

Investigation of Conservative Genes in 168 Archaeobacterial Strains

Dong-Geun Lee and Sang-Hyeon Lee*

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

Received May 6, 2020 / Revised June 4, 2020 / Accepted June 5, 2020

The archaeal clusters of orthologous genes (arCOG) algorithm, which identifies common genes among archaeobacterial genomes, was used to identify conservative genes among 168 archaeobacterial strains. The numbers of conserved orthologs were 14, 10, 9, and 8 arCOGs in 168, 167, 166, and 165 strains, respectively. Among 41 conserved arCOGs, 13 were related to function J (translation, ribosomal structure, and biogenesis), and 10 were related to function L (replication, recombination, and repair). Among the 14 conserved arCOGs in all 168 strains, 6 arCOGs of tRNA synthetase comprised the highest proportion. Of the remaining 8 arCOGs, 2 are involved in reactions with ribosomes, 2 for tRNA synthesis, 2 for DNA replication, and 2 for transcription. These results showed the importance of protein expression in archaea. For the classes or orders having 3 or more members, genomic analysis was performed by averaging the distance values of the conservative arCOGs. Classes Archaeoglobi and Thermoplasmata of the phylum Euryarchaeota showed the lowest and the highest average of distance value, respectively. This study can provide data necessary for basic scientific research and the development of antibacterial agents and tumor control.

Key words : Archaeal genome, arCOG (archaeal cluster of orthologous groups of proteins), conservative gene, ortholog

서 론

생명체가 지구에 출현한 후 생물들은 변화하는 지구환경에서 살아가면서 유전자들을 획득, 변이, 소실하였다. 각 생물 종들의 공통 조상이 갖고 있던 공통조상유전자(ancestral gene)는 종 분화(speciation) 및 복사(duplication)를 통하여, 일부 아미노산 서열의 변화를 겪으며 각 생물 종의 유전체에 분포하였을 것이다. 따라서 현재 지구상에 존재하는 생명체를 이해하려면 모든 생명체에 공통적인 유전자와 함께 각 분류 단위에 독특한 유전자의 이해가 필요하다[3, 12].

공통조상유전자에서 종 분화로 다른 생물 종에 분포한 유전자들의 집합을 orthologs라고 하며, 동일 ortholog를 구성하는 단백질들은 서열의 유사성과 동일한 기능을 갖는다[3, 16]. COG (Cluster of Orthologous Groups of protein)는 동일 ortholog에서 발견되는 단백질들의 집합으로, 각 COG는 하나의 공통조상유전자 유래로 3가지 이상의 생물에 분포한다[3]. Galperin 등[3]에 의하면 COG 기법을 이용하여 게놈서열에서 유전자의 파악, 연관된 생물 종에서 ortholog의 파악, 대사경로의 파악과 유전체의 비교 등이 가능하다. Ortholog 관련 연구는 COG

에서 출발하여 강 등[4]은 43 종류의 미생물에서 보존적 유전자를, 이 등[10]은 66개의 미생물의 보존적 유전자를, 이와 이 [9]는 711 종류의 원핵생물들의 보존적 유전자를 보고하였다. 한편 자료의 수가 늘어남에 따라 컴퓨팅 파워의 한계 등으로 고세균의 arCOGs [16] 및 phage의 POGs (phage orthologous groups) [8] 등을 따로 분리하였다.

arCOGs 자료는 2007년에 41 균주의 고세균 유전체에서 7,538개의 arCOGs를 설정한 것을 시작으로[11] 2012년에 120 균주의 고세균 유전체에 존재하는 총 10,335개의 arCOG로 개편되었고[16], 2015년에는 168 균주의 미생물 유전체에 포함된 13,443개의 arCOG로 분류하였다[12].

2019년 10월에는 기존 arCOG 데이터베이스의 업데이트가 진행되었다[2]. 고세균이 진정세균보다 지구에서의 역사가 오래된 것으로 알려져 있으므로 이들 고세균에 보존적인 유전자를 파악하면 지구상 생명체의 이해를 도울 수 있을 것이다. 이 논문에서는 2019년 10월까지의 168 균주의 고세균 유전체에 존재하는 보존적 유전자들의 종류와 기능 그리고 보존성의 정도를 파악하고자 하였다.

재료 및 방법

재료

고세균 유전체들의 유전자 유사성을 파악하기 위해 arCOGs에서 정리된 자료를 이용하였다[2]. 각 고세균이 가진 arCOG 자료를 확보한 후에, 168개의 고세균 모두에 공통적으로 존재하는 arCOG의 종류 등도 파악하였다. 2019년 10월까지의 168

*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.

균주의 고세균 유전체에 포함된 총 1,048,576개의 유전자들을 12,769개의 arCOG 그룹으로 분류해 놓았다[2]. Table 1에 분석한 168 균주의 고세균들의 분류학적 위치와 각 분류군에 해당하는 고세균들의 개수 등을 나타내었다. Table 1에서 Diaforarchaea group과 unclassified로 표시된 것은 아직 강(class)이나 목(order) 분류체계의 확립이 안된 고세균들이다.

아미노산 서열 분석

각 arCOG에 속하는 단백질들의 아미노산 서열 분석은 이와 이[9]의 방법을 따랐다. 분석대상 고세균 모두에 존재하는 보존적 arCOG에 속하는 단백질들의 서열을 NCBI 공개 데이터베이스에서 추출하고, 각 보존적 arCOG에 속하는 단백질들

은 ClustalX (ver. 2.1) 프로그램으로 다중서열비교를 수행한 후 bootstrap NJ method (n=1,000)를 통해 distance value를 담고 있는 '*.phb' 파일을 생성하였다[6]. 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 고세균에 공통적인 arCOG가(Table 2) 나타내는 distance value의 평균과 표준편차를 각 고세균에 대하여 구하였다. 각 고세균이 나타내는 평균과 표준편차를 강(class) 혹은 문(order) 수준에서 계산하였다.

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

Phylum	Class or order	# of organisms	Abbreviations
Crenarchaeota	Thermoprotei	50	Ctp
	Archaeoglobi	5	Eag
Euryarchaeota	Diaforarchaea group	2	Edg
	Halobacteria	27	Ehb
	Methanobacteria	12	Emb
	Methanococci	15	Emc
	Methanomicrobia	22	Emm
	Methanopyri	1	Emp
	Thermococci	16	Etc
	Thermoplasmata	7	Etp
	Thaumarchaeota	Cenarchaeales (order)	1
Nanoarchaeales (order)		2	Tna
Nitrosopumilales (order)		4	Tnp
Nitrososphaeria		1	Tns
Unclassified		1	Tut
Candidatus Korarchaeota	Unclassified	1	CaK

Table 2. The number of arCOGs of common and their functional categories among 168 archaebacterial genomes

Functional category	The number of archaebacterial genomes			
	168	167	166	165
J	00033, 00357, 00402, 00404, 00405, 00406, 00410, 00412	00109, 00401, 00403, 00487	00078	00047, 00397
L	00470, 00488	00280, 00459, 00469	00415, 00439, 00551	
K	00675	00579		
R	00358, 00501, 00543	00035, 00187		
D			00042	
F			00067, 00689	00063, 00603
H			00069	00638
C			00570	
G				00014, 00018, 00496

Prefix arCOG was omitted at each 5-digit number. Functional categories are J (Translation, ribosomal structure and biogenesis), L (Replication, recombination and repair), K (Transcription), R (General function prediction only), D (Cell cycle control, cell division, chromosome partitioning), F (Nucleotide transport and metabolism), H (Coenzyme transport and metabolism), C (Energy production and conversion) and G (Carbohydrate transport and metabolism).

결과 및 고찰

보존적 유전자의 분포

고세균 168 균주에 있는 각 arCOG를 구성하는 단백질의 수는 arCOG00194 (ABC-type multidrug transport system)가 1,167개로 최대였고, 단백질의 수가 3개인 arCOG는 arCOG00055 (Spermidine synthase), arCOG00484 (4-hydroxybenzoate polyprenyltransferase), arCOG00597 (Nitrogenase molybdenum-iron protein), arCOG00714 (PIN domain containing protein), arCOG00750 (ABC-type permease component) 등 5개였다. 각 arCOG는 동일한 공통조상유전자 유래로 3가지 이상의 고세균에 분포하는 단백질들의 집합이다[3, 11].

168개 고세균의 보존적 유전자를 파악한 후 168, 167 균주 등의 고세균에 보존적인 유전자를 Table 2에 나타내었다. 분석대상 168 균주 모두에서 공통적인 arCOG는 14개였다. 기능그룹으로 분류하면 J (Translation, ribosomal structure and biogenesis)에 8개의 arCOG로 가장 많았고, R (General function prediction only)에 3개, L (Replication, recombination and repair)에 2개, K (Transcription)에 1개로 나타났다. 번역이나 리보솜 구성에 관계된 단백질이 8개로 최대였다. 167 균주에 10개, 166 균주에 9개, 165 균주에 8개의 arCOG가 보존적이었다. 165~168 균주에 보존적인 41개의 arCOG 중에서 번역, 리보솜 구성물과 생성에 관련된 arCOG가 13개로 가장 많았고, 다음으로는 DNA의 복제와 재조합 및 수복 관련된 arCOG가 10개였다.

분석대상 계층의 수와 보존적 유전자를 비교하면 43 종류의 미생물 계층에서 72개의 보존적 유전자[4], 66 종류의 미생물 계층에서 62개의 보존적 유전자[10], 711 종류의 원핵생물 계층에서 1개의 보존적 유전자가 존재하는데[9], 본 연구에서는 168 종류의 고세균 계층에서 14개의 보존적 유전자들이 파악되었다. 이는 분석대상 생물 수와 보존적 유전자 수가 반비례한다는 보고[10]와 일치한다.

이와 이[9]는 711 종류의 원핵생물 중 700 종류 이상의 원핵생물에서 보존적인 총 58개의 COG들을 구하였는데, 본 연구에서는 168 종류의 고세균 중 165 종류 이상이 고세균에서 보존적인 41개의 arCOG들을 구하였다.

가장 적은 종류의 arCOG를 가지는 고세균은 Nanoarchaeales 목의 *Nanoarchaeum equitans* Kin4로 69개의 arCOG에 속하는 73개의 단백질을 가졌고, 가장 많은 종류의 arCOG를 가지는 고세균은 Halobacteria 강의 *Haloarcula marismortui*로 330개의 arCOG에 속하는 615개의 단백질을 가졌다. 한편, Halobacteria 강의 *Haloterrigena turkmenica* DSM 5511에서는 329개의 arCOG에 속하는 785개의 단백질을 가졌다. arCOG의 수보다 단백질의 수가 많은 것은 일부 arCOG에 속하는 단백질이 여러 개 있다는 것을 의미하므로, 해당 고세균에서 중복을 나타내는 arCOG의 기능이 강하거나 다양하다고 할

수 있을 것이다.

168 종류의 고세균에 모두 공통적인 arCOG 별로 구성하는 단백질의 수를 보면 각 고세균당 168~266개로 나타났다. 각 고세균당 하나의 단백질을 갖는 arCOG는 arCOG00675 (DNA-directed RNA polymerase, subunit E'/Rpb7)뿐이었고, arCOG00033 (Wybutosine biosynthesis enzyme Trm5)이 266 단백질로 가장 많았다.

167 종류의 고세균에 보존적인 arCOG는 arCOG00109, arCOG00401, arCOG00403 등 10개였다(Table 2). 168 종류의 고세균 중에서 각 arCOG가 없는 고세균은 arCOG00109 (Methylase of polypeptide chain release factors)의 경우는 Euryarchaeota 문(phylum)의 Thermococci 강에 속하는 *Pyrococcus* sp. NA2이었고, arCOG00401 (Threonyl-tRNA synthetase)의 경우는 Halobacteria 강에 속하는 *Salinarchaeum laminariae* Harcht Bsk1이었고, arCOG00403 (Seryl-tRNA synthetase)의 경우는 Nitrosopumilales 목인 *Candidatus Nitrosopumilus koreensis* AR1이었다.

보존적 유전자의 기능

분석대상 168개의 고세균 모두에서 발견되는 14개의 보존적 arCOG들의 기능을 Table 3에 나타내었다. 이들을 분류하면 tRNA synthetase가 6개로 최다이고, 리보솜과 관련된 arCOG00357과 arCOG00358도 있었다. arCOG00033은 wybutosine 합성 관련 효소이다. Wybutosine은 고세균과 진핵생물의 phe-tRNA에서 발견되는 변형된 염기로, de Crécy-Lagard 등[1]은 wybutosine이나 wyosine의 합성에 관련된 5개의 효소 중 4개가 62개의 고세균에 보존적이라고 보고하였는데, 본 연구에서는 168개의 고세균에서 1개의 효소만이 보존적이었다. arCOG00501은 Ribonuclease Z로 tRNA 전구체의 3' 말단에 작용하여 tRNA 성숙에 관계가 있다[14].

DNA 복제와 관련된 arCOG는 arCOG00470과 arCOG00488 등 2개였고, clamp loader는 DNA 복제를 촉진하는 ATPase로 오랑체(pentamer)이다[5]. 전사와 관련된 arCOG00675와 arCOG00543까지 고려하면, 전체 14개의 보존적 arCOG 중 번역 관련이 10개였으며, 복제와 전사 관련이 각 2개로 번역을 통한 단백질의 발현이 고세균에서의 중요성을 알 수 있었다. arCOG00358, arCOG00501, arCOG00543은 기능그룹이 R (General function prediction only)로 분류되어 있지만(Table 2), Table 3처럼 각 단백질의 세부 기능을 파악하여 전사와 관련된 것으로 간주하였다.

이 등[10]은 진핵미생물 3종을 포함한 66개 종류의 미생물에서 공통적인 62개의 COG 중 82.5%의 COG가 번역과 관련이 있고, 이와 이[9]는 700 종류 이상의 원핵생물에 공통적인 58개 COG들의 86.2%가 번역관련임을 보고하였는데, 본 연구에서는 71.43%의 arCOG가 번역과 관련이 있었다. 이와 이[9]의 연구 대상은 고세균 83 균주를 포함한 711개의 원핵생물이

Table 3. Conservative arCOGs found in all 168 archaeobacterial genomes

# of arCOG	Function of protein
arCOG00033	Wybutosine (yW) biosynthesis enzyme Trm5
arCOG00357	Ribosome-binding ATPase YchF, GTP1/OBG family
arCOG00358	Ribosome-interacting GTPase 1
arCOG00402	Prolyl-tRNA synthetase
arCOG00404	Histidyl-tRNA synthetase
arCOG00405	Glycyl-tRNA synthetase (class II)
arCOG00406	Aspartyl/asparaginyl-tRNA synthetase
arCOG00410	Phenylalanyl-tRNA synthetase alpha subunit
arCOG00412	Phenylalanyl-tRNA synthetase beta subunit
arCOG00470	Clamp loader ATPase, large subunit
arCOG00488	DNA polymerase sliding clamp subunit (PCNA homolog)
arCOG00501	Ribonuclease Z, beta-lactamase superfamily hydrolase
arCOG00543	Transcription termination factor, consists of metallo-beta-lactamase domain and an RNA-binding KH domain
arCOG00675	DNA-directed RNA polymerase, subunit E'/Rpb7

었고, 본 연구에서는 168 종류의 고세균만을 대상으로 하였는데, 모두 번역과 관련된 유전자의 보존 비율이 높은 것으로 나타났다. 이 등[10]은 물질대사를 담당하는 효소같은 단백질들의 합성에 연관된 유전자들의 보존 비율이 높아, 이들은 원시생명체부터 물질대사를 주로 수행하는 것으로 추론하였다.

이 등[10]과 이와 이[9]의 연구에서는 보존적 유전자 중 ribosomal subunit를 구성하는 COG들이 30개 정도로 많았는데, 본 연구에서는 이들의 보존 비율은 낮아 Table 2의 arCOG 00780 (Ribosomal protein L18E) 1개뿐이었다. Ribosome-interacting GTPase들은 리보솜 생성, 세포의 성장 조절, GTP/GDP 비율의 감지에 관여하며 50S subunit의 구조변화에 관여하는데[15, 17], arCOG00358은 ribosome-interacting GTPase 이다(Table 3). ATPase YchF/Ola 1은 DNA 수복, 종양 생성, 세포자살과 관련이 있고 인간에서도 발견된다[7]. arCOG00357은 ATPase YchF에 해당하므로 향후 중앙 퇴치 등의 연구에 단서로 활용될 수 있을 것이다.

arCOG00501은 beta-lactamase superfamily에 속하는데 Reddy 등[13]은 *Bacillus subtilis*에 totarol을 처리하여 ribosomal subunit 등의 단백질 발현 변화와 함께 항균작용을 확인하였는데, 이는 COG나 arCOG 자료로 새로운 항균제를 찾는 데 활용이 가능할 것으로 판단된다.

유전체의 보존성

각 arCOG의 구성원들의 서열 변이로 파악되는 distance value의 합이 낮으면, 고세균들 사이의 아미노산 서열 차이가 작다는 것으로 보존성이 높다고 간주한다[4, 9, 10]. 분석대상 14개의 arCOG 중에서 보존성이 가장 높은 arCOG는 ribosome-binding ATPase (arCOG00357)였고, 보존성이 가장 낮은 arCOG는 DNA polymerase sliding clamp (arCOG00488)이었다.

각각의 고세균 유전체가 나타내는 보존성을 분석하기 위하

여, 14개의 보존적 유전자들이 나타내는 distance value를 Table 1에 제시된 분류단위별로 평균과 표준편차를 구하였다. 16S rDNA는 분류에서의 유용성이 높지만, 분석대상 유전자 수가 많으면 한 개나 몇 개의 유전자로 파악하지 못하는 유전체의 특성을 파악할 수 있다[4, 6, 10]. 본 연구에서는 14개 arCOG의 보존성으로 고세균 유전체의 보존성을 분석하였다. 2015년에 arCOG의 확장 보고가 있었고[12] COG에 대한 최신 보고는 2019년에 있었다[3]. 2019년에 arCOG의 수정이 있었지만, 아직 168개의 각 고세균이 가진 arCOG들의 아미노산 서열은 구할 수 없는 상태이다. 이에 Fig. 1은 168 종류의 고세균이 아닌 120 종류의 고세균의(Crenarchaeota 32개, Euryarchaeota 82개, Thaumarchaeota 5개, 기타 1개) distance value를 분류단위별로 분석한 결과이다. 각 고세균별 distance value의 범위는 0.2841~0.3605이었고 평균은 0.3231이었다.

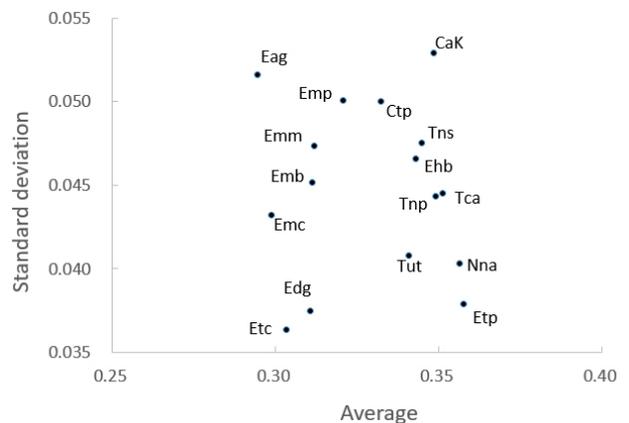


Fig. 1. Distribution pattern of archaeobacterial genomes by distance value with conserved orthologs. X-axis represents distribution of distance averages and Y-axis shows the variance of distance averages for each class or order (Table 1).

Distance value의 평균이 가장 낮아 보존성이 가장 높은 고세균은 Euryarchaeota 문의 Methanococci 강에 속하는 *Methanoterris igneus* Kol5였고, 보존성이 가장 낮은 고세균은 Euryarchaeota 문의 Thermoplasmata 강에 속하는 *Ferroplasma acidarmanus fer1*이었다.

구성원의 수가 적으면 평균이나 편차의 오류 가능성이 클 수 있다. Table 1에서 3개 이상의 구성원을 포함하는 것을 문 (phylum) 수준에서 파악하면 distance value의 평균은 Thaumarchaeota (0.3470), Crenarchaeota (0.3327), Euryarchaeota (0.3179)로 구성원의 개수가 가장 많은 Euryarchaeota의 보존성이 가장 높았다. 하지만 표준편차는 Thaumarchaeota (0.0445)가 Crenarchaeota (0.0500)나 Euryarchaeota (0.0484)에 비해 낮았다. Thaumarchaeota 구성원들은 distance value의 평균은 높고 표준편차는 낮아 보존성은 낮지만 진화단계에서 외부 유전자 유입이 적었다고 할 수 있었다[10]. 그리고, 3개 이상의 구성원을 포함하는 강이나 목 수준의 분류 단위 7개(Ctp, Eag, Ehb, Emm, Emb, Emc, Etp)로 보면 distance value의 범위는 0.2944~0.3575이었고 평균은 0.3190이었다. 7개의 강이나 목 분류단위에서의 변이 평균은 Archaeoglobi 강(Fig. 1의 Eag)에서 최저였고 Thermoplasmata 강(Etp)에서 최고였다. 하지만, 표준편차는 Archaeoglobi 강(Eag)에서 최대였고 Thermoplasmata 강(Etp)에서는 하위 2위였다. 수평적 유전자 전달이 일어나면 하나의 유전체에 존재하는 각 COG들의 distance value는 큰 편차를 보일 수 있다[10]. 따라서 Thermoplasmata 강(Etp)의 보존적 arCOG들이 다른 분류 그룹의 보존적 arCOG들에 비해 아미노산 서열의 변이 정도가 높지만, Etp의 각 arCOG 사이의 변이의 정도는 크지 않으므로 수평적 유전자 전달에 의한 영향없이 공통조상에서 유래된 것으로 유추할 수 있었다.

유전체의 보존성을 나타내는 Fig. 1에서 보이는 특징은 첫째 Euryarchaeota 문의 Thermoplasmata 강(Etp)과 Halobacteria 강(Ehb)을 제외하고는 모두 0.33 이하의 distance value의 평균값을 보였다. 이와 이[9]는 문이나 강 수준의 분류 그룹을 이용하여 고세균이 나타내는 변이의 평균값이 진정세균의 변이의 평균값보다 높다고 보고하였는데, 본 연구에서는 강이나 목 수준의 분류 그룹을 이용하였고 변이의 평균값이 낮은 것으로 유추되었다. 두 번째 특징은 각 분류 단위를 구성하는 유전체의 수와 유전체 변이의 정도는 비례하지 않았는데, 이것은 이와 이[9]의 연구와 일치하였다. 각각 5개와 4개의 유전체로 구성된 Archaeoglobi 강(Eag)과 Thermoplasmata 강(Etp)의 표준편차가 상위와 하위에 있는데, 31개의 유전체로 구성된 Crenarchaeota 문의 Thermoprotei 강(Ctp)과 19개의 유전체로 구성된 Methanomicrobia 강(Emm) 및 21개의 유전체로 구성된 Halobacteria 강(Ehb) 등은 중위에 있었다. 이것이 수평적 유전자 전달없이 공통조상에서 가지고 있던 유전체의 변이에 의한 결과인지, 하나의 강에서 여러 목의 분류 단위

로 나뉘지 않아 나타난 결과인 지에 대해 추후 연구가 필요할 것으로 판단되었다. 이와 이[9]는 Firmicutes와 Proteobacteria 문에 속하는 여러 강이 분포에서 서로 인접하지 않는다는 결과를 보고하였다.

본 연구에서 도출된 14개의 보존적 유전자들은 메탄 생성, 고온성, 호염성 등의 다양한 환경에서 살아가더라도 고세균의 생명현상에 공통적으로 필요한 것으로 판단된다. 이와 이[9]는 보존적 유전자들을 첫째, 원시 생명체의 종 분화(speciation) 이전부터 유지되었거나 둘째, 환경 변화에 따라 유전자를 추가하는 방법 등으로 적용하였거나 셋째, ortholog가 아닌 유전자에 의해 기능이 대체되는 유전자의 기능대체현상(gene displacement)에 의한 것으로 예측하여 보고하면서, 보존적 유전자는 현재 지구의 환경에 필요한 유전자로 생각하였다. 본 연구는 고세균들의 진화 과정에서 보존된 유전자들의 기능으로 기초 생물학 연구에 필요한 자료 제공과 함께 항균제[13] 및 종양[7] 관련 연구에도 활용할 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

- de Crécy-Lagard, V., Brochier-Armanet, C., Urbonavicius, J., Fernandez, B., Phillips, G., Lyons, B., Noma, A., Alvarez, S., Droogmans, L., Armengaud, J. and Grosjean, H. 2010. Biosynthesis of wyosine derivatives in tRNA: an ancient and highly diverse pathway in Archaea. *Mol. Biol. Evol.* **27**, 2062-2077.
- <ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG>
- Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I. and Koonin, E. V. 2019. Microbial genome analysis: the COG approach. *Brief Bioinform.* **20**, 1063-1070.
- 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.
- Kelch, B. A., Makino, D. L., O'Donnell, M. and Kuriyan, J. 2012. Clamp loader ATPases and the evolution of DNA replication machinery. *BMC Biol.* **10**, 34.
- Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press
- Koller-Eichhorn, R., Marquardt, T., Gail, R., Wittinghofer, A., Kostrewa, D., Kutay, U. and Kambach, C. 2007. Human OLA1 defines an ATPase subfamily in the Olg family of GTP-binding proteins. *J. Biol. Chem.* **282**, 19928-19937.
- 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.
9. Lee, D. G. and Lee, S. H. 2015. Investigation of conservative genes in 711 prokaryotes. *J. Life Sci.* **25**, 1007-1013.
 10. 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.
 11. Makarova, K. S., Sorokin, A. V., Novichkov, P. S., Wolf, Y. I. and Koonin, E. V. 2007. Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol. Direct.* **2**, 33.
 12. Makarova, K. S., Wolf, Y. I. and Koonin, E. V. 2015. Archaeal clusters of orthologous genes (arCOGs): An update and application for analysis of shared features between Thermococcales, Methanococcales, and Methanobacteriales. *Life* **5**, 818-840.
 13. 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.
 14. Schiffer, S., Rösch, S. and Marchfelder, A. 2002. Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. *EMBO J.* **21**, 2769-2777.
 15. Shajani, Z., Sykes, M. T. and Williamson, J. R. 2011. Assembly of bacterial ribosomes. *Ann. Rev. Biochem.* **80**, 501-526.
 16. 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.
 17. Zhang, X., Yan, K., Zhang, Y., Li, N., Ma, C., Li, Z., Zhang, Y., Feng, B., Liu, J., Sun, Y., Xu, Y., Lei, J. and Gao, N. 2014. Structural insights into the function of a unique tandem GTPase EngA in bacterial ribosome assembly. *Nucleic Acids Res.* **42**, 13430-13439.

초록 : 168개 고세균 균주들의 보존적 유전자에 관한 연구

이동근 · 이상현*

(신라대학교 의생명과학대학 제약공학과)

고세균 유전체들 사이의 공통적 유전자를 파악하는 archaeal clusters of orthologous genes (arCOG) 알고리즘으로, 168 균주의 고세균들에 공통적인 보존적 유전자를 파악하고자 하였다. 보존된 ortholog의 수는 168, 167, 166 및 165 균주에서 각각 14, 10, 9 및 8개였다. 이들 41개의 arCOG 중에서 번역, 리보솜 구성 및 생합성에 관련된 arCOG가 13개로 가장 많았고 DNA의 복제와 재조합 및 수복에 관련된 arCOG가 10개로 다음으로 많았다. 168개의 고세균 균주들에 보존적인 14개의 arCOG들은 tRNA synthetase가 6개로 가장 많았다. 리보솜과 반응, tRNA 합성, DNA 복제, 전사와 관련한 arCOG가 각 2개씩으로 고세균에서 단백질 발현의 중요성을 알 수 있었다. 보존적 arCOG 구성원들의 distance value의 평균으로 3개 이상의 구성원을 가지는 강(class)과 목(order) 수준에서 유전체를 분석한 결과, Euryarchaeota 문의 Archaeoglobi 강과 Thermoplasmata 강이 각각 최저치와 최고치를 나타내었다. 본 연구는 기초과학 연구와 함께 항균제 개발 및 종양 제어 등에 필요한 자료를 제공할 수 있을 것이다.