

## Comparison of Mitochondria-related Conserved Genes in Eukaryotes and Prokaryotes

Dong-Geun Lee\*

Department of Pharmaceutical Engineering, College of Medical and Life Sciences, Silla University, San 1-1, Gwaepop-dong, Sasang-gu, Busan 617-736, Korea

Received March 5, 2014 / Revised June 26, 2014 / Accepted July 16, 2014

Sixty-two conserved orthologous groups (OGs) of proteins, in 63 prokaryotes and seven eukaryotes were analyzed to identify essential proteins in the mitochondria of eukaryotes, and their counterparts in prokaryotes. Twenty OGs were common in eukaryotic mitochondria, and all were translation related. *Encephalitozoon cuniculi*, an obligate parasitic eukaryote, shares no common mitochondrial OGs with the other 69 organisms. Seventeen conserved OGs were mitochondria related in the 69 organisms. Mitochondria related- and nonrelated-OGs were divided into prokaryotic genomes ( $p < 0.001$ , paired  $t$ -test) unlike eukaryotic genomes in the distance value analysis. The most commonly conserved mitochondria-related OG was COG0048-KOG1750 (ribosomal small subunit S12), whereas it was COG0100-KOG0407 (ribosomal small subunit S11) in nonrelated OGs. These results could be applied in scientific research to determine phylogenetic relationships and in areas such as drug development.

**Key words** : Clusters of Orthologous Groups of proteins (COG), conserved gene, euKaryotic Orthologous Group (KOG), mitochondria

### 서 론

생물의 게놈서열을 파악하고 유전자의 갯수와 종류를 밝히는 genome annotation 작업이 오늘날 많이 수행되고 있다[21]. Annotation은 상동성(homology)에 기반하는데 상동서열(homolog)은 ortholog와 paralog로 나눌 수 있다[9]. 서로 다른 생물종들에 분포하지만 공통조상 유전자에서 유래한 유전자들의 집합이 ortholog이며, 동일 ortholog에 속하는 단백질들은 서열과 기능이 유사하거나 동일하다[16]. 동일 유전체내에서 유전자 복사(gene duplication)로 생긴 유전자들의 집합이 paralog이며, 이들 사이에 공통 기능은 거의 없고 새로운 기능이 생긴다[17]. 따라서 ortholog/paralog의 구분은 학문적으로 진화관계의 재구축에, 응용적 측면에서는 게놈에서 미지 단백질의 기능 유추를 포함하는 genome annotation [13, 16, 17], 병충해 방지 농약의 개발[21], 항생제와 치료제 연구[2, 14] 등에 필수적이다[13].

OG (Orthologous Group of proteins)는 ortholog들에서 유래된 단백질의 집합으로 구조와 기능이 유사하다. COG (Clusters of OG)가 미생물에서 보고된 이후 진핵생물의 KOG

(euKaryotic OG)와 orthoDB, 파지의 POG (Phage COG), 세포내 기생세균인 mollicutes의 OG (MOG), 고세균의 arCOG (archaeobacterial COG), 그리고 세균, 고세균, 진핵생물 등을 가리지 않고 OG를 파악하는 eggNOG, OMA 등의 database가 보고되고 있다[15].

진핵생물의 유전물질은 핵에 주로 분포하고 있으며 세포소기관인 미토콘드리아에, 식물은 추가로 엽록체에도 분포하고 있다. 유전물질을 가지는 세포소기관들의 유래는 생물공생설에서는 원핵생물에서 유래한 것으로 파악하고 있다. *Euryarchaeobacteria*와 *α-Proteobacteria*의 자매(sister) 유전자가 진핵생물 핵내에 많이 분포하며[17] 미토콘드리아 유전자는 핵내 유전자와 다른 진화속도 및 경로를 거쳤다는 보고[10], 미토콘드리아는 *α-Proteobacteria*에 속하는 *Rickettsia* 종(species) 혹은 해양의 dwelling clade에서 유래되었다고 한다[7]. 또한 미토콘드리아는 생체에너지인 ATP 생산에 중요하며 미토콘드리아 DNA의 이상은 근육과 뇌 등 신경계의 질병, 당뇨병과 청각 및 시각의 상실 등을 야기시키는 등 인간에게도 매우 중요하다[4].

변화하는 환경에 적응하기 위해 생물들은 유전자 획득과 상실로 공통조상의 유전적 조성과 달라졌지만[20], 현재의 생명체 이해를 위해서는 모든 생명체의 공통적 유전자에 대한 이해가 중요하다[18, 27]. 이 등[12]은 진핵생물 3종을 포함한 66종의 미생물 모두에서 63개의 ortholog가 보존적이라고 보고하였다. 이[11]는 66종의 미생물과 7종의 진핵생물에 공통적인 진핵미생물 3종을 기반으로 진핵생물 7종에서 공통적으로 보존되는 유전자들의 종류와 기능 및 보존성의 정도를 파악하

#### \*Corresponding author

Tel : +82-51-999-6282, Fax : +82-51-999-5636

E-mail : ldg@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.

였다. 원핵과 진핵생물에 공통적인 OG중 미토콘드리아 관련 OG에 관한 연구는 부족한 실정이다.

이에 본 연구는 원핵과 진핵생물에 공통적으로 보존적인 OG를 미토콘드리아 관련 OG와 비관련 OG로 파악함으로써 현재 생명체에 대한 이해를 높이고자 하였다. 이를 위하여 70종 생물의 유전체에서 진핵생물 기준 70개의 OG를 추출하고 아미노산 서열분석을 하였다. 보존성은 ribosomal small sub-unit가 높은 것을 확인하였다. 고세균, 진정세균, 진핵생물 각 그룹별로 분석하여 원핵생물은 미토콘드리아 관련 OG와 비관련 OG로 구분되지만 진핵생물은 그렇지 않음을 확인하였다.

### 재료 및 방법

#### 재료

유전체 정보는 COGs, KOGs 등 NCBI의 공개 서버에서 추출하였다[5]. 원핵생물과 진핵생물의 유전자 유사성에 관한 자료는 COGs와 KOGs에서 정리된 자료를 이용하였는데, COGs는 2013년 8월 현재 63종의 원핵생물과 3종의 단세포 진핵생물 등 총 66종의 미생물 유전체에 포함된 총 144,324개의 유전자들을 4,873개의 COG 그룹으로 분류해 놓았다[6]. 그리고 KOGs는 2012년 12월 현재 7종의 진핵생물 유전체에 포함된 총 121,518개의 유전자들을 4,852개의 KOG 그룹으로 분류해 놓았다[6]. Table 1은 실제로 분석한 자료인 70종의 생물들이다. COGs와 KOGs에서 단세포 진핵생물인 *Saccharomyces cerevisiae* (빵효모), *Schizosaccharomyces pombe* (분열효모), *Encephalitozoon cuniculi* (뇌회백염원충) 3종은 공통이었다. 따라서

총 70종의 유전체 자료를 분석하였다.

#### 아미노산 서열 분석

66종의 미생물들이 공통적으로 함유하는 보존적 COG 63종류에 속하는 단백질들의[12] 이름과 서열을 기반으로 KOG 공개 데이터베이스에서 원핵과 진핵생물에 보존적이며 진핵생물에 분포하는 KOG를 추출하였으며, 원핵생물 63종과 진핵생물 7종 모두에 보존적인 COG는 62개였다[11]. 추출된 KOG와 그에 상응하는 COG 각각에서 단백질 서열을 확보하였고 ClustalX ver. 2.1을 이용한 다중서열정렬을 수행한 후 phylogenetic tree를 작성하면서 bootstrap (n=1,000) 분석으로 신뢰도를 높였다[10]. Bootstrap 분석으로 distance value를 담고 있는 '\*.phb' 파일을 작성하였다. 서열정렬의 parameter는 기본값을 이용하였다. Phylodraw 프로그램(ver 0.8) [3]을 이용하여 각 단백질의 distance value를 구하였고, distance value를 포함한 자료의 분석과 정리에는 MS사의 엑셀 프로그램을 이용하였다.

#### 보존성 분석

70종의 생물에 보존적인 COG 기준 62개, KOG 기준 71개의 OG (Orthologous Group of proteins)가 나타내는 distance value의 평균과 분산을 각 OG와 생물에 대하여 분석하여 보존성 정도를 파악하였다. 보존적 OG를 미토콘드리아 관련 OG와 비관련 OG로 나누어 분석하였다. 진핵생물의 미토콘드리아에 존재하는 OG와 그에 상응하는 원핵생물의 OG를 미토콘드리아 관련 OG로 간주하였다.

Table 1. Origin of studied 70 genomes. Abbreviations are shown in parenthesis of Eucarya

Phylogenetic Group	Organism
Archaea	<i>Aeropyrum pernix</i> , <i>Archaeoglobus fulgidus</i> , <i>Halobacterium</i> sp. NRC-1, <i>Methanobacterium thermoautotrophicum</i> , <i>Methanococcus jannaschii</i> , <i>Methanopyrus kandleri</i> , <i>Methanosarcina acetivorans</i> C2A, <i>Pyrobaculum aerophilum</i> , <i>Pyrococcus abyssi</i> , <i>Pyrococcus horikoshii</i> , <i>Sulfolobus tokodaii</i> , <i>Thermoplasma acidophilum</i> , <i>Thermoplasma volcanium acidophilum</i> , <i>Thermoplasma volcanium</i> , <i>Methanosarcina acetivorans</i> C2A
Bacteria	<i>Agrobacterium tumefaciens</i> , <i>Aquifex aeolicus</i> , <i>Bacillus halodurans</i> , <i>Bacillus subtilis</i> , <i>Borrelia burgdorferi</i> , <i>Brucella melitensis</i> , <i>Buchnera</i> sp. APS, <i>Campylobacter jejuni</i> , <i>Caulobacter crescentus</i> , <i>Chlamydia pneumoniae</i> , <i>Chlamydia trachomatis</i> , <i>Clostridium acetobutylicum</i> , <i>Corynebacterium glutamicum</i> , <i>Deinococcus radiodurans</i> , <i>Escherichia coli</i> K12, <i>E. coli</i> O157:H7 EDL933, <i>E. coli</i> O157:H7, <i>Fusobacterium nucleatum</i> , <i>Haemophilus influenzae</i> , <i>Helicobacter pylori</i> 26695, <i>Helicobacter pylori</i> J99, <i>Lactococcus lactis</i> , <i>Listeria innocua</i> , <i>Mesorhizobium loti</i> , <i>Mycobacterium leprae</i> , <i>M. tuberculosis</i> CDC1551, <i>Mycobacterium tuberculosis</i> H37Rv, <i>Mycoplasma genitalium</i> , <i>M. pneumoniae</i> , <i>M. pulmonis</i> , <i>Neisseria meningitidis</i> MC58, <i>N. meningitidis</i> Z2491, <i>Nostoc</i> sp., <i>Pasteurella multocida</i> , <i>Pseudomonas aeruginosa</i> , <i>Ralstonia solanacearum</i> , <i>Rickettsia prowazekii</i> , <i>Salmonella typhimurium</i> , <i>Sinorhizobium meliloti</i> , <i>Staphylococcus aureus</i> , <i>Streptococcus pneumoniae</i> , <i>St. pyogenes</i> , <i>Synechocystis</i> sp., <i>Thermotoga maritima</i> , <i>Treponema pallidum</i> , <i>Ureaplasma urealyticum</i> , <i>Vibrio cholerae</i> , <i>Xylella fastidiosa</i> , <i>Yersinia pestis</i>
Eucarya	<i>Arabidopsis thaliana</i> (A), <i>Caenorhabditis elegans</i> (C), <i>Drosophila melanogaster</i> (D), <i>Encephalitozoon cuniculi</i> (E), <i>Homo sapiens</i> (H), <i>Saccharomyces cerevisiae</i> (Sc), <i>Schizosaccharomyces pombe</i> (Sp)

결과 및 고찰

미토콘드리아 보존적 유전자

Table 2에 분석대상 70종의 생물에 보존적인 OG를 미생물에 보존적인 COG와 이에 상응하는 진핵생물의 KOG로 표시하면서, KOG는 핵과 미토콘드리아 기원에 따라 나누어 제시하였다. 71개의 보존적 KOG중[11] 20개의 KOG가 미토콘드리아 관련 OG였고 이는 COG로도 20개에 해당하는 것이었다. 이들은 모두 번역(translation)에 관여하였다. 한편 20개 COG 관점에서는 43개 KOG와 상응했는데 이러한 OG 수의 차이는

하나의 COG가 다수의 KOG와 상응하기 때문으로[11] 하나의 COG 유전자가 미토콘드리아 유전자 및 핵내 유전자 모두 분포하기 때문이었다. 실제로 COG0180은 트립토판 tRNA synthetase로 진핵생물의 세포질 tRNA synthetase인 KOG2145와 미토콘드리아 tRNA synthetase인 KOG2713과 상응했으며, translation initiation factor인 COG0532는 세포질용 KOG1144와 미토콘드리아에 존재하는 KOG1145와 상응하였다(Table 2). 진핵생물 단백질의 상당부분은 원핵생물 단백질에서 유래했거나 공통적인 것으로 보고하고 있는데[11, 18] 본 연구는 진핵생물의 미토콘드리아와 핵 그리고 원핵생물 사이에 모두

Table 2. Comparison and functional category of conserved orthologs between prokaryotic COG and eukaryotic KOG. Mitochondrial KOG was bolded. C and K in front of number is COG and KOG, respectively. Underlined KOGs are not distributed in all seven

Functional category	Function : Match of (COG #, KOG #s)		
Cell cycle control, mitosis and meiosis	ATPase for cell cycle control	(C0037, K2840)	
	tRNA synthetase	(C0008, K1147, K1148, <b>K1149</b> )	(C0180, K2145, <b>K2713</b> )
Translation	(C0016, <u>K2783</u> , K2784)	(C0060, <u>K0433</u> , K0434)	(C0072, K2472, <u>K2783</u> )
	(C0143, <u>K0436</u> , <u>K1247</u> )	(C0162, K2144, <u>K2623</u> )	(C0442, <u>K2324</u> , K4163)
	(C0495, <u>K0435</u> , K0437)	(C0012, K1491)	(C0013, K0188)
	(C0018, <u>K1195</u> )	(C0124, K1936)	(C0172, K2509)
	(C0441, K1637)	(C0525, K0432)	
	Ribosomal large subunit	(C0080, K0886, <b>K3257</b> )	(C0087, K0746, <b>K3141</b> )
	(C0088, <u>K1475</u> , <b>K1624</b> )	(C0090, <b>K0438</b> , K2309)	(C0091, <b>K1711</b> , K3353)
	(C0094, K0397, <b>K0398</b> )	(C0097, <b>K3254</b> , K3255)	(C0102, <b>K3203</b> , K3204)
	(C0197, K0857, <b>K3422</b> )	(C0200, <b>K0846</b> , K1742)	
	(C0081, <u>K1569</u> , K1570)	(C0244, K0815, K0816)	(C0089, K1751)
	(C0093, K0901)	(C0198, K3401)	(C0255, K3436)
		(C0255, K3436)	(C0256, K0875)
	Ribosomal small subunit	(C0048, K1749, <b>K1750</b> )	(C0052, K0830, <b>K0832</b> )
	(C0103, <b>K1697</b> , K1753)	(C0184, K0400, <b>K2815</b> )	(C0185, K0898, <b>K0899</b> )
	(C0199, <b>K1741</b> , K3506)		
	(C0049, K3291)	(C0092, K3181)	(C0096, K1754)
(C0098, K0877, <u>K2646</u> )	(C0099, K3311)	(C0100, K0407)	
(C0522, K3301, K4655)			
Translation elongation factor	(C0480, <b>K0465</b> , K0467, K0468, K0469)	(C0231, K3271)	
Translation initiation factor	(C0532, K1144, <b>K1145</b> )	(C0361, K3403)	
rRNA methylation	(C0030, K0820)		
Polypeptide chain release factor	(C2890, <u>K2904</u> , K3191)		
Transcription	RNA polymerase	(C0085, K0214, K0215, K0216)	(C0086, K0260, K0261, K0262)
Replication, recombination and repair	Exonuclease	(C0258, K2518, K2519, K2520)	
	Topoisomerase	(C0550, K1956)	
	DNA polymerase	(C0592, K1636)	
Posttranslational modification, protein turnovers	Protease with possible chaperon activity	(C0533, <u>K2707</u> , K2708)	
General function prediction only	EMAP domain	(C0073, K2241)	
Intracellular trafficking and secretion	Preprotein translocase	(C0201, K1373)	
	Signal recognition GTPase	(C0541, K0780)	(C0552, K0781)

보존적인 단백질이 있음을 나타냈다.

미토콘드리아에 공통적 KOG는 20개였고 K0398 (C0094), K3254 (C0097), K0899 (C0185) 등 3개 KOG는 애기장대(Table 1의 Eucarya의 약어 A), 빵효모(Sc), 분열효모(Sp)에만 존재하였다. 총 71개의 보존적 KOG중[11] 17개 KOG가 진핵생물 6종에 보존적이었는데, 특이한 생물은 뇌회백염원충(E)으로 20개 KOG를 하나도 보유하지 않았다. 절대기생성 진핵생물인 뇌회백염원충의 미토콘드리아에 해당하는 기능이 결여되어 있다고 판단할 수 있었다. Tsaousis 등[19]은 뇌회백염원충에는 세포질과 미토콘드리아 사이에 ATP 등 여러 기질의 이동을 매개하는 MCF (Mitochondrial Carrier Family)가 부족하며 ATP의 흐름이 기생숙주, 뇌회백염원충의 세포질, 뇌회백염원충의 퇴화된 미토콘드리아인 미토솜(mitosome) 방향이라고 보고하였다. 이로써 뇌회백염원충이 생명현상에 필요한 ATP를 미토솜에서 얻지 못하는 것도 숙주세포 내부에 기생하는 원인 중 하나라고 판단할 수 있었다. 각 생물체가 보유한 단백질 코딩 유전자수나 KOG의 종류도 뇌회백염원충이 비교대상 7종의 진핵생물 중 가장 적었다[11]. 미토콘드리아 관련 보존적 OG를 구성하는 단백질의 총 개수는 원핵생물이 20~24개였으며, 진핵생물은 미토콘드리아와 핵 유래를 모두 합쳐 뇌회백염원충(E) 24개, 초파리(D) 44개, 꼬마선충(C) 45개, 빵효모(Sc) 56개, 분열효모(Sp) 60개, 애기장대(A) 106개, 사람(H) 229개로 뇌회백염원충이 다른 종들에 비해 현저히 낮았다.

총 20개의 미토콘드리아 보존적 KOG는 모두 번역(translation)에 관련되었다. Integral membrane protein 관련 유전자는 핵에서 미토콘드리아로 이동하였다는 보고도 있었다[7]. 즉 미토콘드리아의 2중막 구조에서 외막은 통과하고 내막에 위치할 전자전달계 등의 막단백질을 세포질에서 공급할 수 없고 미토콘드리아가 공급하므로 이들의 합성에 필요한 번역관련 OG 등이 미토콘드리아에 보존적일 것이다(Table 2).

Makarova 등[13]은 유전자를 생물체 사이에 전달이 없는 hard core 유전자군, 전달이 거의 없는 soft core 유전자군, 전달이 가능한 shell 유전자군으로 분류하였으며, 대장균 종(species)의 균주 사이에 유전자들의 40%만 공통이었다[20]. 이들은 기원이 다른 생물체 사이에서 horizontal gene transfer (HGT)가 광범위하다고 하였다[13, 20]. HGT가 드물어 생물계통에서 백만 년에 한번 일어나도 원핵생물의 진화기간인 20억 년 동안 2000개가 전달될 수 있다[13]. 미토콘드리아 보존적 OG에도 HGT가 포함될 수는 있지만 비교대상 생물들에 필수적일 가능성은 매우 높을 것이다.

**생물별 미토콘드리아 관련 유전자군의 보존성**

개별 유전자가 아닌 유전자 수가 많아지면 distance value로 유전체의 특성을 나타낼 수 있다[8]. Fig. 1과 Fig. 2는 원핵생물 63종과 진핵생물 7종에 모두 보존적인 OG를 대상으로 phylogenetic tree에서 나타나는 distance의 평균과 분산을 이

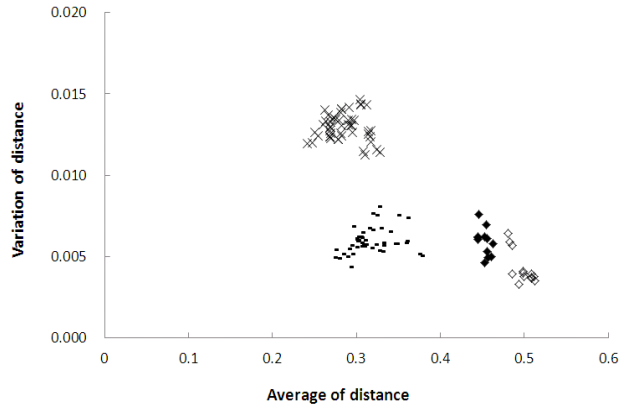


Fig. 1. Distribution pattern of prokaryotic genomes by distance value with all conserved orthologs between 63 prokaryotes and 7 eukaryotes. Symbol (phylum, site of orthologous groups in eukaryotes) is × (eubacteria, non-mitochondria), - (eubacteria, mitochondria), ◇ (archaeobacteria, mitochondria), and ◆ (archaeobacteria, non-mitochondria), respectively. Distance value represents the number of substitutions as a proportion of the length of the alignment.

용하여 표시한 개별 유전체의 분포 결과로, 미토콘드리아 관련 OG와 미토콘드리아 비관련 OG로 나누어 표시하였다. 진핵생물의 미토콘드리아 KOG에 상응하는 원핵생물의 COG를 미토콘드리아 관련 OG로 분류하였다.

고세균과 진정세균 모두(Fig. 1) 미토콘드리아 관련 OG와 비관련 OG 등 각각 2개의 그룹으로 나누어졌고( $p < 0.001$ , paired  $t$ -test), 진핵생물은(Fig. 2) 그렇지 않았다( $p > 0.05$ , paired

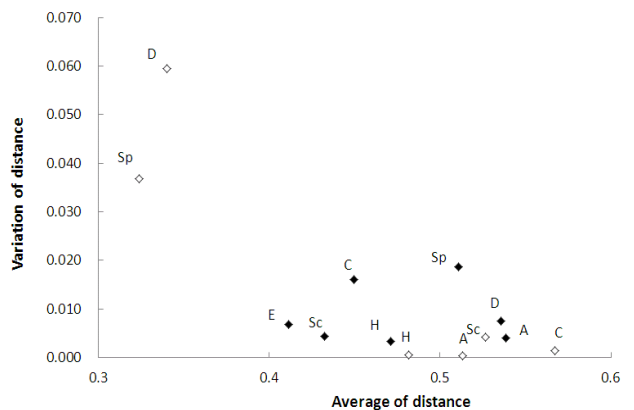


Fig. 2. Distribution pattern of eukaryotic genomes by distance value with all conserved orthologs between 63 prokaryotes and 7 eukaryotes. Symbol (site of orthologous groups) is ◇ (mitochondria) and ◆ (non-mitochondria), respectively. See Fig. 1 for distance value. Each abbreviation represents organism such as *Arabidopsis thaliana* A, *Caenorhabditis elegans* C, *Drosophila melanogaster* D, *Encephalitozoon cuniculi* E, *Homo sapiens* H, *Saccharomyces cerevisiae* Sc, and *Schizosaccharomyces pombe* Sp.

t-test). 고세균의 경우 평균이 두 OG에서 큰 차이를 보였고, 진정세균과 진핵생물은 큰 차이가 없었다. OG의 distance 평균을 t-test로 분석하니 미토콘드리아 관련 OG와 비관련 OG 모두 진핵생물과 진정세균, 고세균과 진정세균 사이에 유의성이 아주 높은( $p < 0.001$ ) 차이가 있었다. 진핵생물과 고세균 사이에 나타나는 OG의 distance 평균의 차이는 미토콘드리아 관련 OG ( $p < 0.05$ )와 미토콘드리아 비관련 OG ( $p < 0.01$ ) 모두 유의성을 보였다. ANOVA 결과는 미토콘드리아 관련 OG와 비관련 OG 모두 진핵생물, 진정세균, 고세균 사이에  $p < 0.001$  수준의 차이가 있었다. 즉 보존적 OG의 distance 평균은 진정세균이 고세균 및 진핵생물과 차이가 통계적 유의성이 높았다. 41개 고세균의 arCOG 분석으로 arCOG의 40%는 기존의 COG와 구분되어 고세균과 진정세균이 서로 다른 진화경로를 지녔는데[13] 본 연구에서는 각 OG를 구성하는 구성원의 distance 즉 변이도 역시 차이를 보였다. 개별 단백질의 변이속도가 생물 사이에 큰 차이를 보이지 않는다고 가정하면 평균이 높다는 것은 고세균의 탄생이 진정세균보다 빨랐다는 것으로 판단되었다.

한편 분산이 적은 것은 모든 단백질의 변이도가 비슷하다는 뜻일 것이다. 즉 고세균은 단백질들의 변이 정도가 골고루 크며, 진정세균의 경우 지구상에서의 출현은 고세균보다 늦지만 미토콘드리아 관련 OG에 속하는 각 단백질들의 변이속도는 차이가 크고 미토콘드리아 비관련 OG는 유사하다는 것이다.

미토콘드리아 유전자는 핵내 유전자와 다른 진화속도 및 경로를 겪었다는 보고가 있었다[10]. 본 연구에서는 초파리(Fig. 2D), 꼬마선충(C)에서는 진화속도가 많이 다르고 애기장대(A)와 사람(H)에서는 그 차이가 크지 않았다(Fig. 2). OG의 distance 평균을 보면 꼬마선충(C), 빵효모(Sc), 사람(H)은 미토콘드리아 관련 OG가 미토콘드리아 비관련 OG보다 높았고 분열효모(Sp), 초파리(D), 애기장대(A)는 반대였다(Fig. 2).

한편 *Aquifex aeolicus*를 제외한 모든 진정세균들은 미토콘드리아 관련 OG가 비관련 OG보다 distance의 평균이 적었고, 고세균은 반대였다( $p < 0.001$ , paired t-test). 이는 다음과 같은 유추가 가능하게 하였다. 첫째, 변이의 속도가 미토콘드리아 관련 OG들과 미토콘드리아 비관련 OG들 사이에 차이가 있었을 가능성이다. 이는 다시 DNA 결합단백질 등에 의한 DNA 보호 정도와 유전자 길이 등으로 나눌 수 있었다. DNA 결합단백질 등에 의한 DNA 보호와 관련하여 미토콘드리아 관련 유전자군들의 상시발현과 비관련 유전자군들의 조건발현을 생각할 수 있었다. 상시발현을 위해서는 DNA의 packing과 DNA 결합단백질들의 결합이 느슨하여 미토콘드리아 관련 유전자군들이 변이를 겪을 가능성이 증대될 것이다. 하지만 Table 2에서 진핵생물 6종과 원핵생물 63종에 모두 보존적인 미토콘드리아 관련 OG는 COG 관점에서 tRNA synthetase 2개, ribosomal large subunit 10개, ribosomal small subunit 6개 등이다. 진핵생물과 원핵생물 모두에 보존적이며 핵 유래

인 OG의 갯수가 COG 관점에서 tRNA synthetase 14개, ribosomal large subunit 7개, ribosomal small subunit 7개 등이었다(Table 2). Ribosome의 large- 그리고 small-subunit는 수십 개 단백질의 결합체이므로 ribosome이 제대로 기능을 하기 위해서는 모든 subunit가 있어야, 즉 관련 유전자 모두가 발현되어야 하므로 미토콘드리아 관련 OG와 비관련 OG의 distance 평균이 차이를 보이는 것은 조건발현으로 설명이 되지 않는다. DNA 복구효소나 활성산소종 등 산화적 스트레스 저항 효소에 의한 미토콘드리아 DNA의 보호 가능성 등도 제기되지만, 핵막이 없는 원핵생물은 이들 효소가 모든 유전자들을 동일하게 보호할 것이다.

원핵생물은 핵막이 없으므로 미토콘드리아 관련 유전자군과 비관련 유전자군 각각이 유전자 길이당 변이원에 노출되는 확률은 동일할 것이다. 아미노산 갯수를 비교하면 미토콘드리아 관련 OG들이 평균 265.13, 중앙값 179, 표준편차 208.26이고 미토콘드리아 비관련 OG들이 평균 401.34, 중앙값 327, 표준편차 304.86이었다. 이는 미토콘드리아 비관련 OG들이 미토콘드리아 관련 OG들에 비해 아미노산 갯수 즉 유전자가 더 크지만(t-test,  $p < 0.001$ ), distance의 평균이 고세균에서는 적고 진정세균에서는 컸다. 즉 유전자 길이에 의한 변이의 차이 가능성은 특히 고세균에서 낮다는 것을 알 수 있었다.

한편 미토콘드리아 관련 OG는 비관련 OG에 비해 유전자의 길이는 짧고 발현정도는 비슷한데, distance의 평균은 미토콘드리아 관련 OG가 크다는 것은 기원이 다른 두 유전체의 결합가능성이 제기된다. 특히 고세균의 경우 진핵생물의 내부공생설처럼 에너지 대사를 주로 하는 원시생명체와 그렇지 않은 생명체의 결합으로 고세균의 원시조상이 탄생했을 수도 있을 것이지만 추후의 연구가 더 필요할 것이다. 고세균은 진핵생물과 공통적인 유전정보처리관련 core 유전자들을 가지고 물질대사, 구조, 신호전달 관련 유전자군은 shell 유전자군으로 진정세균과 유사하여 두 유전자군들의 진화적 경로가 다르다는 보고도 있었다[13].

#### 유전자의 보존 정도

Fig. 3은 63종의 원핵생물과 7종의 진핵생물에 분포하는 OG의 보존 정도를 미토콘드리아 관련 OG ( $\diamond$ )와 비관련 OG ( $\blacklozenge$ )로 나누어 나타내었다. 미토콘드리아 관련 여부와 상관없이 OG들의 변이 정도는 유사한 것으로 판단되었다.

Fig. 3A는 ortholog와 paralog를 합친 분포이고 Fig. 3B는 ortholog만의 분포이다. 분산에서 차이를 보이는 이유는 ortholog만을 추출한 후 이들만으로 정렬을 다시 수행한 것 때문일 것이다. Paralog까지 합친 경우 미토콘드리아 관련 OG의 distance가 최소 0.315, 최대 0.512, 평균 0.385, 중앙값 0.394, 분산 0.002였으며 미토콘드리아 비관련 OG가 최소 0.308, 최대 0.510, 평균 0.383, 중앙값 0.389, 분산 0.002로 두 OG 그룹 사이에 차이가 없었고 ortholog만 고려한 경우도 차이가 없었

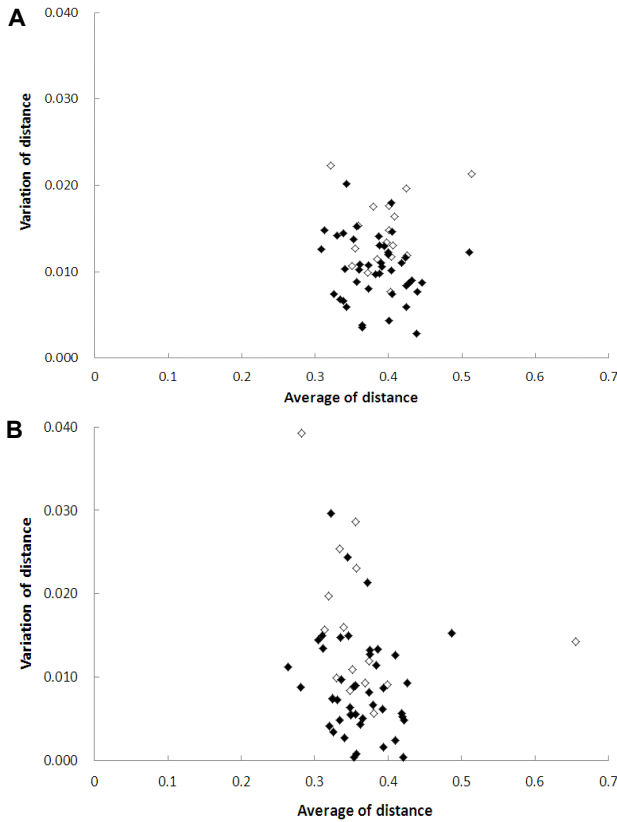


Fig. 3. Distribution of 62 common OGs by distance value with (A) orthologs and paralogs and (B) orthologs only. Symbol  $\diamond$  and  $\square$  represents mitochondria-related and nonrelated OGs, respectively. See Fig. 1 for distance value.

다( $p > 0.05$ ). 동일 OG의 distance를 paralog와 ortholog를 합친 경우와 ortholog만 고려한 경우로 나누어 비교하면 COG0052와 COG0552만 ortholog만 고려한 것이 높았다. 유전체에 paralog가 없으면 ortholog의 보존성이 낮고, paralog가 있으면 높다는 보고가 있었는데[18] 본 연구의 결과와 대체로 부합되었다. Ortholog+paralog와 ortholog의 변이평균을 paired *t*-test로 비교하면 미토콘드리아 관련 OG는 유의성  $p=0.047$  수준에서, 미토콘드리아 비관련 OG는 유의성  $p=4.76 \times 10^{-10}$  수준에서 차이를 보여 미토콘드리아 비관련 OG가 변이의 통계적 유의성이 훨씬 높은 것으로 즉 미토콘드리아 관련 OG는 paralog가 있어도 변이의 통계적 유의성이 아주 크지 않음을 알 수 있었다. 미토콘드리아 관련 OG와 비관련 OG를 구성하는 단백질의 수를 *t*-test로 비교하면  $p=0.01$  수준에서 차이가 없었다. Paralog를 제외한 ortholog들만으로 distance 평균을 비교하면, 미토콘드리아 관련 OG에서는 COG0048-KOG1750 (ribosomal small subunit S12)이 0.3157로 최소였고 COG0052-KOG0832 (ribosomal small subunit S2)가 0.5120으로 최대였으며, 미토콘드리아 비관련 OG에서는 COG0100-KOG0407 (ribosomal small subunit S11)이 0.3087로 최소였고

COG0086-K0260-K0261-K0262 (subunit of DNA-directed RNA polymerase)가 0.5100으로 최대였다. 이는 해당 유전자의 존재가 미토콘드리아 DNA 혹은 핵 DNA에 상관없이 미토콘드리아 자체의 ribosome을 구성하는 subunit의 보존성이 진핵생물의 활동에 중요하다는 것을 알려주는 것으로 판단되었다.

결론적으로 진핵과 원핵생물 유래 70종의 유전체에서 진핵생물 기준으로 70개의 미토콘드리아 OG들을 구성하는 각 구성원에 대한 분석을 통하여 첫째, 미토콘드리아 보존적 OG는 모두 미토콘드리아 자체의 번역(translation)에 관련되었고 둘째, 분석대상 OG들이 고세균과 진정세균 즉 원핵생물에서는 미토콘드리아 관련 OG와 비관련 OG로 보존성이 구분되지만 진핵생물은 구분되지 않았다. 셋째, 보존성은 미토콘드리아 ribosomal small subunit가 높은 것을 확인하였다.

본 연구로 도출한 17개의 진핵생물 미토콘드리아 보존적 KOG는 현재 진핵생물의 미토콘드리아에서 생명체의 기능 유지에 필수적 역할을 담당하는 것으로 간주할 수 있으며 원핵생물에도 분포한다. 본 연구결과는 미토콘드리아 관련 유전자와 원핵생물의 보존성과 함께 항생제의 부작용을 줄이기 위해 공통적이지 않은 유전자를 타겟으로 한 치료 대상 단백질 연구[11, 14, 19] 등 기능적 연계에 대한 이해를 높일 수 있을 것이다.

### References

- Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., Naslund, A. K., Eriksson, A. S., Winkler, H. H. and Kurland, C. G. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 134-140.
- Buysse, J. M. 2001. The role of genomics in antibacterial target discovery. *Curr Med Chem* **8**, 1713-1726.
- Choi, J. H., Jung, H. Y., Kim, H. S. and Cho, H. G. 2000. PhyloDraw: a phylogenetic tree drawing system. *Bioinformatics* **16**, 1056-1058.
- Finsterer, J. 2007. Hematological manifestations of primary mitochondrial disorders. *Acta Haematol* **118**, 88-98.
- Fraser, C. M., Eisen, J. A. and Salzberg, S. L. 2000. Microbial genome sequencing. *Nature* **406**, 799-803.
- <http://www.ncbi.nlm.nih.gov/COG/>
- Huynen, M. A., Duarte, I. and Szklarczyk, R. 2013. Loss, replacement and gain of proteins at the origin of the mitochondria. *Biochim Biophys Acta* **27**, 224-231.
- Kimura, M. 1983. The neutral theory of molecular evolution. *Cambridge University Press*.
- Kristensen, D. M., Kannan, L., Coleman, M. K., Wolf, Y. I., Sorokin, A., Koonin, E. V. and Mushegian, A. 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* **26**, 1481-1487.
- Kuramae, E. E., Robert, V., Snel, B. and Boekhout, T. 2006.

- Conflicting phylogenetic position of *Schizosaccharomyces pombe*. *Genomics* **88**, 387-393.
11. Lee, D. -G. 2013. Investigation of conserved genes in eukaryotes common to prokaryotes. *J Life Sci* **23**, 595-601.
  12. 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.
  13. 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.
  14. Moon, E. K., Chung, D. I., Hong, Y. C., Ahn, T. I. and Kong, H. H. 2008. *Acanthamoeba castellanii*: gene profile of encystation by ESTs analysis and KOG assignment. *Exp Parasitol* **119**, 111-116.
  15. Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., Jensen, L. J., von Mering, C. and Bork, P. 2012. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* **40**, D284-D289.
  16. 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 Bioinformatics* **4**, 41.
  17. Tatusov, R. L., Koonin, E. V. and Lipman, D. L. 1997. A genomic perspective on protein families. *Science* **278**, 631-637.
  18. Thiergart, T., Landan, G., Schenk, M., Dagan, T. and Martin, W. F. 2012. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biol Evol* **4**, 466-485.
  19. Tsaousis, A. D., Kunji, E. R., Goldberg, A. V., Lucocq, J. M., Hirt, R. P. and Embley, T. M. 2008. A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. *Nature* **453**, 553-556.
  20. Welch, R. A., Burland, V., Plunkett, G. 3<sup>rd</sup>, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., Perna, N. T., Mobley, H. L., Donnenberg, M. S. and Blattner, F. R. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* **99**, 17020-17024.
  21. Wibberg, D., Jelonek, L., Rupp, O., Hennig, M., Eikmeyer, F., Goesmann, A., Hartmann, A., Borriess, R., Grosch, R., Pühler, A. and Schlüter, A. 2013. Establishment and interpretation of the genome sequence of the phytopathogenic fungus *Rhizoctonia solani* AG1-IB isolate 7/3/14. *J Biotechnol* **20**, 142-155.

## 초록 : 진핵생물과 원핵생물의 미토콘드리아 관련 보존적 유전자 비교

이동근\*

(신라대학교 제약공학과)

원핵과 진핵생물에 공통 보존적인 OG (Orthologous Group of proteins)를 미토콘드리아 관련 OG와 비관련 OG로 나누어 분석하였다. 62개의 원핵-진핵생물 공통적 COG (Clusters of OG)중 20개가 미토콘드리아 관련 OG였고 이들은 모두 번역관련 OG로 생명현상에서의 단백질의 중요성을 확인할 수 있었다. 세포내 절대기생체인 뇌회백염원충은 비교대상 다른 생물들 모두에 공통적인 미토콘드리아 관련 OG가 전혀 없었다. 뇌회백염원충을 제외한 6개 진핵생물과 원핵생물 63중에 모두 보존적인 미토콘드리아 관련 OG는 17개였다. Phylogenetic tree의 distance 분석을 수행하니 보존적 OG가 원핵생물에서 미토콘드리아 관련 OG와 비관련 OG 등 각각 2개의 그룹으로 나누어 졌고( $p < 0.001$ , paired *t*-test) 진핵생물은 그렇지 않았다( $p > 0.05$ , paired *t*-test). 보존성이 가장 높은 ortholog는 미토콘드리아 관련 OG에서는 COG0048-KOG1750 (ribosomal small subunit S12)이었고, 미토콘드리아 비관련 OG에서는 COG0100-KOG0407 (ribosomal small subunit S11)이었다. 본 연구결과는 진화관계 등의 기초학문적 연구와 치료제 개발 등의 자료가 될 수 있을 것이다.