

Conserved COG Pathways and Genes of 122 Species of Archaea

Dong-Geun Lee and Sang-Hyeon Lee*

Department of Pharmaceutical Engineering, Silla University, Busan 46958, Korea

Received September 1, 2023 /Revised September 27, 2023 /Accepted October 4, 2023

The purpose of this study was to identify conserved metabolic pathways and conserved genes in 122 archaeal species. Using the Clusters of Orthologous Groups of Proteins (COG) database of conserved genes, we analyzed whether 122 species had 63 COG metabolic pathways, the 822 COGs that compose them, and a total of 4,877 COGs. Archaeal ribosomal proteins were the most conserved in metabolic pathways. 46 COGs in seven COG pathways among 63 COG pathways and 20 COGs in others were conserved in 122 species. Some genes involved in cell wall and extracellular matrix synthesis, replication, transcription, translation, and protein metabolism were common to all 122 species. When the distance value of the phylogenetic tree was analyzed at the phylum level or class level, the average was the lowest at the class Halobacteria of the phylum Euryarchaeota. Standard deviation was high for the class Nitosphaeria of the phylum Thaumarchaeota, the unclassified members of phylum Thaumarchaeota, the class Halobacteria of the phylum Euryarchaeota, the class Thermoprotei of the phylum Crenarchaeota, and other archaea. Furthermore, the phylogenetic tree analysis revealed six commonalities. The results of this study, along with data on conserved genes, could be used for drug development and gene selection for strain improvement.

Key words : Archaea, COG (cluster of orthologous groups of proteins), conserved gene, conserved metabolic pathways

서 론

진화의 결과로 다양한 생명체들이 지구에 존재하며, 생명체들 사이에는 물질대사와 번식 등의 몇 가지 공통점이 있다. 생명체들 사이에서 공통된 기능을 수행하는 단백질이 있다면 유전자의 염기서열이 달라도 공통된 유전자를 갖고 있다고 할 수 있다. 공통조상의 유전자는 종분화(speciation)로 여러 자손 종들에 분포하고 동일한 기능을 나타내는데, 이런 유전자들의 집합을 ortholog라고 하고 동일한 ortholog에서 유래하여 3종(species) 이상에 존재하는 단백질의 집합을 COG (Cluster of Orthologous Groups of proteins)라고 하며, 2023년 현재 4,877개의 COG가 보고되었다[13]. 이러한 보존적 유전자에 대한 연구는 생명에 대한 기초 자료를 제시할 뿐만 아니라 응용 측면에서 균주의 개선을 위해 선택하는 유전자의 선택[9]과 항암 의약품 개발[3] 등에 활용될 수 있는 유용성도 있다.

Lee와 Lee [11]는 원핵생물 711개에 보존적인 유전자를 보고하면서 ribosomal large subunit인 COG0080이 711개의 원핵생물에 공통적이라고 보고하였다. Lee와 Lee [12]는 168개의 고세균 모두에 14개의 보존적 유전자가 있고 이들이 tRNA synthetase, ribosome, phe-tRNA의 변형된 염기인 wybutosine 합성 효소, tRNA 성숙과 연관이 있다고 보고하였지만 하나의 종에 10개의 균주가 사용되는 등 종의 수는 168개에 훨씬 모자랐다. 한편, COG 데이터베이스는 2020년에 업데이트 되어 1종에 1개의 균주로 하며 1,309종의 원핵생물을 포함하고 있다. Lee와 Lee [13]는 1,309종의 원핵생물이 보유한 COG들을 보고하였고, Lee 등[14]은 1,309종의 원핵생물에 대해 COG 자료에서 제시하는 물질대사인 63개의 COG pathways에 대해 보고하였다.

원핵생물은 의학, 산업, 기초과학의 측면에서 중요한 연구대상이다. 이들은 주변 환경에 따라 생명현상을 조절하며 현재 각 서식지에서 역할을 수행하고 있다[14]. 고세균은 지구에 최초로 출현한 생명체로 알려져 있어 이들 고세균들에 대한 이해는 지구의 생명체를 이해하는데 도움이 될 것이다. 본 연구에서는 2023년 현재의 COG 자료를 구성하는 122종의 고세균에 대하여 63개의 COG pathways와 이 경로들을 구성하는 822개의 COG들을 파악하여 고세균과 생명에 대한 이해를 높이고자 하였다.

*Corresponding author

Tel : +82-51-999-5624, Fax : +82-51-999-5628

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.

재료 및 방법

분석재료

COG 데이터베이스는 Lee와 Lee [12]의 보고 후에 업데이트되어 2023년 현재까지 유지되는 데이터베이스를 이용하였고[13] 1,309종의 원핵생물 중에서 고세균은 총 122종이고 전체 COG는 4,877개였다. 고세균에 속하는 122종의 자료를 생성하고 분석에 이용하였다. 분석대상 고세균이 가지는 COG 데이터베이스를 생성한 후, COG pathways [6]에서 정한 63개의 대사경로와 63개의 대사경로를 구성하는 822개의 COG 종류들을 파악하였으며, 전체 4,877개의 COG 데이터베이스에서 각 고세균이 가지는 COG 들을 확인하였다. Table 1은 실제로 분석한 122종의 고세균을 강(class) 수준에서 각 강을 구성하는 종(species)의 개수를 나타내고 있다. 문(phylum)은 파악되었지만 강을 알 수 없는 종은 Euryarchaeota 문에 1개(EU-Ot), Thaumarchaeota 문에 5개(TH-Ot)였다. Crenarchaeota 문의 25개 종은 모두 Thermoprotei 강에 속하였다(Table 1).

대사경로와 대사경로별 COG 보유 여부 파악

Perl 프로그램과 엑셀 프로그램을 이용하여 COG에서 제시하는 63개의 COG pathways [6], 63개의 COG pathways의 각 대사경로를 구성하는 822개의 COG와 전체 4,877개의 COG를 각 고세균 종이 보유하는지를 파악하였다.

대사경로 orthologs를 이용한 유전체 분석

대사경로 및 대사경로 관련 보존적 유전자 등의 보유 계통수를 작성하였다. 행렬로 만든 분석대상 고세균들의 자료들을 MegaX 프로그램[10]으로 다중서열비교를 수행한 후, Maximum-Likelihood와 Neighbor-joining 방법으로 Presence-Absence phylogenetic tree를 제작하였다. 이때 bootstrap method (n=1,000)를 이용하였고 distance value를 담고 있는 nwk 파일을 생성하였다. Distance value는 Dendroscope 프로그램(ver 0.8)으로 얻었고[7], distance value 등의 분석에는 엑셀 프로그램을 이용하였다. *E. coli* K12 MG1655균주를 outgroup 으로 설정하였다.

결과 및 고찰

COG pathways의 수와 구성 COG의 수

Lee 등[14]은 완전한 하나의 대사경로가 보유하는 COG의 개수를 분모로 하고 각 세균종이 보유하는 해당 대사경로의 COG의 개수를 분자로 하는 대사경로 구성율의 개념을 제시하면서 63개의 COG pathway [7]에 대해 1,309종의 원핵생물 각 구성원이 0~26개만이 완전한 “대사경로 구성률”을 보이며, 63개의 COG pathways 모두에 결합이 있는 원핵생물이 53종이라고 보고하면서 대사경로를 구성하는 COG가 완전하지 않아도 대사가 이루어지는 가능성을 토의하였다. 또한, 대사경로에 대한 정보를 담고 있는 MetaCyc에서도 하나의 대사경로에서 빠진 효소를

Table 1. Phylogenetic groups, numbers of studied species, and average (range) among 63 COG pathways, 822 COGs in 63 COG pathways, and all 4,877 COGs

Phylogenetic Group Phylum Class	Abbreviations	# of species	Average (range) among		
			63 COG pathways	822 COGs of 63 COG pathways	all 4,877 COGs
Crenarchaeota					
Thermoprotei	CR	25	11.1(3~18)	364.9 (273~428)	947.1(783~1119)
Euryarchaeota					
Archaeoglobi	EU-A	3	14.3(13~15)	405.7(399~411)	1232.0(1176~1266)
Halobacteria	EU-H	31	20.3(16~25)	424.4(339~450)	1253.6(860~1413)
Methanobacteria	EU-Mb	5	13.0(11~15)	385.0(372~405)	1071.4(960~1212)
Methanococci	EU-Mc	4	12.8(11~14)	392.5(386~399)	1115.8(1068~1150)
Methanomicrobia	EU-Mm	18	15.6(11~19)	401.4(373~443)	1257.2(1084~1521)
Methanopyri	EU-Mp	1	11	381	1044
Thermococci	EU-Tc	4	11.8(10~15)	368.8(352~395)	1097.3(1077~1108)
Thermoplasmata	EU-Tp	10	12.9(11~16)	371.4(354~396)	916.2(861~966)
Candidatus Nanohaloarchaea	EU-Cn	1	2	326	918
Other	EU-Ot	1	8	202	557
Candidatus Thermoplasmatota	EU-CT	1	9	431	1309
Thaumarchaeota					
Nitrososphaeria	TH-Ns	7	14.0(13~15)	368.3(342~386)	927.9(811~1026)
Others	TH-Ot	5	12.5(12~13)	360.6(346~370)	862.3(827~895)
Other Archaea	OA	6	11.7(3~23)	240.7(126~433)	661.3(332~1289)

pathway hole이라고 정의하며[2], genome에서 해당 유전자가 발견되지 않았거나 다른 경로를 통해 대사를 완성할 가능성을 제시하고 있다[8]. 이에 본 연구에서는 각 대사 경로를 구성하는 COG의 비율이 70% 이상이면 해당 대사 경로를 가진 것으로 설정하였다. 이 기준으로는 122종의 고세균 모두가 25개의 대사경로가 없었으며, 122종의 고세균 모두에 공통적인 대사경로는 없었다. COG pathways는 MetaCyc와 달리 리보솜을 구성하는 단백질과 번역인자들을 대사경로에 포함시켰다[14]. 가장 많은 고세균들에 공통적인 대사경로는 archaeal ribosomal proteins로 120종의 고세균에 존재하였고 archaeal ribosomal proteins를 70% 미만으로 가지는 2개의 종은 other Archaea의 *Candidatus Nanopusillus acidilobi*와 Eubacteria의 *Candidatus Nanohaloarchaea* 강에 속하는 *Nanohaloarchaea archaeon* SG9로 각각 58%와 67%였다. 그 다음으로 많은 고세균에 분포하는 대사는 pyrimidine biosynthesis가 117종에, A/V-type ATP synthase가 112종, ribosome 30S subunit가 111종의 고세균에 존재하여 단백질, 핵산, 그리고 ATP 합성이 고세균에 중요한 것으로 유추되었다.

고세균의 강 단위에서 대사경로의 수가 가장 많은 것은 Euryarchaeota 문의 Halobacteria 강이 20.3개이고 Crenarchaeota 문의 Thermoprotei 강이 11.1개로 최저였다(Table 1). 63개의 COG pathways 중 0개를 가져 최소인 고세균은 other Archaea의 *Candidatus Nanopusillus acidilobi*로 이 종은 Crenarchaeota에 속하는 *Acidilobus* sp. 7A를 숙주로 하여 세포외부에 기생하며 세포에 필요한 전구체 생산을 숙주에 의존하는 것으로 알려져 있다[17]. 다음으로 3개의 대사경로를 가진 고세균은 2종으로 Crenarchaeota 문의 *Thermofilum adornatus*가 archaeal ribosomal proteins, molybdopterin biosynthesis, pyrimidine biosynthesis를 가졌고, other Archaea에 속하는 *Nanoarchaeum equitans* Kin4-M이 aminoacyl-tRNA synthetases, archaeal ribosomal proteins, ribosome 30S subunits를 가졌다. Archaeal ribosomal proteins는 두 고세균에서 공통으로 다른 대사경로는 없어도 고세균 특이적인 리보솜은 필요한 것을 알 수 있었다.

63개의 COG pathways를 구성하는 822개의 COG에 대한 보유 평균이 가장 많은 강은 Halobacteria 강(424.4개)이었고, 최소는 other Archaea (240.7개)였다(Table 1). 분류가 가능한 수준에서는 Thaumarchaeota 목에서 강을 알 수 없는 고세균들이 360.6개로 최소였고, Crenarchaeota 목의 Thermoprotein 강이 364.9개로 다음이었다. 822개의 COG 중에서 보유한 COG 수가 적은 종은 모두 other Archaea에 속하였는데, 최소부터 나열하면 *Candidatus Nanopusillus acidilobi* (126개), *Nanoarchaeum equitans* Kin4-M (149개), *Candidatus Mancarchaeum acidiphilum* Mia14 (179개), archaeon GW2011 AR15 (231개), *Candidatus Korarchaeum cryptofilum* OPF8 (334개)로 이들 균주들은 단독생활이 불가능하여 기생 혹은 체외공생을 영위하여 독립적인 대사경로가 부족한 것으로 토의되었다[14].

보존적 유전자의 기능

Table 2에 연구대상 122종의 고세균 모두에서 발견되는 COG들을 63 COG pathways의 대사경로를 구성하는 COG와 그 외(Others)로 구분하였다. Others는 전체 4,877개의 COG들 중에서 63 COG pathways에 속하는 822개의 COG를 제외하였을 때, 122종의 고세균 모두에 공통적인 것이었다. 63개의 COG pathways 중에서 7개의 COG pathways에 속하는 COG가 46개, 그리고 others가 20개였다. Ribosome을 구성하는 3개의 pathways에 속하는 것이 29개, tRNA synthetase와 전사인자가 각각 5개, RNA polymerase를 구성하는 16개 중 3개, 그리고 tRNA modification에 관련된 59개 중에 2개의 COG가 공통적이었다. 63개의 COG pathways에 속하지 않고 122종의 고세균에 공통적인 보존적 유전자는 복제(COG0258, COG0358, COG0417), 전사(COG0250, COG1094, COG1405, COG2101), 번역(COG0012, COG0201, COG1990, COG2511, COG2890), 세포벽과 세포외기질의 합성(COG0438, COG1047), 번역후변형(COG0024, COG1899), ATP 등의 뉴클레오티드의 대사(COG0464, COG1936), 단백질 접힘(COG0459), 단백질 분해(COG0638) 등에 관련되었다. 고세균은 지구에 최초로

Table 2. Common COGs and their metabolic pathways in all 122 archaeobacteria (Numbers in parenthesis are total numbers of conserved genes in that pathway and prefix COG was omitted in COGs columns)

Pathways	COGs
P53 Ribosome 30S subunit (21)	0092, 0522, 0098, 0049, 0096, 0103, 0051, 0100, 0099, 0186, 0185
P05 Archaeal ribosomal proteins (33)	1358, 2058, 1727, 2147, 1997, 1631, 1890, 1471, 2007, 2238
P54 Ribosome 50S subunit (33)	0081, 0087, 0088, 0244, 0080, 0093, 0089, 1841
P04 Aminoacyl-tRNA synthetases (26)	0017, 0124, 0180, 0423, 0495
P61 Translation factors (21)	0532, 1093, 1503, 5256, 5257
P51 RNA polymerase (16)	0086, 1095, 1761
P63 tRNA modification (59)	3642, 2520
Others	0012, 0024, 0201, 0250, 0258, 0358, 0417, 0438, 0459, 0464, 0638, 1047, 1094, 1405, 1899, 1936, 1990, 2101, 2511, 2890,

출현한 생물체로 생각되며, 현재 발견되는 고세균들의 보존적 유전자는 초기의 고세균이 가진 보존적 유전자와 변화한 지구환경에서 필요한 보존적 유전자의 조합일 가능성이 높을 것이다. 그런 관점에서 외부와 세포질을 구분 짓는 세포벽과 세포외기질의 합성, 복제, 전사, 번역, 단백질의 대사와 관련된 유전자들이 공통적인 core genome이 되고, 각 고세균이 접한 환경에서 필요한 유전자들이 차이를 보인다고 할 것이다.

분류그룹별 COG pathways 등의 분포

분석대상 고세균들의 대사경로 혹은 COG의 보유 유무에 따른 presence/absence 계통수를 maximum-likelihood와 bootstrap (n=1,000)을 적용하여 작성한 결과를 토대로 계통수에서 구한 각 고세균의 distance value의 평균과 표준편차를 Table 1의 분류단위별로 구하였다(자료미제시). 공통점은 EU-H (Euryarchaeota문의 Halobacteria강)의 평균이 가장 낮으며 다른 분류그룹과 분리되어 있었다. 어느 분류단위의 distance value의 평균이 낮다는 것은 phylogenetic tree의 root에 가깝다는 것이다. Outgroup으로 사용한 *E. coli* K12의 root distance가 가장 높았다. 따라서 대장균과 고세균들의 공통조상에서 변이가 가장 적은 것이 EU-H라고 해석할 수 있었다. 각 구성원이 1개인 EU-Cn, EU-Ot, EU-CT, EU-Mp 등을 제외하면 63개의 COG pathways에서는 OA와 TH-Ns, TH-Ot가, 63개의 COG pathways를 구성하는 822개의 COG에서는 EU-Tc와 EU-Mc가, 전체 4,877개의 COG에서는 OA와 EU-Mb가 평균값이 높아 대장균과 각각 가장 거리가 먼 것으로 해석되었다.

분류단위가 문(phylum)까지 확정된 것들은 비교하면 표준편차가 높은 분류그룹은 COG pathways 등의 분석기준에 따라 각각 달랐다. 표준편차는 63개의 COG pathways에서는 TH-Ns (0.367), TH-Ot (0.351), EU-H (0.293), OA (0.243)의 순서로 높았고, 63개의 COG pathways를 구성하는 822개의 COG에서는 OA (0.275), CR (0.085), TH-Ot (0.084)의 순서로 높았으며, 전체 4,877개의 COG에서는 OA (0.154) 다음으로 EU-Mb (0.035)와 EU-H (0.035)가 공동 2위였다. 계통에서 개별 유전자의 생성 혹은 소실의 속도가 종에 상관없이 일정하다고 가정하면 진화적으로 뒤에 생성된 강(class)보다 먼저 생성된 강을 구성하는 각 생물종들이 유전자 보유 유무에 편차가 높을 것이며, 여러 개의 유전자로 구성된 대사경로의 보유는 상대적으로 편차가 낮을 것이다. 하지만 63개의 COG pathways보다 이를 구성하는 822개의 COG가 더 낮은 편차를 보였고 4,877개의 COG가 가장 낮은 편차를 보였다. 정량적으로 각 분류단위의 구성원의 보유범위(Table 1)를 비교대상의 수로 나누면 63개의 COG pathways는 $22/63=0.35$, 822개의 대사구성 COG는 $203/822=0.25$, 4,877개의 전체 COG는 $710/4877=0.15$ 였다. 따라서 예상과 다르게 편차가 나타난

이유는 비교하는 형질의 수가 다른 점과 비교 형질의 수에 비해서 각 분류단위 내에서 형질 차이의 범위에 의해 나타난 결과로 유추되었다.

분류단위는 일정하지만 63, 822, 4,877개 등의 비교하는 형질 숫자에 차이가 있어도 산포도를 확인할 수 있는 변동계수를 구하여 비교하였다. 전반적으로 EU-H, OA, TH-Ot가 높았다. 63개의 COG pathways에서는 EU-H (142.05), TH-Ot (64.97), TH-Ns (60.92), OA (35.88)의 순서로 높았고, 63개의 COG pathways를 구성하는 822개의 COG에서는 OA (46.31), EU-H (44.29), TH-Ot (24.93), CR (22.48)의 순서로 높았으며, 전체 4,877개의 COG에서는 OA (35.27), EU-H (27.02), EU-Mb (8.19)의 순서였다.

계통수

Fig. 1은 분석대상 고세균들의 대사경로 혹은 COG의 보유 유무에 따른 presence/absence 계통수를 maximum-likelihood와 bootstrap (n=1,000)을 적용하여 작성한 것으로 122개의 고세균을 Table 1에서 약어로 표시된 분류단위 2개까지 각 분지 말단에 표시하였고 분류단위 앞의 숫자는 해당 분류단위에 속하는 종의 개수이다. 63개의 COG pathways로 작성한 계통도는 122종의 고세균이 57개의 분지 말단으로 복잡하여 생략하였다. 63개의 COG pathways를 구성하는 822개의 COG를 이용한 계통수(Fig. 1A)와 4,877개의 COG를 이용한 계통수(Fig. 1B)의 공통점은 첫째, outgroup인 *E. coli* 가까이 EU-H가 위치하고, 둘째, OA05가 EU-H와 함께 위치하고, 셋째, EU-Mc, EU-Mb, EU-Tc, EU-A의 구성원끼리 가장 가까우며, 넷째, TH-Ot1 (*Candidatus Caldiarchaeum subterraneum*)이 OA01 (*Candidatus Korarchaeum cryptofilum* OPF8)과 가장 가까우면서 둘다 CR (Crenarchaeota)과 연관성이 높은 것이며, 다섯째, TH-Ot1을 제외한 나머지 Thaumarchaeota 문의 구성원들끼리가 가장 유사하였고, 여섯째, OA01과 OA05를 제외한 나머지 OA들끼리가 가장 유사하였다. 이러한 공통점은 neighbor-joining 방법을 이용한 계통수에서도 동일하였다(자료미제시).

계통수에서 EU-H와 유사한 것으로 파악된 OA05 (archaeon AArc-SI)는 other archaeobacteria로 COG site에서 분류하고 있지만 NCBI의 taxonomy에서는 Euryarchaeota문의 Halobacteria강에 속하는 *Halalkaliarchaeum desulfuricum*로 재분류가 되어있어 보존적 유전자의 산물인 COG를 분류에 이용할 수도 있을 것으로 판단되었다. OA01 (*Candidatus Korarchaeum cryptofilum* OPF8)은 NCBI의 taxonomy에서 *Candidatus Korarchaeota* 문으로 분류되어 있다. 고세균은 Euryarchaeota와 Crenarchaeota 두 개의 문만 존재하다가 중온성 Crenarchaeota인 Thaumarchaeota 문이 생기고 [1], 분류학적으로 불완전하며 호냉균부터 호열균까지 다양한 환경에서 생활하며 다양한 물질 대사를 수행하는

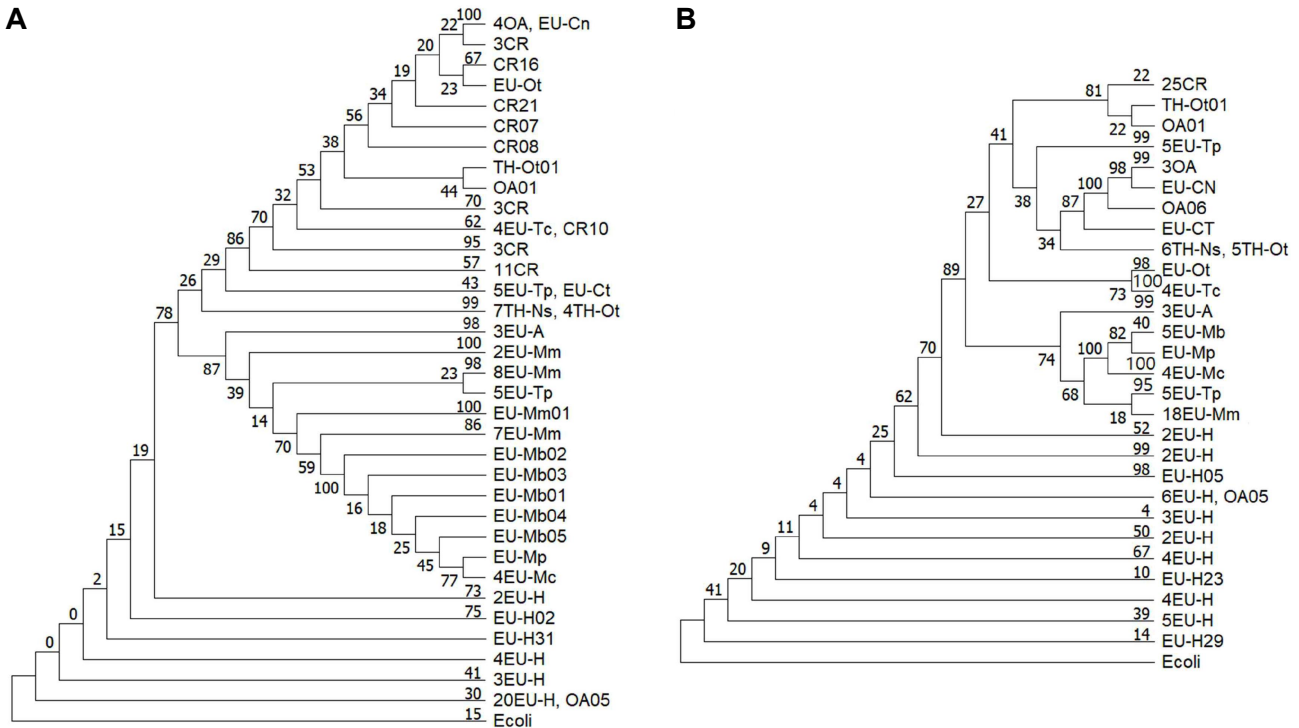


Fig. 1. The ML (Maximum Likelihood) phylogenetic tree of 122 species of archaeobacteria in the sense of presence or absence of the union 822 COGs consisting 63 COG pathways (A), and all 4,877 COGs (B).

주로 혐기성 및 화학합성 고세균의 모임인 TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota)이 제안되었다[5, 15].

Crenarchaeota는 대부분 호열균이며 Euryarchaeota는 대부분 메탄생성균이며 일부는 호염균과 호열균으로 알려져 있다. 본 연구에서 Crenarchaeota는 25종이고 Euryarchaeota는 78종이며 이 중에 메탄 관련(Methanobacteria (EU-Mb), Methanococci (EU-Mc), Methanomicrobia (EU-Mm), Methanopyri (EU-Mp) 강)이 28종, 호염성(Halobacteria (EU-H) 강)이 29종, 호열성(Thermococci (EU-Tc), Thermoplasmatata (EU-Tp) 강)이 14종, 해저 열수구에 서식하는 혐기성이며 초고온성인 Archaeoglobi (EU-A) 강이 3종이었다 (Table 1). 두 계통도 모두에서 메탄 관련의 EU-Mb, EU-Mc, EU-Mp, EU-Mm 강은 서로 인접하였고 Euryarchaeota 이면서 호열성인 EU-Tc와 EU-Tp 그리고 EU-Ot가 호열성이 대부분인 CR이 분포한 지역에 존재하였으며 초고온성인 EU-A는 메탄관련의 Euryarchaeota와 인접하였다(Fig. 2). 중온성 Crenarchaeota로 분류되었다가 독립되어 암모니아를 산화시키며 해양, 토양, 민물에 서식하는 Thaumarchaeota는 Nitrososphaeria (TH-Ns) 강[16]의 7종과 미분류된 (TH-Ot) 5종이었다. 두 계통도에서 TH-Ns와 TH-Ot는 모두 CR이 분포한 지역에 존재하였다(Fig. 1). Euryarchaeota 문에 속하다가 Candidatus Thermoplasmatota 문으로 독립 하려는 EU-CT는 EU-Tp 혹은 Thaumarchaeota (TH)와 인접

하였다.

고세균은 지구상에 최초로 출현한 생물체로 생각되며, 현재 발견되는 고세균들은 호염성, 호열성, 호냉성, 메탄 생성 등의 극한 환경 이외에도 중온성인 인간의 장내에서도 1,167개의 고세균 계통이 발견되어 고세균이 사람의 건강과 관련이 있을 것으로 보고되었다[4]. 본 연구에서 파악된 고세균의 보존적 COGs는 63 COG pathways를 구성하는 46개, 그리고 나머지 20개였다. 분석에 사용된 고세균들이 호열성, 호중성, 호염성, 메탄생성 등(Table 1)의 다양한 환경에 분포하더라도 고세균들의 생명현상을 유지하는데 필요한 것으로 유추된다. Lee와 Lee [11]는 원시 공통생명체가 보유했던 유전자가 현재까지 유지되거나, 환경 변화에 따라 공통된 유전자 추가나 소실의 결과 등이 현재의 생물체에 공통적 유전자의 기원으로 보고하였다. 본 연구는 고세균이 겪은 수십억 년의 진화과정에서 보존된 유전자의 종류에 대한 자료와 함께 의약품 개발 [3], 균주 개선을 위한 유전자 선택[9] 등에 필요한 기초자료를 제공할 수 있을 것이다.

The Conflict of Interest Statement

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Brochier-Armanet, C., Boussau, B., Gribaldo, S. and Forterre, P. 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* **6**, 245-252.
2. Caspi, R., Billington R., Keseler I. M., Kothari, A., Krummenacker, M., Midford, P. E., Ong, W. K., Paley, S., Subhraveti, P. and Karp, P. D. 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.* **48**, D445-D453.
3. Chen, D., Zhang, Z., Li, M., Wang, W., Li, Y., Rayburn, E. R., Hill, D. L., Wang, H. and Zhang, R. 2007. Ribosomal protein S7 as a novel modulator of p53-MDM2 interaction: binding to MDM2, stabilization of p53 protein, and activation of p53 function. *Oncogene* **26**, 5029-5037.
4. Chibani, C. M., Mahnert, A., Borrel, G., Almeida, A., Werner, A., Brugère, J.-F., Gribaldo, S., Finn, R. D., Schmitz, R. A. and Moissl-Eichinger, C. 2022. A catalogue of 1,167 genomes from the human gut archaeome. *Nat. Microbiol.* **7**, 48-61.
5. Guy, L. and Ettema, T. J. 2011. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580-587.
6. <https://www.ncbi.nlm.nih.gov/research/cog/pathways/>
7. Huson, D. H. and Scornavacca, C. 2012. Dendroscope 3 : An interactive viewer for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061-1067.
8. Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M. and Subhraveti, P. 2019. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform.* **20**, 1085-1093.
9. Klein-Marcuschamer, D., Santos, C. N., Yu, H. and Stephanopoulos, G. 2009. Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes. *Appl. Environ. Microbiol.* **75**, 2705-2711.
10. Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547-1549.
11. Lee, D. G. and Lee, S. H. 2015. Investigation of conservative genes in 711 prokaryotes. *J. Life Sci.* **25**, 1007-1013.
12. Lee, D. G. and Lee, S. H. 2020. Investigation of conservative genes in 168 archaeobacterial strains. *J. Life Sci.* **30**, 813-818.
13. Lee, D. G. and Lee, S. H. 2021. Investigation of COGs (Clusters of Orthologous Groups of proteins) in 1,309 species of prokaryotes. *J. Life Sci.* **31**, 834-839.
14. Lee, D. G., Kim, J. H. and Lee, S. H. 2022. Metabolic pathways of 1309 prokaryotic species in relation to COGs. *J. Life Sci.* **32**, 249-255.
15. Liu, Y. and Li, M. 2022. The unstable evolutionary position of Korarchaeota and its relationship with other TACK and Asgard archaea. *mLife* **1**, 218-222.
16. Ren, M. and Wang, J. 2022. Phylogenetic divergence and adaptation of Nitrososphaeria across lake depths and freshwater ecosystems. *ISME J.* **16**, 1491-1501.
17. St John, E., Liu, Y., Podar, M., Stott, M. B., Meneghin, J., Chen, Z., Lagutin, K., Mitchell, K. and Reysenbach, A. L. 2019. A new symbiotic nanoarchaeote (*Candidatus Nanoclepta minutus*) and its host (*Zestospaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst. Appl. Microbiol.* **42**, 94-106.

초록 : 고세균 122종의 보존적 COG pathways와 유전자

이동근 · 이상현*

(신라대학교 제약공학과)

이 연구의 목적은 122종의 고세균 중에 보존된 대사 경로와 보존된 유전자를 확인하는 것이었다. 각각의 122개 고세균이 63개의 COG 대사 경로, 이를 구성하는 822개의 COG, 총 4,877개의 COG를 보유하고 있는지 분석했다. 대사경로에서는 archaeal ribosomal proteins만이 가장 보존적이었다. 122종의 고세균 모두에 공통적인 COG는 7개의 COG pathways에서 46개, 그리고 그 외가 20개였다. COG pathways에서는 ribosome을 구성하는 29개, tRNA synthetase와 전사인자가 각각 5개, RNA polymerase를 구성하는 3개, 그리고 tRNA modification에 관련된 2개의 COG가 공통적이었다. COG pathways에 속하지 않고 122종의 고세균에 공통적인 보존적 유전자까지 고려하면 외부와 세포질을 구분 짓는 세포벽과 세포외기질의 합성, 복제, 전사, 번역, 단백질 대사에 관련된 유전자들 중에서 일부가 공통적이었다. 계통수에서 구한 각 고세균의 distance value를 분류단위로 보면 Euryarchaeota 문의 Halobacteria강의 평균이 가장 낮았고 표준편차는 Thaumarchaeota 문의 Nitrososphaeria강, 강을 알 수 없는 Thaumarchaeota문의 고세균, Euryarchaeota 문의 Halobacteria 강, Crenarchaeota 문의 Thermoprotei 강, 기타 고세균(OA)이 높았다. 계통수 분석으로 6가지의 공통점을 찾았다. 본 연구결과는 보존된 유전자에 관한 자료 외에도 의약품 개발, 균주 개선을 위한 유전자의 선택 등에 활용될 수 있을 것이다.