ISSN (Print) 1225-9918
ISSN (Online) 2287-3406
Journal of Life Science **2017** Vol. 27. No. 6. 694~701 DOI : https://doi.org/10.5352/JLS.2017.27.6.694

- Note -

Conservative Genes of Less Orthologous Prokaryotes

Dong-Geun Lee*

Major in Pharmaceutical Engineering, Division of Bio-industry, College of Medical and Life Science, Silla University, Kwaebop-dong 1-1, Busan 617-736, Korea

Received January 10, 2017 / Revised February 6, 2017 / Accepted February 24, 2017

Mycoplasma genitalium represents the smallest genome among mono-cultivable prokaryotes. To discover and compare the orthologs (conservative genes) among *M. genitalium* and 14 prokaryotes that are uncultivable and have less orthologs than *M. genitalium,* COG (clusters of orthologous groups of protein) analyses were applied. The analyzed prokaryotes were *M. genitalium*, one hyperthermophilic exosymbiotic archaeon *Nanoarchaeum equitans*, four intracellular plant pathogenic eubacteria of *Candidatus* Phytoplasma genus, and nine endosymbiotic eubacteria of phloem- and xylem-feeding insects. Among 367 orthologs of *M. genitalium,* 284 orthologs were conservative between *M. genitalium* and at least one other prokaryote. All 15 prokaryotes commonly have 29 orthologs, representing the significance of proteins in life. They belong to 25 translation-related, including 22 ribosomal proteins, 3 subunits of RNA polymerase, and 1 protein-folding‐related. Among the 15 prokaryotes, 40 orthologs were only found in all four *Candidatus* Phytoplasma. The other nine *Candidatus,* all endosymbionts with insects, showed only a single common COG0539 (ribosomal protein S1), representing the diversity of orthologs among them. These results might provide clues to understand conservative genes in uncultivable prokaryotes, and may be helpful in industrial areas, such as handling prokaryotes producing amino acids and antibiotics, and as precursors of organic synthesis.

Key words : *Candidatus* Phytoplasma, conservative gene, COG (cluster of orthologous groups of proteins), *Mycoplasma genitalium,* orthologs

서 론

지금까지 보고된 원핵생물은 분포와 특징 등이 아주 다양하며, 이들은 기초적 생명현상의 파악 및 생태와 경제적 측면 등에서 매우 중요하다[2]. 이들의 생명현상에는 유전자가 필수적이며, 유전자가 많다는 것은 많은 효소 등 다양한 단백질 등의 생산이 가능하며 다양한 환경에서 생존할 수 있다는 것을 나타낸다[5]. 게놈크기와 유전자의 개수는 비례하며 일반적으로 원핵생물의 유전자는 1 kilo-base pair (Kbp) 당 하나씩 존재한다[16]. 현재까지 보고된 원핵생물의 유전체 크기는 *Candidatus* Nasuia deltocephalinicola str. NAS-ALF의 최소 0.11 mega-base pair (Mbp) [7]에서 *Sorangium cellulosum*의 최대 14.8 Mbp로[5] 최소와 최대가 100배 이상 차이를 보인다.

한편 생존에 다른 생물을 필요로 하는 원핵생물들이 있다. 숙주의존성 원핵생물들은 생명활동에 필요한 부분 중 일부를 숙주에 의존하여, 자유생활 원핵생물에 비해 유전자의 수가 작아도 현재까지 생존할 수 있었을 것이다. 실제 일반적으로 자유생활을 하는 원핵생물에 비해 공생, 기생, 병원성 등 숙주의존성 원핵생물들은 게놈크기가 작다[14, 16].

하나의 공통조상 유전자에서 기원하여 다양한 생물종 들에 분포하는 보존적 유전자(Conservative gene)들을 orthologs로 정의하고, orthologs에서 유래한 단백질들을 OG (Orthologous Group of proteins)라고 하며 하나의 OG는 구조와 기능이 유사하다[4]. 염색체 서열분석 기술이 발전함에 따라 배양 없이도 원핵생물의 게놈서열, 유전자의 개수와 기능의 유추 등이 가능하다[17]. 3가지 이상의 생물종에 존재하는 OG를 COG (Clusters of Orthologous Group of proteins)라고 한다[4]. COG 기법을 이용하면 분석대상 생물종에 공통적인 보존적 유전자 파악[12], 진화과정에서 유전자의 변화[14], 배양없이 원핵생물의 기능유추[17], 인공 원핵생물의 설계[8], 내성균주 개발[10] 등 여러 방면에 적용할 수 있다.

2017년 1월 현재까지 COG database에 보고된 COG의 개수는 4,631개 이며 원핵생물 균주는 총 711개 이다[4, 12]. 이들 중 *Mycoplasma genitalium* G37은 같은 속의 *M. pneumoniae*에 비해 200개 적은 유전자를 보유하고 있으며[6], 자연계에서 분리하여 인공적으로 단독배양이 가능한 원핵생물들 중 최소 크기의 게놈(0.58 Mbp)과 367개의 COG를 보유하고 있어[3] 많은 연구의 비교대상이 되었다[6, 8]. COG database에 보고된 711종의 균주들 중에서 *M. genitalium* 보다 작은 수의 COG

***Corresponding author**
Tel : +82-51-999-6282, Fax : +82-51-999-5636
E-mail : ldg@silla.ac.kr

를 가진 원핵생물들은 14종이었다[3, 4]. 보유한 COG 즉 orthologs 수가 작은 원핵생물은 다른 원핵생물들과의 보존적 유전자 개수가 작다는 것으로 생명현상의 공통점이 작다고 할 수 있지만, 동시에 이들이 보유한 orthologs들은 각 원핵생물의 생명현상에 필수적이라고 할 수 있을 것이다.

이 논문에서는 COG database에 보고된 711종의 균주들 중에서 *M. genitalium* 보다 작은 수의 orthologs를 가진 원핵생물들이 가지는 보존적 유전자를 파악하고 상호 비교하여 공통점과 차이점을 파악하여 생명현상과의 관계를 알아보고자 하였다.

재료 및 방법

분석대상 원핵생물

분석대상 15개 원핵생물들을 Table 1에 나타내었다. 고세균은 *Nanoarchaeum equitans* Kin4-M (이하 Nanequ)만 있었다. 각 원핵생물이 함유하고 있는 보존적 유전자인 COG 자료는 COG database에서 확보하였다[3]. 유전체의 크기, 총 유전자 개수, 단백질 유전자 개수 등의 자료는 BioCyc Database Collection을 이용하였다[1]. *M. genitalium* (이하 Mycgen)은 세포내기생도 하며 단독시험관배양도 가능하다[12]. Nanequ는 초고온성 고세균으로 세포외공생을 하며, *Candidatus* Phytoplasma 속 4개의 원핵생물은 식물의 세포 내부에 기생하는 병원균이다. 나머지 9개의 원핵생물들은 식물의 수액을 섭취하는 곤충의 세포 내부에 공생한다.

Orthologs 비교

분석대상 원핵생물들이 보유한 orthologs 비교를 위하여 COG database에서[3] 확보한 자료를 이용하였다. 분석대상 생물종이 전체 4,631개의 각 COG를 보유하고 있는 지를 MS사의 엑셀프로그램(ver. 10)을 이용하여 행렬로 작성하였다[12]. 이후 15개의 분석대상 원핵생물 사이에 공통적인 COG 번호를 통해 보존적 유전자를 검색하였고, Mycgen과 공통적인 것과 Mycgen 이외의 분석대상에만 존재하는 COG 등으로 나누어 분석하였다.

Ortholog 보유 계통수(ortholog content tree)

행렬로 작성된 분석대상 원핵생물들이 4,631개의 각 COG를 보유하는 지를 토대로 ortholog 보유 계통수를 작성하였다. Mega 프로그램(ver. 6.06)의 phylogeny analysis를 이용하여 phylogenetic tree를 작성하면서 bootstrap method (n=1,000)로 분석하였다.

결과 및 고찰

게놈과 유전자 수 비교

Table 1에 게놈의 크기를 기준으로 분석 대상 각 원핵생물의 게놈 크기, GC 비율(%), Mycgen과의 공통 COG의 개수와 비율(%) 등을 나타내었다. 분석대상 원핵생물들은 명명법이 *Candidatus*로 시작되는 것이 총 13개로, 이들은 시험관에서 배양이 불가능하다. *Candidatus* Phytoplasma 속(genus)은 형태적으로 mycoplasma와 유사한 원핵생물들을 이르는 용어로 구성원 서로간의 유연관계가 높지 않다[2].

게놈의 크기는 112,091~959,779 염기쌍(base pairs, bp), GC 비율은 13.5~42.1%, Mycgen의 367개 COG와 공통인 COG의 개수는 66~244개, 각 원핵생물이 보유한 전체 단백질 중 COG가 차지하는 비율은 29.36~98.24% 범위였다. Trephe, Uzidia, Sulmue, Syndia, Mycgen, Phyaus 등의 원핵생물은 단백질 유전자 수와 RNA 유전자 수 그리고 전체 유전자 수가 일치하지 않았는데, 이는 가짜 유전자(pseudogene)까지 유전자로 계산하여 나타난 결과이다[1]. 게놈의 크기를 유전자의 총수로 나누면 663~1,210 bp/gene의 범위(평균 954 bp/gene)였고 Nasdel (663 bp/gene), Hodcic (765), Carrud (735), Trephe (805) 등 게놈의 크기가 작을수록 유전자 하나의 평균 bp도 짧았다. 이는 유전자의 크기가 실제로 작거나 염기서열 일부를 다수의 유전자가 공유하는 중복유전자(overlapping gene)의 존재를 알려주는 것으로 판단되었다[11].

GC 비율은 절대세포내 원핵생물에서 감소한다는 보고[13, 14]가 있었는데, 연구대상 게놈들의 GC 비율은 13.5~42.1%로 나타났다. 17% 이하의 GC 비율을 보인 원핵생물은 Nasdel, Carrud, Zinins의 3종이고 30% 이상의 비율은 Trephe (42.1%), Uzidia (30.2%), Nanequ (31.6%), Mycgen (31.7%)의 4종으로, 초고온성 고세균인 Nanequ는 중온성인 Mycgen과 유사하였다. 대장균이 50% 수준의 GC 비율을 보이므로 Trephe와 고온성인 Nanequ는 높지 않다고 할 수 있었다. GC 비율은 환경에의 적응을 반영한다. 즉 자유생활을 하는 원핵생물은 높은 GC 비율을 갖고, 영양분이 제한되거나 부족한 환경에서 생활하는 원핵생물은 낮은 GC 비율을 갖는다[13, 14]. 혐기성에 비해 호기성이, 세포내공생체에 비해 병원균이, 중온성에 비해 고온성세균이 높은 GC 비율은 보이며 평균적으로 자유생활세균(49.1%), 통성 숙주의존성세균(43%), 세포내기생세균(37.8%), 세포내공생세균(26.8%) 순으로 GC 비율이 감소한다는 보고[14]와 비교하면, 초고온성 기생세균인 Nanequ의 31.6%는 낮다고 할 수 있었다.

현재까지 인공적으로 단독배양이 가능하다고 보고된 원핵생물들 중에서 Mycgen은 최소의 게놈크기를 가지고 있고 367개의 COG를 가지고 있으며 비교대상으로 많이 활용된다[6, 8]. Table 1에서 Mycgen과 공통적인 COG의 수를 보면 Mycgen보다 게놈이 큰 4개의 원핵생물들은 240~246개 COG가 (74.62~76.64%), Mycgen보다 게놈이 작은 원핵생물들은 66~180개(27.17~64.08%)의 COG가 Mycgen과 공통이었다. 그리고 각 원핵생물이 보유한 COG 종류와 Mycgen의 COG와의 공통비율도 Mycgen보다 게놈이 큰 4개의 원핵생물들은

Table 1. Lineage and the number of COGs, proteins and genes of analyzed 15 procaryotes

Organism (Abbreviation)	Host	Genome size (bp)	GC ratio (%)	# of COGs		% of common COGs with Mycgen	# of genes for			COG / Proteins (%)
				Total	Common with Mycgen		Proteins	RNA	Total	
Candidatus Nasuia deltocephalinicola str. NAS-ALF (Nasdel)	Leafhopper	112,091	16.9	103	66	64.08	137	32	169	75.18
Candidatus Hodgkinia cicadicola Dsem (Hodcic)	Cicada	143,795	46.4	125	68	54.40	169	19	188	73.96
Candidatus Tremblaya phenacola PAVE (Trephe)	Mealybugs	171,500	42.1	165	93	56.36	175	35	213	94.29
Candidatus Carsonella ruddii DC (Carrud)	Citrus psyllid	174,014	15.1	155	85	54.84	207	31	238	74.88
Candidatus Zinderia insecticola CARI (Zinins)	Spittlebugs	208,564	13.5	186	106	56.99	202	28	230	92.08
Candidatus Uzinura diaspidicola str. ASNER (Uzidia)	Armored scale insects	263,431	30.2	223	125	56.05	227	34	261	98.24
Candidatus Sulcia muelleri CARI (Sulmue)	Leafhopper	276,511	22.7	226	134	59.29	246	33	279	91.87
Candidatus Portiera aleyrodidarum BT-QVLC (Porale)	Whitefly	357,472	26.1	235	117	49.79	252	36	288	93.25
Candidatus Profftella armatura (Syndia)	Citrus psyllid	464,857	24.2	355	180	50.70	372	37	410	95.43
Nanoarchaeum equitans Kin4-M (Nanequ)	*Ignicoccus hospitalis*	490,885	31.6	346	94	27.17	541	47	588	63.96
Mycoplasma genitalium G37 (Mycgen)	-	579,677	31.7	367	367	100.00	475	43	518	77.26
Candidatus Phytoplasma asteris AYWB (Astyel)	Onion	723,970	26.8	316	240	75.95	693	37	730	45.60
Candidatus Phytoplasma asteris OY-M (Oniyel)	Onion	853,092	27.8	327	244	74.62	749	39	803	43.66
Candidatus Phytoplasma australiense (Phyaus)	Strawberry	879,959	27.3	321	246	76.64	684	43	727	46.93
Candidatus Phytoplasma australiense NZSb11 (Strlet)	Strawberry	959,779	27.3	323	245	75.85	1100	41	1141	29.36

The data of COGs and genes were from COGs [4] and BioCyc (http://biocyc.org) database, respectively.
M. genitalium G37 (Mycgen) could be mono cultured *in vitro*. *N. equitans* Kin4-M (Nanequ) is the only archaeon and extracellular symbiotic hyperthermophile and the others are obligate intracellular symbionts or pathogens.

74.62~76.64%, Mycgen보다 게놈이 작은 원핵생물들은 27.17~64.08%의 범위였다. 즉 게놈크기가 작을수록 Mycgen과의 공통 COG의 수와 비율이 작아 생명현상이 Mycgen과 차이를 보일 가능성이 있는 것으로 판단되었다.

rRNA 유전자의 수는 Hodcic의 19개, Zinins의 28개를 제외하면 모두 31~47개 사이였다. rRNA 유전자 수가 많으면 활용이 가능한 자원에 대응하는 속도가 빨라지며, 성장속도가 빠르면 rRNA 유전자 수가 많다고 한다[6]. 실제로 *Escherichia coli* O157:H7은 140개, *Bacillus subtilis* 168은 184개의 rRNA 유전자를 보유하고 있었다[7]. 따라서 분석대상 원핵생물들은 모두 성장속도가 느리며 외부환경의 변화에 빠른 대응이 힘들 것으로 사료되었다. Mycgen을 제외한 분석대상 14개 균주 중 Nanequ를 제외한 나머지 13개 균주는 모두 세포내 병원체 혹은 공생체로 숙주 세포내부의 환경은 세포외부 환경보다 변화가 적을 것이며, rRNA 유전자 개수가 작아도 생존이 가능하였을 것이다.

전체 단백질 중 COG가 차지하는 비율은 *Candidatus* Phytoplasma 속의 원핵생물들이 29.36~49.63%로 비교대상 다른 원핵생물들의 63.96~98.24%보다 낮았다(Table 1). COG는 3종류 이상의 원핵생물에 분포하므로 이들은 비교대상 다른 원핵생물들과 구분되는 독자적인 생명현상을 많이 나타낼 것으로 유추되었다.

보존적 유전자 개수 비교

Table 2에 Mycgen에도 분포하며 나머지 분석대상 14종의 원핵생물들 1~14종에 존재하는 COG의 개수를 나타내었다. 분석대상 15종 모두에 존재하는 COG의 수는 29개 이고, Mycgen에만 존재하는 COG의 수는 83개 였다. 따라서 Mycgen이 보유한 367개의 COG 중에서 Mycgen과 나머지 비교대상 원핵생물들이 공유하는 COG의 종류는 총 284개 였다. 한편 Mycgen에 없고 11종 이상의 원핵생물들이 공통적으로 보유하는 COG는 없었고, Mycgen에 없고 1~10종의 원핵생물이 보유하는 COG는 총 632개 였다(Table 2). 이중 Mycgen에 없고 단 하나의 원핵생물에 존재하는 COG의 수가 382개로 전체 632개중 60.44%로 나타났다. 382개의 COG를 균주별로 보면 Nanequ가 219개, Syndia가 78개였고 나머지는 1~21개 사이였다. 즉 초고온성 고세균인 Nanequ와 polyketide 독소를 합성하여 공생곤충을 보호하는 Syndia가[15] Mycgen과 많은 생리적 차이를 보인다고 할 수 있었다.

분석대상 원핵생물 중 초고온성 고세균인 Nanequ를 제외하고 나머지는 진정세균이었다. Table 2의 괄호 안의 숫자는 Nanequ를 제외한 나머지 원핵생물들을 비교한 결과이다. Mycgen과 나머지 비교대상 원핵생물들 사이의 공통적인 COG는 284개 였다. 이것의 66.9%인 190개 COG를 Nanequ가 보유하지 않았고, 33.1%인 94개 COG만 공통적이었다. Mycgen에 없는 632개 COG의 60.1%인 380개 COG가 Nanequ에 없었다.

Table 2. Number of common COGs with or without *M. genitalium* (Mycgen) and *N. equitans* Kin4-M (Nanequ) among 15 analyzed prokaryotes

# of bacteria	# of common COGs	
	with Mycgen (+ without Nanequ)	without Mycgen (+ without Nanequ)
15	29 (NP)	NP (NP)
14	17 (11)	0 (NP)
13	15 (8)	0 (0)
12	15 (10)	0 (0)
11	11 (9)	0 (0)
10	13 (6)	2 (1)
9	15 (9)	2 (2)
8	17 (12)	7 (6)
7	23 (15)	12 (12)
6	28 (20)	20 (12)
5	68 (66)	41 (36)
4	6 (5)	54 (53)
3	10 (8)	38 (34)
2	17 (11)	74 (61)
1	83 (83)	382 (163)
Total	367 (273)	632 (380)

Values in parenthesis are the number of common COGs without archaeon Nanequ.
M. genitalium G37 (Mycgen) could be mono cultured *in vitro*. *N. equitans* Kin4-M (Nanequ) is the only archaeon and extracellular symbiotic hyperthermophile and the others are obligate intracellular symbionts or pathogens.
NP means not possible to compare.

또한 Mycgen에 없고 단 하나의 원핵생물에 존재하는 COG의 수는 382개 인데 이중 219개(57.3%)가 Nanequ에만 존재하였다. 이로서 비교대상 원핵생물들 중 유일한 고세균인 Nanequ가 다른 원핵생물들에 비해 독특한 단백질과 생명현상을 나타내는 것으로 유추할 수 있었다.

분석대상 모두의 보존적 유전자

Table 3에 분석대상 15종의 원핵생물 게놈 모두에 존재하는 COG의 기능범주와 COG 번호를 나타내었다. 총 29개의 COG가 15종의 원핵생물에 공통적이었다. 관련 COG의 기능과 기능범주(functional category)로 나타내면 가장 개수가 많은 것은 번역관련(기능범주 J)으로 총25개 중 tRNA-synthetase (COG0008), 번역개시인자(COG0361), 번역신장인자(COG0480)가 1개씩이고 22개가 리보솜 구성단백질이었다(Table 2). 전사관련(기능범주 K) COG는 COG0085, COG0086, COG0202로 모두 RNA 중합효소의 소단위체였고, 기능범주 O의 COG0459는 chaperonin GroEL로 단백질의 접힘(folding)에 관여한다. 연구대상 균주들의 공통 COG는 모두 RNA와 리보솜 그리고 단백질 접힘에 관계하여 생명현상에서 단백질이 중요함을 알려주는 것으로 사료되었다. 전사관련(기능범주 K)

COG0202는 RNA 중합효소의 α 소단위체로 전사의 핵심역할을 수행하는데, α 소단위체의 돌연변이는 부탄올 등의 용제에 내성이 강한 세균 등 다양한 표현형을 나타내었다[10]. 동일 COG라도 아미노산서열에 차이를 보이는데 서열이 다른 α 소단위체가 다양한 환경에서 각 원핵생물들의 생존에 도움을 주는 것으로 판단할 수 있었다.

비교대상 원핵생물 711종 모두에 공통적인 COG는 COG 0080 (Ribosomal protein L21) 1개, 708종 이상에서는 22개의 COG와 703종 이상은 49개의 COG가 공통이었다[12]. 일반적으로 비교대상 생물의 수가 증가하면 공통 COG의 수는 감소하는데, 본 연구에서는 분석대상 원핵생물의 수가 15개로 708개에[12] 비해 월등히 작지만 공통 COG의 수는 크게 차이가 나지 않았다. 또한 본 연구에서 파악된 공통적 COG 29개는 모두 703종 이상의 공통적 COG에 포함되어[12], 본 연구의 공통적 COG가 원핵생물의 생명유지에 중요한 것으로 판단되었다.

각 분석대상그룹의 보존적 유전자

Mycgen 이외의 분석대상그룹에 존재하는 보존적 유전자를 비교하였다. 분석대상 15개 원핵생물 중 *Candidatus* Phytoplasma에 속하는 균주는 Astyel, Oniyel, Phyaus, Strlet 등 4개였다(Table 1). 이들은 Mycgen과 함께 Mollicutes 강(class)에 속하며 식물의 세포내부에 기생하는 병원체이다. 분석대상 15개 원핵생물 중 *Candidatus* Phytoplasma 속 4개 균주 모두에서만 존재하는 40개 COG의 기능범주와 번호를 Table 3에 나타내었다. 총 100개의 COG가 분석대상 15개 원핵생물 중 하나 이상의 *Candidatus* Phytoplasma에만 존재하여, *Candidatus* Phytoplasma속 4개 균주 사이에도 보존적 유전자의 다양성이 있는 것으로 판단되었다. 또한 *Candidatus* Phytoplasma에 속하는 구성원 서로간의 유연관계가 높지 않다는 보고[2]도 있었다. Mycgen에는 없고 *Candidatus* Phytoplasma에만 존재하는 기능범주와 COG는 기능범주 Q (Secondary metabolites biosynthesis, transport and catabolism) 관련 COG4869 (Propanediol utilization protein), 기능범주 U (Intracellular trafficking, secretion, and vesicular transport) 관련 COG1272 (Predicted membrane channel-forming protein YqfA, hemolysin III family), 기능범주 X (Mobilome: prophages, transposons) 관련 COG2801 (Transposase InsO and inactivated derivatives), 그리고 복합기능인 기능범주 FV 관련 COG0756 (dUTPase)과 기능범주 ET 관련 COG0834 (ABC-type amino acid transport/signal transduction system, periplasmic component/domain) 등 총 5개 였다. 분석대상 *Candidatus* Phytoplasma들은 식물에 기생하고 Mycgen은 동물에 기생하는데 보유하는 COG에서도 *Candidatus* Phytoplasma들에만 존재하는 COG가 있어, 이들은 Mycgen 및 분석대상 다른 원핵생물들과 구별되는 생명현상을 나타나는 것으로 유추할 수 있었다. 하지만 COG 데이타베이스의 711개 원핵생물 중 COG4869는 69개, COG1272는 308개, COG2801은 461개, COG0756은 492개, COG0834는 531개의 원핵생물에 분포하여[3] 이들 COG가 식물체에의 기생과 관련되는 지는 확신할 수 없었다.

분석대상 15개 원핵생물 중 *Candidatus* Phytoplasma를 제외한 나머지 9개의 *Candidatus*들은 곤충에 공생한다는 공통점 외에 속(genus)은 모두 달랐다(Table 1). 이 곤충에 공생하는 9개의 *Candidatus*들에서만 발견되는 COG는 Ribosomal protein S1인 COG0539 하나뿐이었다. COG0539는 비교대상 나머지 원핵생물에는 없지만 COG 데이타베이스의 711개 원핵생물 중 614개의 원핵생물에 존재하여, COG0539의 분포와 *Candidatus*들의 특징을 설명하기는 어려웠다. 또한 총 291개의 COG가 분석대상 15개 원핵생물 중 하나 이상의 곤충에 공생하는 *Candidatus*들에 있어, 이들 사이에 보존적 유전자의 분포 즉 생리적 현상 등이 다양할 가능성이 있는 것으로 유추되었다.

하지만 이러한 결과는 각 그룹별로 분석대상 원핵생물들의 수가 적다는 한계가 있어, 향후 분석대상 원핵생물들의 수를 확대하면 각 분석대상그룹의 보존적 유전자 파악에 도움이 될 것으로 판단되었다. 또한 속- 혹은 종-특이적으로 존재하며 다른 계보(lineage)에는 존재하지 않는 고아 유전자(Orphan genes)가 특정 생물들의 독특한 생명현상에 관여할 수도 있다[18]. COG는 3가지 이상의 생물들에 분포하므로[4, 12] 속-특이적 COG 등이 존재하는 지 향후 연구를 통해 밝힐 수 있을 것이다.

Ortholog 보유 계통수(ortholog content tree)

Ortholog에서 유래한 COG의 보유유무를 bootstrap (n= 1,000)을 적용하여 작성한 분석대상 원핵생물들의 Maximum-Likelihood (ML) 계통수를 Fig. 1에 나타내었다. Neighbor joining, UPGMA, maximum parsimony 계통수에서도 아주 유사한 형태를 보였다. 결과를 보면 첫째 *Candidatus* Phytoplasma 속 4균주와 곤충공생 원핵생물 9균주가 서로 다른 분류그룹으로 나누어졌다. 둘째 보유 COG의 수가 103, 125, 186개인 Nasdel, Hodcic, Zinins 그리고 165, 155, 223, 226개인 Trephe, Carrud, Uzidia, Sulmue 등이 높은 유연관계를 보여 단순히 COG의 보유개수가 계통수에 반영된 것은 아닌 것으로 유추되었다. 이는 분석대상 원핵생물들이 보유한 ortholog의 분포가 서로 차이를 보이며, 서로 다른 생명현상과 연관이 있을 것으로 판단되었다. 각 분기점에서 생물종까지의 거리를 보면 *Candidatus* Phytoplasma 속에 비해 곤충공생 원핵생물 9균주가 길었다. 이는 *Candidatus* Phytoplasma 속들의 유연관계가 곤충공생 원핵생물들 보다 높은 것을 나타내는 것으로 '각 분석대상그룹의 보존적 유전자'의 결과의 공통 COG의 수와 상통하는 것으로 판단되었다.

Table 3. The function and the number of COGs common in all 15 procaryotes and only in all 4 *Candidatus* Phytoplasma

Function (Functional category)		The number of COG	
		in all analyzed prokaryotes	only in all *Candidatus* Phytoplasma
Energy production and conversion	(C)		0240 0778 3493
Cell cycle control, cell division, chromosome partitioning	(D)		2088 4942
Amino acid transport and metabolism	(E)		0765 1126 1135 2011
Nucleotide transport and metabolism	(F)		0504 1051
Carbohydrate transport and metabolism	(G)		3839
Lipid transport and metabolism	(I)		0688 1183
Translation, ribosomal structure and biogenesis	(J)	0008 0048 0049 0051 0052 0080 0087 0088 0090 0092 0093 0094 0096 0097 0098 0099 0100 0103 0185 0186 0197 0200 0361 0480 0522	0802 3481
Transcription	(K)	0085 0086 0202	1191
Replication, recombination and repair	(L)		0338 0749 1039 1198 1484
Cell wall/membrane/envelope biogenesis	(M)		1970
Posttranslational modification, protein turnover, chaperones	(O)	0459	0265
Inorganic ion transport and metabolism	(P)		0803 1108 1121 1464
Secondary metabolites biosynthesis, transport and catabolism	(Q)		4869
General function prediction only	(R)		0388 1054 2179 4122
Function unknown	(S)		3870 4720
Signal transduction mechanisms	(T)		3887
Intracellular trafficking, secretion, and vesicular transport	(U)		1272
Mobilome: prophages, transposons	(X)		2801
Others			0756(FV) 0834(ET)

Prefix COG was omitted at each number and double letters in others column are mixed functional category.

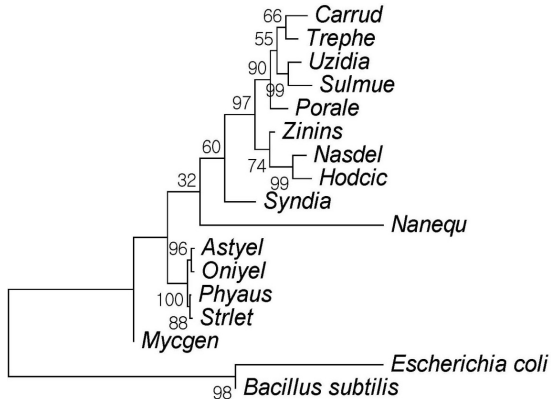

Fig. 1. ML (Maximum Likelihood) phylogenetic tree of 15 prokaryotes and outgroups of *E. coli* K-12 and *B. subtilis* in the point of presence or absence of 4,631 COG. Bootstrap values at each node are expressed as a percentage of 1,000 trials and values lower than 30% were not expressed.

연구결과 활용 가능성

Hutchison III 등[8]은 Mycgen (*M. genitalium*)과 *Haemophilus influenza*의 유전자를 비교하여 배양 가능한 원핵생물 중 최소게놈인 Mycgen도 배양에 비필수적인 유전자가 존재한다고 보고하였다. 또한 438개의 단백질, 35개의 RNA 등 총 473개의 유전자를 가진 531 Kbp의 인공염색체를 합성하고 *M. capricolum* 기반의 인공세균을 개발하여 배지에서 생장과 세포분열이 가능한 것을 보았다[8]. 즉 보존적 유전자를 이용하여 인공의 원핵생물을 합성하였다.

전사공학(transcriptional engineering)과 균주공학(strain engineering)으로 RNA 중합효소의 소단위체에 돌연변이를 유도하여 부탄올 등의 용제에 내성이 강한 균주 생산[10] 외에 항생제 대량생산도 가능한 것으로 보고되고 있다. 균주개량에는 서열기반법과 무작위법의 두 가지가 있는데 본 연구에서 활용한 COG는 기능과 서열에 기반하므로[4, 17] 균주개량의 기초자료가 될 수 있을 것이다. 즉 전사공학, 균주공학, 합성생

물학 등을 통하여 아미노산, 항생제, 의약품, 유기합성 전구체 등을 효율적으로 합성하는 영역 등에 기초자료로 활용이 가능할 것이다.

References

1. Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Pujar, A., Shearer, A. G., Travers, M., Weerasinghe, D., Zhang, P. and Karp, P. D. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*. **40**, D742-D753.
2. Firrao, G., Gibb, K. and Streten, C. 2005. Short taxonomic guide to the genus 'Candidatus Phytoplasma'. *J. Plant Pathol*. **87**, 249-263.
3. ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data
4. 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.
5. Han, K., Li, Z. F., Peng, R., Zhu, L. P., Zhou, T., Wang, L. G., Li, S. G., Zhang, X. B., Hu, W., Wu, Z. H., Qin, N. and Li, Y. Z. 2013. Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Sci. Rep*. **3**, 2101.
6. Himmelreich, R., Plagens, H., Hilbert, H., Reiner, B. and Herrmann, R. 1997. Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res*. **25**, 701-712.
7. http://biocyc.org/organism-summary?object=NEQU228908
8. Hutchison III, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N. and Deerinck, T. J., et al. 2016. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253.
9. Klappenbach, J. A., Dunbar, J. M. and Schmid, T. M. 2000. rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol*. **66**, 1328-1333.
10. 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.
11. Koressaar, T. and Remm, M. 2012. Characterization of species-specific repeats in 613 prokaryotic species. *DNA Res*. **19**, 219-230.
12. Lee, D. G. and Lee, S. H. 2015. Investigation of conservative genes in 711 prokaryotes. *J. Life Sci*. **25**, 1007-1013.
13. Mann, S. and Chen, Y. P. 2010. Bacterial genomic G + C composition-eliciting environmental adaptation (Review). *Genomics* **95**, 7-15.
14. Merhej, V., Royer-Carenzi, M., Pontarotti, P. and Raoult, D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct*. **10**, 13.
15. Nakabachi, A., Ueoka, R., Oshima, K., Teta, R., Mangoni, A., Gurgui, M., Oldham, N. J., van Echten-Deckert, G., Okamura, K., Yamamoto, K., Inoue, H., Ohkuma, M., Hongoh, Y., Miyagishima, S., Hattori, M., Piel, J. and Fukatsu, T. 2013. Defensive bacteriome symbiont with a drastically reduced genome. *Curr. Biol*. **23**, 1478-1484.
16. Ochman, H. and Davalos, L. M. 2006. The nature and dynamics of bacterial genomes. *Science* **311**, 1730-1733.
17. Shoji, S., Dambacher, C. M., Shajani, Z., Williamson, J. R. and Schultz, P. G. 2011. Systematic chromosomal deletion of bacterial ribosomal protein genes. *J. Mol. Biol*. **413**, 751-761.
18. Tautz, D. and Domazet-Lošo, T. 2011. The evolutionary origin of orphan genes. *Nat. Rev. Genet*. **12,** 692-702.

초록 : Orthologs 수가 적은 원핵생물들의 보존적 유전자

이동근*
(신라대학교 의생명과학대학 바이오산업학부 제약공학전공)

알려진 단독배양이 가능한 원핵생물 중 최소게놈을 가지고 있는 *Mycoplasma genitalium*보다 보존적 유전자 수가 적은 14개 원핵생물의 유전자를 보존적 유전자 관점의 COG (Clusters of Orthologous Group of proteins)로 검토하였다. 분석대상은 *M. genitalium*, 초고온성 고세균으로 세포외공생을 하는 *Nanoarchaeum equitans,* 진정세균으로 식물의 세포내에 기생하는 병원균인 *Candidatus* Phytoplasma 속 4개와 식물의 수액을 섭취하는 곤충의 세포내에 공생하는 9종이었다. *M. genitalium*이 가진 367개의 보존적 유전자 중에서, 284개가 비교대상 다른 원핵생물과 공통이었다. *M. genitalium* 등 분석대상 원핵생물 모두에 보존적 유전자는 29개로, 이들은 리보솜 구성단백질 22개 등 번역관련 25개, RNA 중합효소의 소단위체 3개, 단백질 접힘관련 1개 등으로 단백질의 중요성을 알 수 있었다. 분석대상 15개 원핵생물 중 *Candidatus* Phytoplasma속 4개 균주 모두에만 존재하는 COG는 40개 였다. 속(genus)이 서로 다른 나머지 9개의 *Candidatus*는 곤충에 공생한다는 공통점이 있지만 COG0539 (Ribosomal protein S1) 하나만 공통적이었고, 이는 곤충 세포내 공생체들 사이에 보존적 유전자가 다양함을 나타내는 것으로 판단되었다. 본 연구의 결과는 배양이 불가능한 세균의 보존적 유전자 이해에 대한 단서와 함께 아미노산, 항생제, 의약품, 유기합성 전구체 등을 효율적으로 합성하는 원핵생물의 조작에 필요한 기초자료로 활용이 가능할 것이다.