

## Uridylate kinase as a New Phylogenetic Molecule for Procaryotes

Dong-Geun Lee, Jin-Ok Lee, Jae-Hwa Lee

Department of Bioscience and Biotechnology, College of Engineering, Silla University

TEL: +82-51-999-5748, FAX: +82-51-999-5636

### Abstract

For the phylogenetic analysis of procaryotes, 16S rRNA gene has been used. In spite of it's common use, so high conservative of 16S rRNA gene limited resolving power, hence other molecule was applied in this study and the result was compared with that of 16S rRNA. COG (Clusters of Orthologous of protein) algorithm revealed that three COGs were only detected in 42 procaryotes ; transcription elongation factor (COG0195), bacterial DNA primase (COG0358) and uridylate kinase (COG0528). Uridylate kinase gene was selected owing to the similarity and one single copy number in each genome. Phylogenetic tree of 16S rRNA gene and uridylate kinase showed similarities and differences. Uridylate kinase may help the problem of very high conservative of 16S rRNA gene in phylogenetic analysis and it would help to access more accurate discrimination and phylogenetic analysis of bacteria.

### 서 론

원핵생물(procaryote) 분류에 이용되고 진화적인 관계를 나타내는 분자들의 조건은 비교대상 미생물에서 모두 발견될 것, 보존성이 아주 높을 것, 동시에 각 생물이 구분될 수 있는 변이를 나타낼 것, 그리고 크기가 적당할 것 등이다<sup>1)</sup>. 이러한 기준에 부합되며 분자생물학적 방법의 발달에 따라 현재는 16S rRNA 유전자 염기서열을 이용한 분류법이 널리 사용되고 있다<sup>2)</sup>. 하지만 16S rRNA 유전자의 너무 큰 보존성으로 진화적 경향을 파악할 때 생기는 제한된 해상력, 각 원핵생물마다 유전자의 수가 다른 것, 하나의 원핵생물이 갖는 16S rRNA 유전자들의 염기서열이 다른 경우 등으로 진화적 경향을 제시할 때 과대 혹은 과소 평가가 될 수 있다.

생화학적 분류법과 전체게놈의 염기서열에 기초한 분류법의 중간자적 위치에

있는 방법으로 COG (Clusters of Orthologous Groups of protein)를 이용한 분류도 있었다<sup>3)</sup>. 본 논문에서는 COG 알고리즘을 이용하여 새로운 분류분자들을 검색하고, 그 결과를 16S rRNA 유전자를 이용한 분류와의 비교를 목적으로 하였다.

## 재료 및 방법

### 재료

원핵생물 42종의 유전체 (microbial genome)는 National Center for Biotechnology Information (NCBI)의 공개 데이터베이스에서 추출하였다<sup>4)</sup>.

### 원핵생물의 보존적 유전자 선정 및 서열 확보

보존적 유전자 탐색 과정을 진핵생물인 효모를 포함하는 43종의 미생물과 42종의 원핵생물에 대하여 수행하여<sup>5)</sup> 원핵생물만 공통적으로 보유하는 COG를 선정하였다. 선정된 COG에 해당하는 핵산염기서열과 아미노산 서열은 Genbank에서 구하였고 16S rRNA 유전자의 염기서열은 RDP-II에서 확보하였다.

### 다중서열 정렬 및 유연관계 분석

선정된 COG의 염기서열과 16S rDNA의 염기서열들은 Clustal (ver. 1.64b) 프로그램을 이용하여 다중정렬을 수행하였다. Neighbor-joining method와 bootstrap method (n=1000)로 유전자 간의 상관관계를 분석하였다.

## 결과 및 고찰

### 원핵생물 한정의 보존적 유전자

원핵생물에서만 발견되는 3종류의 COG들은 COG0195, COG0358, COG0528로 각각 transcription elongation factor, bacterial DNA primase 그리고 uridylyate kinase의 기능을 갖는 것이었다.

### 유사도 및 보존성 비교

전체적으로 보면 핵산의 유사도보다 아미노산의 유사도가 높았고, 고세균 (archaeobacteria)은 COG0195가 진정세균 (eubacteria)은 COG0358이 핵산과 아미노산의 변이도가 큰 것을 알 수 있었다.

### 새로운 분류 분자

16S rRNA 유전자를 이용한 분류의 단점을 보완하기 위하여 다른 분자들이 이용되어 왔다<sup>7)</sup>. Table 1과 같이 COG0195와 COG0358의 경우는 한 생물내에 중복되는 유전자가 관찰되는 반면 COG0528의 경우 하나의 미생물에 하나의 유전자만 존재하므로 각 미생물은 유전자의 중복성에 의한 문제를 피할 수 있다는 것을 알

수 있었고 세균 각 종에 대해 정량적인 비교와 진화적 관점의 파악이 유리하다고 사료되었다.

Table 1. Organisms, conserved COGs and nucleotide length of conserved COGs in all procaryotes tested. Multiple number represents their multiple paralogs. (Ar = Archaeobacteria, Eu = Eubacteria)

	Organisms	Abbreviation	COG0195	COG0358	COG0528
Ar	<i>Aeropyrum pernix</i>	Ape	435	1281	741
	<i>Archaeoglobus fulgidus</i>	Afu	420	1200	660
	<i>Halobacterium sp. NRC-1</i>	Hbs	420	1353	726
	<i>Methanobacterium thermoautotrophicum</i>	Mth	432, 552	1146	675
	<i>Methanococcus jannaschii</i>	Mja	552	1275	723
	<i>Pyrococcus abyssi</i>	Pab	189, 438, 513	1344	678
	<i>Pyrococcus horikoshii</i>	Pho	189, 438, 513	1344	681
	<i>Thermoplasma acidophilum</i>	Tac	429	1305	690
	<i>Thermoplasma volcanium</i>	Tvo	429	1293	684
Eu	<i>Aquifex aeolicus</i>	Aae	1014	1497	723
	<i>Synechocystis</i>	Syn	1377	1908	783
	<i>Bacillus halodurans</i>	Bha	1149	1800	720
	<i>Bacillus subtilis</i>	Bsu	1116	1812	723
	<i>Lactococcus lactis</i>	Lla	1149	1914	717
	<i>Mycobacterium leprae</i>	Mle	1044	1929	840
	<i>Mycobacterium tuberculosis</i>	Mtu	1044	1920	786
	<i>Mycoplasma genitalium</i>	Mge	1596	657, 1824	732
	<i>Mycoplasma pneumoniae</i>	Mpn	1623	639, 1863	708
	<i>Streptococcus pyogenes</i>	Spy	1158	1815	729
	<i>Ureaplasma urealyticum</i>	Uur	1377	1926	708
	<i>Chlamydia pneumoniae</i>	Cpn	1305	1773	747
	<i>Chlamydia trachomatis</i>	Ctr	1305	1788	738
	<i>Buchnera sp. APS</i>	Buc	1491	1734	729
	<i>Campylobacter jejuni</i>	Cje	1089	1818	720
	<i>Caulobacter crescentus</i>	Ccr	1647	1932	741
	<i>Escherichia coli K12</i>	Eco	1488	1746	726
	<i>Escherichia coli O157</i>	EcZ	1602	1746	726
	<i>Haemophilus influenzae</i>	Hin	1488	1782	714
	<i>Helicobacter pylori 26695</i>	Hpy	1188	1680	723
	<i>Helicobacter pylori J99</i>	jHp	1188	1680	723
	<i>Mesorhizobium loti</i>	Mlo	1596	1959	744
	<i>Neisseria meningitidis MC58</i>	Nme	1503	1773	720
	<i>Neisseria meningitidis Z2491</i>	NmA	1518	1773	720
	<i>Pasteurella multocida</i>	Pmu	1485	1749	729
	<i>Pseudomonas aeruginosa</i>	Pae	1482	1995	738
	<i>Rickettsia prowazekii</i>	Rpr	1512	1851	729
	<i>Vibrio cholerae</i>	Vch	1488	1764	732
	<i>Xylella fastidiosa</i>	Xfa	1512	1053, 1053, 1734	744
	<i>Borrelia burgdorferi</i>	Bbu	1449	1572	690
	<i>Treponema pallidum</i>	Tpa	1458	1818	756
	<i>Thermotoga maritima</i>	Tma	1035	1698	696
<i>Deinococcus radiodurans</i>	Dra	1206	1716	738	

Fig. 1은 42종의 원핵생물을 16S rRNA 유전자를 이용한 분류와 COG0528인 uridylate kinase의 유전자를 이용한 분류를 나타내고 있다. 두 분류의 공통점은 같은 속 (genus)에 속하는 세균들은 같은 위치에 분포, 고세균 내부의 높은 응집성, 고세균 그룹이 진정세균의 Firmicutes 그룹 일부와 상관관계가 상대적으로 높은 것 등이었고 차이점은 Firmicutes에 속하는 세균들의 분산, 기타 진정세균으로 분류하였던 세균들의 분산, proteobacteria의 alpha와 epsilon 그룹의 분류도에서의 위치, gamma-Proteobacteria 일부가 고세균과 높은 유연관계 등이었다. COG0528인 uridylate kinase의 경우 16S rRNA 유전자처럼 보존성이 아주 높지는 않아 16S rRNA 유전자의 아주 높은 보존성에 의해서 생기는 유연관계의 과소/과대평가를 줄일 수 있을 것이며 16S rRNA와 상호 보완적인 결과로 원핵생물의 정확한 분류에 기여할 수 있을 것으로 사료되었다<sup>6)</sup>.

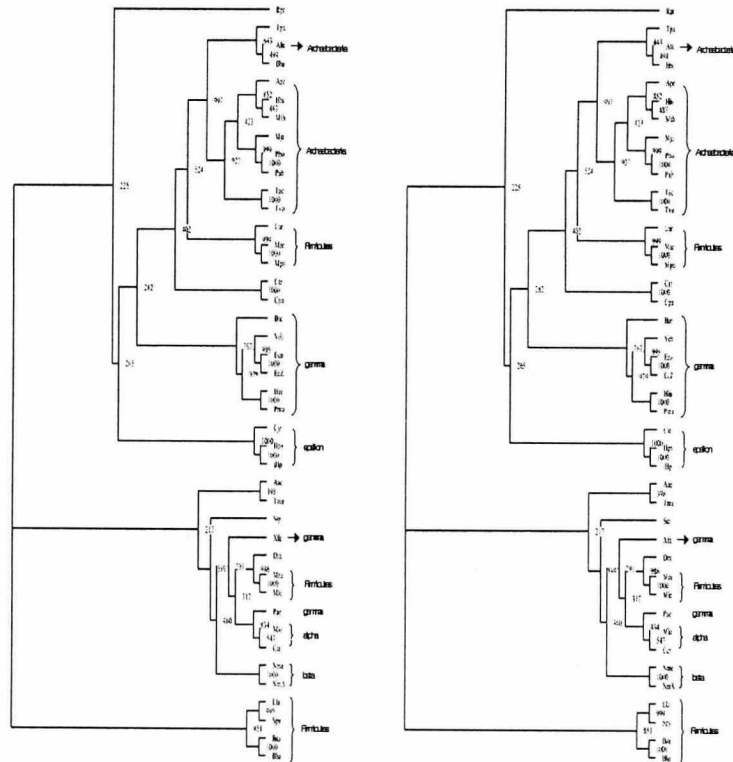


Fig. 1 Comparison of the phylogenetic trees of 42 prokaryotes obtained from neighbor-joining analysis of either 16S rRNA gene sequence (A) and COG0528 gene sequence (B). Bootstrap values at each node are expressed as a number over 1000 trials. *Helicobacter pylori* J99 (jHp) was omitted in 16S rRNA gene owing to the their short length. Terminal Branches have been extended for clarity and their length is therefore not meaningful.

## 요약

원핵생물 (prokaryote)의 분류에 16S rRNA 유전자가 많이 이용되어 있으나 제한된 해상력과 유전자의 수에 차이가 있는 등의 문제가 있어 이를 보완할 수 있는 새로운 생체분자를 찾고 그 분류 결과를 16S rRNA의 결과와 비교하였다. COG (clusters of orthologous of protein) 알고리즘으로 42종의 원핵생물 (prokaryote) 에서만 발견되는 transcription elongation factor (COG0195), bacterial DNA primase (COG0358) 그리고 uridylate kinase (COG0528)를 구하였다. 이중 유사도와 유전자 수를 바탕으로 새로운 분류의 키로 uridylate kinase를 설정하여 분석한 결과 16S rRNA 유전자 결과와 유사점과 차이점을 보여, uridylate kinase를 이용한 분류가 16S rRNA 유전자를 이용한 분류의 문제점을 보완하여 원핵생물의 정확한 분류에 기여할 수 있을 것으로 사료되었다.

## 참고문헌

1. Kimura, M. 1990. Prokaryote Systematics: The Evolution of a Science. Prokaryotes 2nd Edition. pp. 3-18.
2. Woese, C. R. Bacterial evolution. (1987) *Microbiology Review* **51**, 221-271.
3. Lee, D.-G., H.-Y. Kang, S. Kim, S.-H. Lee, C.-M. Kim, S.-J. Kim and J.-H. Lee. Classification of Archaeobacteria and Bacteria using a gene content tree approach. (2003) *Korean J. Biotechnol. Bioeng.* **18**, 39-44.
4. <ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria>
5. Lee, D.-G., H.-Y. Kang, J.-H. Lee and C.-M. Kim. Detection of Conserved Genes in *Proteobacteria* by using a COG Algorithm. (2003) *Korean J. Biotechnol. Bioeng.* **17**, 560-565.
6. Wilson, K., M. A. Schembri, P. D. Baker, and C. P. Saint. Molecular Characterization of the Toxic Cyanobacterium *Cylindrospermopsis raciborskii* and Design of a Species-Specific PCR. (2000) *Applied and Environmental Microbiology* **66**, 332-338.