life Sci.*, 36(3), 258-264, <https://doi.org/10.5352/JLS.2026.36.3.258>
- Lee, D. G., Kim, A., Lee, S. H., 2025, Comparison of conservation of COGs common to over 2,260 prokaryotes, *J. Life Sci.*, 35(12), 1150-1154, <https://doi.org/10.5352/JLS.2025.35.12.1150>
- Naushad, H. S., Lee, B., Gupta, R. S., 2014, Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: Identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya*, *Pectobacterium* and *Brenneria*, *Int. J. Syst. Evol. Microbiol.*, 64(Pt 2), 366-383, <https://doi.org/10.1099/ijs.0.054213-0>
- Segerman, B., 2012, The genetic integrity of bacterial species: The core genome and the accessory genome, two different stories, *Front. Cell Infect. Microbiol.*, 2, 116.
- Wang, Y., Huang, J., 2018, Assessment of molecular markers for classification of bacterial phyla using topological dissimilarity of phylogenetic trees, *J. Phylogenetics Evol. Biol.*, 6(3), e1000204.

-
- Professor. Sang-Hyeon Lee
Division of Culinary Arts, College of Life Science, Silla University
slee@silla.ac.kr
 - Professor. Dong-Geun Lee
Department of Pharmaceutical and Cosmetics, College of Medical and Life Science, Silla University
ldg@silla.ac.kr
 - Professor. Andre Kim
Department of Pharmaceutical and Cosmetics, College of Medical and Life Science, Silla University
adrk@silla.ac.kr