ISSN (Print) 1225-9918
ISSN (Online) 2287-3406
Journal of Life Science **2013** Vol. 23. No. 4. 595~601 DOI : http://dx.doi.org/10.5352/JLS.2013.23.4.595

- Note -

Investigation of Conserved Genes in Eukaryotes Common to Prokaryotes

Dong-Geun Lee*

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

Received January 21, 2013 / Revised March 19, 2013 / Accepted April 10, 2013

The clusters of orthologous groups of proteins (COG) algorithm was applied to identify essential proteins in eukaryotes and to measure the degree of conservation. Sixty-three orthologous groups, which were conserved in 66 microbial genomes, enlarged to 104 eukaryotic orthologous groups (KOGs) and 71 KOGs were conserved at the nuclear genome of 7 eucaryotes. Fifty-four of 71 translation-related genes were conserved, highlighting the importance of proteins in modern organisms. Translation initiation factors (KOG0343, KOG3271) and prolyl-tRNA synthetase (KOG4163) showed high conservation based on the distance value analysis. The genes of *Caenorhabditis elegans* appear to harbor high genetic variation because the genome showed the highest variation at 71 conserved proteins among 7 genomes. The 71 conserved genes will be valuable in basic and applied research, for example, targeting for antibiotic development.

Key words : Conserved gene, genome, ortholog, Clusters of Orthologous Groups of proteins (COG), euKaryotic Orthologous Group (KOG)

서 론

생명의 기원과 진화에 대한 탐구심으로 많은 연구들이 있어 왔다. 그중에서 인간을 포함한 진핵생물의 발생과 진화과정에 대한 많은 가설들이 있어왔지만 크게 두 가지로 나눌 수 있다. 즉 다른 생물의 도움없이 진핵생물의 공통조상에서 많은 진핵생물이 유래되었다는 autogenous model과 다양한 원핵생물들의 공생에 의해 진핵생물이 유래되었다는 symbiogenic model이 있다[18].

진화과정에서 '공통조상 유전자(ancestral gene)'는 종분화(speciation)로 여러 생물의 유전체에 분포하였으며 '유전자 복사(gene duplication)'와 돌연변이에 의해 새로운 유전자가 만들어졌다고 알려져 있다[16]. 공통조상 유전자에서 유래하여 종분화로 나타나 서로 다른 생물종들에 있는 유전자들의 집합을 ortholog라고 하며, 동일 ortholog에 속하는 구성원들은 서열과 기능이 유사하거나 동일하다[15]. 한편 하나의 유전체내에서 하나의 유전자의 복사로 이루어진 유전자들의 집합을 paralog라고 하며, 이들은 복사된 유전자에 새로운 기능이 부여되어 구성원 사이의 공통 기능은 거의 없는 것이다[16].

COG (Clusters of Orthologous Groups of protein)는 ortholog들에서 유래된 단백질의 집합으로 유사한 구조와 기능을 갖는 것으로 알려져 있다. 각 COG는 하나의 공통조상유전자에서 기원하며 3가지 이상의 생물종에 분포하는 단백질들의 집합이며, COG의 파악은 유전체서열에서 파악한 유전자 생성물인 단백질들 사이의 아미노산 서열 비교를 통하여 얻어진다[15, 16]. 한편 유전체(genome) 서열 분석기술의 발달로 많은 생물종의 유전체 서열이 파악되었으며 생명의 신비와 분류학적 관계 파악에 대한 다양한 접근들이 이루어지고 있는데[18, 20] COG 알고리즘을 이용하면 각 단백질 군들에 대한 진화적 분석이라는 순수과학적 의미 이외에도 게놈에서 미지의 단백질 기능 추측과[15, 16] 광범위 혹은 협범위 항생제의 연구[2], 구조유전체학(structural genomics)의 대상 선택[1] 등에 응용된다.

강 등[9]은 COG 분석을 통해 미생물 43종에 72개의 ortholog들이 보존적인 것을 밝혔으며 이 등[11]은 진핵 3종을 포함한 66종의 미생물들에서 63개의 ortholog가 보존적이라고 보고하였다. COG가 원핵생물 63종과 진핵미생물 3종 등 미생물에 국한된 것이라면 진핵미생물 3종과 다세포 진핵생물 4종 등 총 7종의 진핵생물에서 구한 KOG (euKaryotic orthologous group)가 있다[15]. 한편 생명이 탄생한 이후에 변화하는 지구 환경에 적응하려고 다세포 생물들이 새로운 유전자를 획득하고 때로는 상실하며 공통조상의 유전적 형태와 많이 달라졌지만, 미시적 환경에서 서식하는 미생물들은 공통조상과의 공통점을 오랜 시간 유지했다는 보고와 함께[3] 진핵생물이 오히려 공통조상의 유전자를 더 많이 보존한다는 보고도 있다[20].

현재의 생명을 이해하기 위해서는 각 분류단위에 독특한 생명현상과 함께 모든 생명체가 공통적으로 나타내는 필수기능(housekeeping function)에 대한 이해가 중요하다[16]. 한편

***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.

ortholg 관련 보고는 주로 미생물이나[9, 11, 12] 진핵생물들에[15] 국한되었고 원핵과 진핵에 공통적인 ortholog에 대한 연구는 부족한 실정이다. 이 연구에서는 진핵생물 3종을 포함한 미생물 66종에서 보존적인 63개의 ortholog를 기반으로 진핵생물 7종에서 공통적으로 보존되는 유전자들의 종류와 기능 및 보존성의 정도를 파악하고자 하였다.

재료 및 방법

재료

진핵생물 유전자의 유사성에 관한 자료는 KOGs에서 정리된 자료를 이용하였다. 2012년 12월 현재 7종의 유전체에 포함된 총 121,518개의 유전자들을 4,852개의 KOG 그룹으로 분류해 놓았다[4, 7]. Table 1은 실제로 분석한 생물들이다. 진핵미생물은 빵효모인 *Saccharomyces cerevisiae* (Sce)와 분열효모인 *Schizosaccharomyces pombe* (Spo) 그리고 뇌회백염을 일으키는 원충이며 세포내 기생체인(이하 뇌회백염원충) *Encephalitozoon cuniculi* (Ecu) 3종이었다. 동물로는 선충류인 꼬마선충 *Caenorhabditis elegans* (Cel), 초파리 *Drosophila melanogaster* (Dme) 그리고 사람 *Homo sapiens* (Hsa) 이었고 식물로는 애기장대 *Arabidopsis thaliana* (Ath) 였다. Table 1에서 단백질을 코딩(coding)하는 유전자의 수는 KEGG에서 구하였다[6].

아미노산 서열 분석

66종의 미생물들이 공통적으로 함유하는 보존적 COG 63종류에 속하는 단백질들의[11] 이름과 서열을 기반으로 KOG 공개 데이터베이스에서 원핵과 진핵에 보존적이며 진핵에 분포하는 KOG를 추출하고, 각 보존적 KOG에 속하는 단백질들은 ClustalX ver. 2.1을 이용한 다중서열비교를(Fig. 1) 통해 distance value를 담고 있는 '*.dnd' 파일을 작성하였다[10, 11]. Alignment parameter는 기본값을 이용하였다. Phylodraw 프로그램(ver 0.8)을 이용하여 각 단백질의 distance value를 구하였고, distance value를 포함한 자료의 분석과 정리에는 MS사의 엑셀 프로그램을 이용하였다.

보존성 분석

66종의 원핵생물에 보존적인 COG를 바탕으로 분석대상 7종의 진핵생물 모두에서 발견되는 71개의 보존적 KOG가 나타내는 distance value의 평균과 분산을 각 KOG와 생물에 대하여 구하여 보존성 정도를 분석하였다.

결과 및 고찰

단백질 수와 KOG 구성 비율

Table 1에 분석대상 각 진핵생물이 보유하는 KOG 종류의 수(A), KOG를 구성하는 단백질 수(B), 단백질을 코딩하는 전체 유전자 수(C) 등을 나타내었다. 전체 단백질에서 KOG의 구성비율은 사람(Hsa)이 95.33%로 가장 높고 애기장대(Ath)와 꼬마선충(Cel)이 낮은 것을 알 수 있다. 애기장대(Ath)와 꼬마선충(Cel)은 단백질 코딩 유전자수가 사람보다 많지만

Table 1. Studied 7 genomes and their compositions of KOGs and proteins

Organism (Usual name, Abbreviation)	Kingdom (Phylum)	Number of			Proteins per KOG (B/A)	% of KOG proteins (B/C X100)
		KOG type (A)	KOG constituent proteins (B)	total protein genes (C)		
Arabidopsis thaliana (thale cress, Ath)	Viridiplantae (Streptophyta)	3285	13744	27396	4.18	50.17
Caenorhabditis elegans (worm, Cel)	Metazoa (Nematoda)	4235	10582	20519	2.50	51.57
Drosophila melanogaster (fruit fly, Dme)	Metazoa (Arthropoda)	4351	8445	13907	1.94	60.72
Homo sapiens (human, Hsa)	Metazoa (Chordata)	4597	19039	19972	4.14	95.33
Saccharomyces cerevisiae (baker yeast, Sce)	Fungi (Ascomycota)	2668	4003	5907	1.50	67.77
Schizosaccharomyces pombe (fission yeast, Spo)	Fungi (Ascomycota)	2762	3728	5020	1.35	74.26
Encephalitozoon cuniculi (Microsporidia, Ecu)	Fungi (Microsporidia)	1073	1218	1996	1.13	61.02

Total protein genes and the others were from KEGG [4] and KOGs [4] database, respectively.

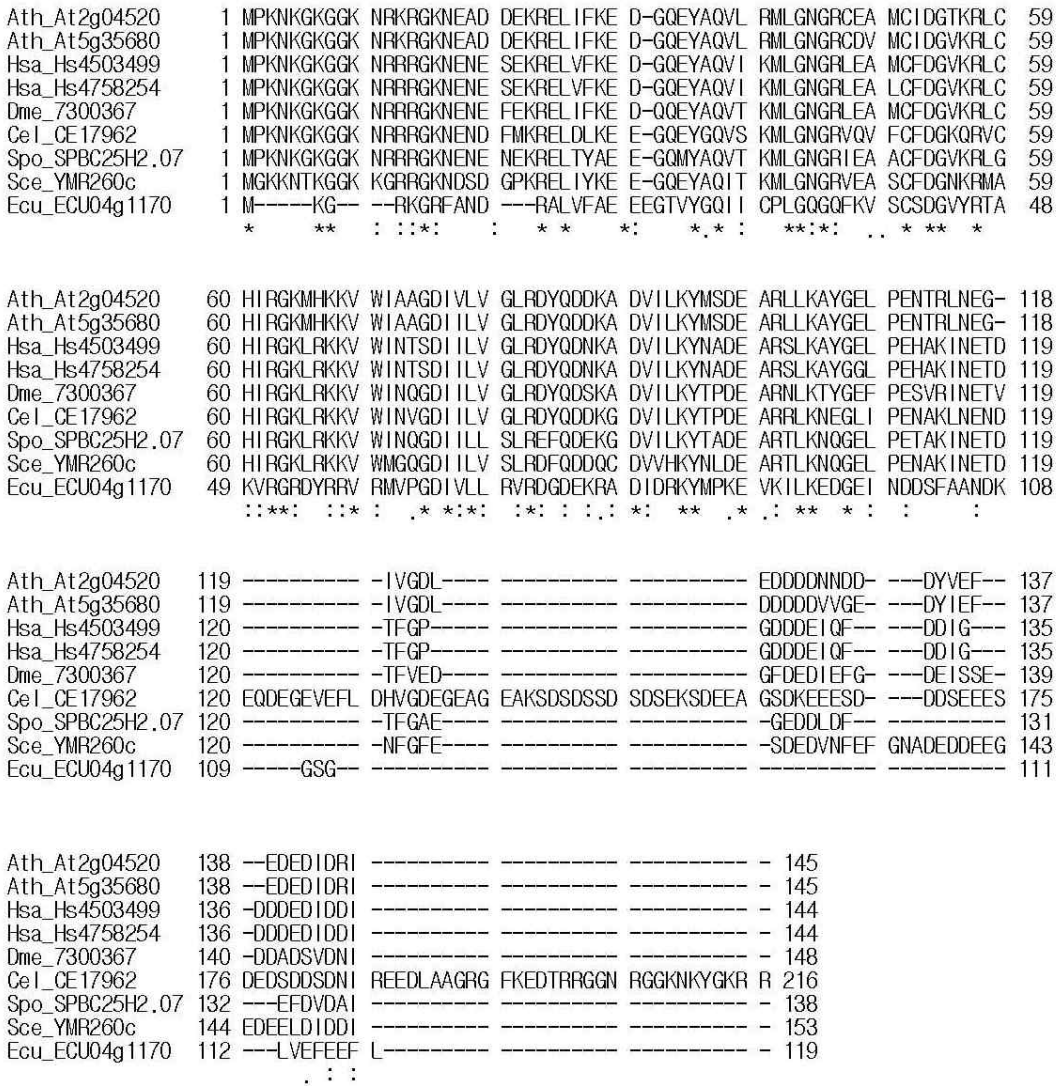

Fig. 1. Alignment of KOG3403 (Translation initiation factor 1A, eIF-1A) among 7 eukaryotes. Abbreviation of each organism (Table 1) and protein name was placed in front of and behind the underline, respectively. Each 71 KOG was aligned and analyzed as KOG3403.

KOG 구성비율이 낮아 연구대상 다른 생물들에 비해 독자적 생명현상을 보이는 단백질이 많다고 할 수 있을 것이다. 각 생물체가 보유한 단백질 코딩 유전자수나 KOG의 종류를 보면 뇌회백염원충(Ecu)가 가장 작았는데 뇌회백염원충(Ecu)가 세포내 기생체인 이유가 생명현상 유지에 필요한 단백질의 부족이라고 판단할 수 있을 것이다. 단일 KOG당 유전자수도 사람이 가장 높게 나타났다. Gabaldon과 Huynen [5]은 '수평적 유전자 전달(lateral gene transfer, LGT)'에 의한 유전자 전달을 보고하였는데 이 관점에서는 비교대상 게놈 중 사람이 가장 많은 LGT를 받았다고 할 것이다.

보존적 KOG

Table 2는 66종의 미생물 모두에서 발견되는 63개의 보존적 COG들을 기반으로 진핵생물의 KOG에서 추출한 결과를 기능별로 분류한 것이다. 66종의 미생물에 보존적인 COG와 비교한 KOG의 분석결과는 다음과 같았다.

첫째, 원핵과 진핵 70종의 생물 모두에서 보존적인 ortholog의 개수는 COG 기준으로 62개로 나타났다. 이 등이 보고한 [11] 66종의 미생물에서 63개의 COG가 보존적이라는 결과와 비교하면 COG0143 (Methionyl-tRNA synthetase)에 상응하는 두 개의 KOG중 KOG0436은 뇌회백염원충(Ecu)에서 KOG1247은 초파리(Dme)에서 해당하는 ortholog를 찾을 수 없었다. 즉 단일 KOG로 연구대상 진핵생물에 모두 분포하지는 않았다. 이러한 결과는 미생물 43종에서 72개의 COG들이 보존적이었고[9] 66종 미생물들은 63개가 보존적이었으며[11]

Table 2. Comparison and functional category of conserved orthologs between COG and KOG

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

C and K in parenthesis represents COG and KOG, respectively. Underlined KOGs are not distributed in all seven eukaryotic genome studied.

본 연구에서는 62개가 보존적이므로 비교대상 생물의 수가 늘어날수록 보존적 ortholog들의 숫자가 줄어든다는 보고와 [11] 상응하는 것이었다. 이 등[11]과 본 연구에서 공통되는 진핵생물은 빵효모(Sce), 분열효모(Spo), 뇌회백염원충(Ecu) 등 진핵미생물 3종이고 본 연구에서는 애기장대(Ath), 꼬마선충(Cel), 초파리(Dme), 사람(Hsa) 등 다세포진핵생물이 추가되었다.

특이한 것은 KOG2783이 COG0016과 COG0072와 겹쳤다. 이는 빵효모(Sce)와 분열효모(Spo)가 보유한 COG0016과 COG0072에 속하는 ortholog들 중에서 동일한 아미노산 서열을 가지고 있어 나타나는 결과로 판단되었다. Thiergart 등[18]은 진핵생물 핵내의 유전자는 euryarchaebacteria와 α-Proteobacteria의 자매(sister) 유전자가 많다고 보고하였는데 이는 진핵과 원핵생물 사이에 ortholog들이 존재한다는 것으로 본 연구와 일정부분 상응하는 것을 알 수 있었다.

둘째, 하나의 COG가 둘 이상의 다양한 KOG로 나타난 경우가 많았다. 66종 미생물에 보존적인 63개의 COG 중 36개의 COG가 둘 이상의 KOG와 연관되었다. 중복된 KOG2783을

하나로 간주하면 63개의 COG가 104개의 KOG와 연관되었다. Table 2에서 밑줄로 나타낸 것은 연구대상 7종의 모든 진핵생물에 분포하지 않고 6종 이하의 진핵생물에서 발견되는 KOG로 총 34개 KOG였다. 즉 미생물과 연구대상 진핵생물 모두에서 보존적인 ortholog들은 미생물 COG로는 62개, 진핵생물 KOG로는 71개로 나타났다. RNA polymerase 관련 COG 2개와 exonuclease 관련 COG 들은 각각 3개의 KOG로 분화되었으며 2개의 KOG로 분화한 COG도 다수이다(Table 2). 각 KOG는 다른 KOG와 다른 기능을 보이거나 유사해도 100% 일치하지 않는다[15]. Nei와 Rooney [14]는 새로운 유전자들은 대개 유전자복사(gene duplication)에 의한다고 보고하였는데 62개의 COG가 71개의 KOG로 분화한 것도 새로운 유전자의 출현으로 판단할 수 있다. 유전자복사에 의해 새로이 출현한 유전자들이 진핵생물의 종분화(speciation) 과정에서도 오랫동안 보존된 것으로 유추할 수 있었다[20]. 확산된 36개의 COG를 기능별로 분류하면 tRNA synthetase 관련 9개, ribosomal large subunit 12개 등 번역(translation) 관련한 COG가 32개로 최다였다(Table 2).

셋째, 보존적 KOG의 수와 기능을 연관시키면 번역 관련 KOG의 수와 비율이 높은 것으로 나타났다. 66종 미생물에 보존적인 63개의 COG 중 3종류 이상의 진핵에 분포하는 KOG는 104개였고 87개가 번역 관련 KOG (83.6%)로 나타났다. 이 등[11]은 66종의 미생물에 보존적인 COG중 번역에 관여하는 유전자들이 총 52개(82.5%)로 비율이 아주 높다고 보고하였는데 본 연구의 결과도 유사하였다. 또한 위 둘째와 관련하여 분화된 COG 36개 중 32개(88.9%)가 번역 관련 COG로 원핵과 마찬가지로 진핵에서도 단백질 합성이 중요하며 이는 생명체가 물질대사를 주로 수행하는 것으로 추측할 수 있었다[11]. 7종의 진핵생물에 공통적인 ribosome 관련 KOG의 수가 30개인데(Table 2) ribosome 구성 단백질들은 단백질합성 외에 전사, DNA 복구, mRNA processing, 세포자살, 발달조절 등 다양한 역할을 나타내어 생명유지에 중요한 역할을 하는 것으로 보고되고 있다[19].

넷째, 하나의 COG가 진핵생물의 핵에 분포하는 KOG와 미토콘드리아나 엽록체에 관련된 KOG로 나뉘는 것을 확인할 수 있었다. Table 2의 각 COG와 KOG를 나타내는 괄호 옆에 * 표시한 총 20개의 COG가 그런 양상을 보였다. 66종의 미생물에 보존적인 63개의 COG로 파악하면 tRNA synthetase 관련 16개의 COG 중 2개, ribosomal large subunit 관련 17개의 COG 중 10개, ribosomal small subunit 관련 13개의 COG 중 6개, translation initiation과 elongation 관련 COG 각 2개 중 하나씩 이런 양상을 보였다. 숫자나 비율에서 ribosome을 구성하는 subunit가 높게 나타났다. 미토콘드리아는 α-Proteobacteria에서 유래한 것으로 알려져 있고 미토콘드리아 독자적인 유전자들이 핵내의 게놈으로 이동하는 것으로 보고되고 있다[18].

유전자의 보존 정도

Fig. 2는 66종 미생물 모두에서 보존적인 63개의 COG를 기반으로 진핵생물 7종 모두에 분포하는 71개 KOG의 보존정도를 나타낸 것이다. 보존정도의 파악을 위하여 distance value의 평균을 이용하였다. 각 KOG들의 평균은 0.220~5.226의 분포를 보였다. 전반적으로 보면 평균과 변이가 낮은 그룹과 그렇지 않은 그룹으로 나뉘는 것을 알 수 있었다. 평균이 낮게 나타나는 경우는 각 종들 사이에 동일한 KOG를 구성하는 단백질의 아미노산서열 차이가 작다는 것을 의미하며, 이는 곧 종간의 유전자 보존성이 높음을 의미한다고 할 수 있다. Distance의 평균이 0.5 이하인 KOG는 15개로 이들은 변이(variation)도 다른 보존적 KOG들에 비해 낮은 것을 알 수 있었다. 이들은 최소값부터 KOG3403 (평균 0.220), KOG3271 (0.253), KOG4163 (0.275), KOG3301 (0.281), KOG3506 (0.287), KOG3436 (0.306), KOG4655 (0.310), KOG3311 (0.321), KOG1148 (0.333), KOG3401 (0.340), KOG3255 (0.347), KOG3204 (0.358), KOG3353 (0.368), KOG2472 (0.394), KOG3291 (0.401)로 15개 였다. KOG4655만 'RNA processing and modification'의 기능을 나타내고 나머지는 모두 'Translation, ribosomal structure and biogenesis'의 기능을 나타내어 번역관련 ortholog들이 보존된 수와 비율도 높고(Table 2) 보존성도 높은 것을 확인할 수 있었다. Fig. 1에 가장 보존성이 높은 KOG3403의 각 구성 단백질들을 정렬한 결과를 나타내었다.

강 등[9]과 이 등[11]은 distance value의 합이 낮은 것을 보존성이 높다는 것으로 간주하였는데 본 연구에서는 합이 작은 것부터 KOG3403 (합 1.976), KOG4163 (2.477), KOG4655 (2.478), KOG1148 (2.667), KOG3271 (3.546), KOG3301 (3.939), KOG2472 (3.944), KOG3506 (4.017) 순으로 distance의 평균과 유사하였다. KOG3403은 Translation initiation factor 1A (eIF-1A), KOG4163은 Prolyl-tRNA synthetase, KOG4655는

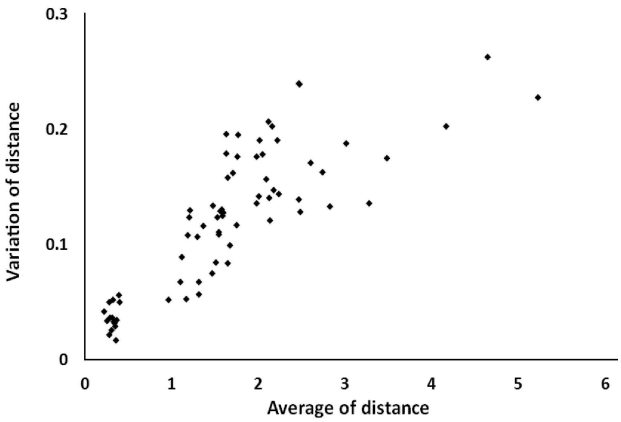

Fig. 2. Distribution of 71 common KOGs by distance value. X-axis and Y-axis represents average and variation, respectively, of distance for each KOG.

U3 small nucleolar ribonucleoprotein (snoRNP) component, KOG1148은 Glutaminyl-tRNA synthetase, KOG3271는 Translation initiation factor 5A (eIF-5A), KOG3301은 Ribosomal protein S4, KOG2472는 Phenylalanyl-tRNA synthetase beta subunit, KOG3506는 40S ribosomal protein S29로 나타나 번역관련 ortholog들의 높은 보존성을 확인할 수 있었다. 보존성이 높은 상위 10개 ortholog를 비교하면 66종의 미생물에서 6위 COG0099와 진핵 7종에서 9위 KOG3311만 동일한 ortholog였고 나머지는 모두 달랐다[11].

KOG0830 (40S ribosomal protein SA (P40)/Laminin receptor 1)은 보존성이 가장 낮아 distance의 평균도 5.525로 최대, 합도 297.88로 최대였다. 높은 합은 57개인 구성원의 수도 기여했을 것이다. 하지만 KOG3401, KOG3255, KOG3204, KOG3353은 25개 이상의 구성원으로도 distance의 평균 0.368 이하, 합 12.147 이하로 나타나는 등 구성원의 수가 낮은 보존성에 크게 기여하지 않는 것을 알 수 있었다.

한편 유전체에 paralog 없이 ortholog 하나씩만 있으면 ortholog의 보존성은 낮고, paralog가 있으면 ortholog의 보존성이 높다는 보고가 있었다[8]. 하지만 본 연구에서는 생물체의 각 유전체에 유전자 하나씩만 존재하는 KOG0261, KOG0262, KOG1147, KOG2708은 평균이 1.362 이하, 합 9.531 이하로 비교적 보존성이 높았다.

유전체의 보존성

Fig. 3는 66종 미생물에서 보존적인 63개의 COG를 기준으로 탐색한 7종의 진핵생물에서 공통적으로 보존적인 71개의 KOG가 나타내는 distance value의 평균과 분산을 이용하여 나타낸 개별 유전체의 분포 결과이다. 개별 유전자들의 distance value 관점에서는 유전체의 특성을 나타내기 어렵지만,

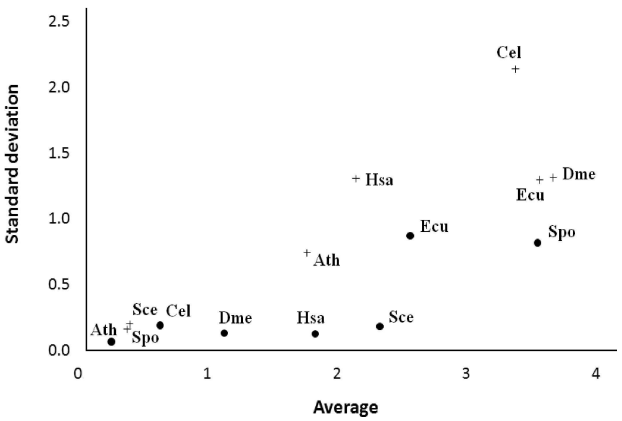

Fig. 3. Distribution pattern of eukaryotic genomes by distance value with all conserved orthologs [●] and sum of each least variated ortholog [+]. X-axis represents distribution of distance averages and Y-axis shows the standard deviation of distance averages for each organism. Sum of distance value for each conserved KOG was divided by 71.

유전자의 수가 많아지면 유전체의 특성을 표현할 있다[9, 10]. Fig. 3에서 + 표시는 공통 KOG에 속하는 모든 구성원을 고려한 경우인데 꼬마선충(Cel)이 ortholog들 평균사이의 편차가 제일 컸다. 뇌회백염원충(Ecu)과 초파리(Dme) 그리고 빵효모(Sce)와 분열효모(Spo)가 평균과 편차가 비슷하였다. 공통적인 단일 KOG에 속한 ortholog들에서 distance value가 최소인 것들만 추출하여 구한 결과로(● 표시) 판단하면 빵효모(Sce)와 분열효모(Spo)를 제외하고 ortholog의 모든 구성원을 고려한 경우보다(+ 표시) distance의 평균과 표준편차가 감소하는 것을 볼 수 있었다.

본 연구로 도출한 71개의 진핵생물 보존적 KOG는 현재 생명체의 본질적 기능에 중요한 역할을 담당하는 것으로 간주할 수 있다. 이들이 원시 진핵생명체의 출발부터 보존적이었는지, 환경변화에 순응하며 추가된 것인지, 진화과정에서 유전자의 수평적 전달에 의한 것인지[5], 유전자의 기능대체현상(gene displacement)에 의한 것인지[13] 정확히 알 수 없다. 하지만 진핵생명체의 진화 과정에서 보존된 유전자는 현재 존재하는 생명체의 생명현상에 필수적일 가능성이 높으므로 본 연구결과는 유전자의 보존성과 함께 항생제의 대상이 되는 단백질 연구[2] 등 기능적 연계에 대한 기초 자료를 제공할 수 있을 것이다. 특히 항생제의 경우 진핵생물 혹은 원핵생물에서만 나타나는 유전자를 대상으로 선택성이 높은 항생제 또는 진균제 개발이 가능할 것이며 사람과 원핵생물 등에서 공통적인 보존유전자는 항생제의 부작용을 막기 위해 항생제 개발의 타겟에서 제외되는 것이 좋을 것이다.

References

1. Brenner, S. E. 2000. Target selection for structural genomics. *Nat Struct Biol* **7**(Suppl), 967-969.
2. Buysse, J. M. 2001. The role of genomics in antibacterial target discovery. *Curr Med Chem* **8**, 1713-1726.
3. Fraser, C. M., Eisen, J. A. and Salzberg, S. L. 2000. Microbial genome sequencing. *Nature* **406**, 799-803.
4. ftp://ftp.ncbi.nih.gov/pub/COG/KOG/
5. Gabaldon, T. and Huynen, M. A. 2003. Reconstruction of the proto-mitochondrial metabolism. *Science* **301**, 609-609.
6. http://www.genome.jp/kegg/kegg2.html
7. http://www.ncbi.nlm.nih.gov/COG/grace/shokog.cgi
8. Jordan, I. K., Wolf, Y. I. and Koonin, E. V. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol* **4**, 22.
9. 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.
10. Kimura, M. 1983. The neutral theory of molecular evolution. *Cambridge University Press.*
11. 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.
12. Lee, D.-G., Kim, C. M., Lee, E. U. and Lee, J. H. 2003. Genetic composition analysis of marine-origin euryarchaeota by using a COG algorithm. *J Life Sci* **13**, 298-307.
13. Mushegian, A. 1999. The minimal genome concept. *Curr Opin Genet* **9**, 709-714.
14. Nei, M. and Rooney, A. P. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* **39**, 121-152.
15. 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.
16. Tatusov, R. L., Koonin, E. V. and Lipman, D. L. 1997. A genomic perspective on protein families. *Science* **278**, 631-637.
17. Tekle, Y. I., Grant, J. R., Kovner, A. M., Townsend, J. P. and Katz, L. A. 2010. Identification of new molecular markers for assembling the eukaryotic tree of life. *Mol Phylogenet Evol* **55**, 1177-1182.
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. Warner, J. and McIntosh, K. 2009. How common are extra-ribosomal functions of ribosomal proteins? *Mol Cell* **34**, 3-11.
20. Zhou, X., Lin, Z. and Ma, H. 2010. Phylogenetic detection of numerous gene duplications shared by animals, fungi and plants. *Genome Biol* **11**, R38.

초록 : 원핵생물과 공통인 진핵생물의 보존적 유전자 탐색

이동근*
(신라대학교 의생명과학대학 제약공학과)

생물들에서 생명의 본질적 기능을 수행하는 단백질들의 종류와 보존성을 밝히기 위해 COG (Clusters of Orthologous Groups of proteins) 알고리즘을 이용하였다. 66종의 미생물에서 보존적인 63개의 ortholog 그룹들은 진핵생물 7종에서 104개의 ortholog들로 확산되었으며, 7종 모두의 핵에 보존적인 KOG (euKaryotic Orthologous Group)은 71개였다. 71개 중 단백질 합성에 관여하는 유전자들이 총 54개로 생명현상에서의 단백질의 중요성을 확인할 수 있었다. Distance value로 보존적 유전자가 생물종 사이에 나타내는 유전자 변이의 정도를 파악하니 'Translation initiation factor'인 KOG3403과 KOG3271 그리고 'Prolyl-tRNA synthetase' (KOG4163) 등이 높은 보존성을 보였다. 보존적 KOG들의 평균과 분산으로 유전체 분석을 수행하여 꼬마선충이 KOG 평균사이의 편차가 제일 커 유전자의 변이가 다양한 것을 알 수 있었다. 본 연구결과는 기초연구와 항생제 개발 등에 이용될 수 있을 것이다.