

COG 알고리즘을 통한 해양성 Euryarchaeota의 유전적 조성 분석

이동근^{1,2} · 김철민^{3,4} · 이은열⁵ · 이재화^{1,2,*}

¹신라대학교 공과대학 생명공학과
²신라대학교 마린바이오산업화지원센터
³부산대학교 유전공학연구소부설 생물정보학센터
⁴부산대학교 의과대학부설 부산지능센터
⁵경성대학교 공과대학 식품공학과

Genetic Composition Analysis of Marine-Origin Euryarchaeota by using a COG Algorithm

Dong-Geun Lee^{1,2}, Cheol-Min Kim^{3,4}, Eun Yeol Lee⁵ and Jae-Hwa Lee^{1,2,*}

¹Department of Bioscience and Biotechnology, Silla University, Busan 617-736, Korea

²Marine Biotechnology Center for Bio-Functional Material Industries, Silla University, Busan 617-736, Korea

³Pusan Bioinformatics and Biocomplexity Research Center, Pusan National University, Busan 609-735, Korea

⁴Busan Genome Center, College of Medicine, Pusan National University, Busan 609-735, Korea

⁵Department of Food Science and Technology, College of Engineering, Kyungsoong University, Busan 608-736, Korea

Abstract

To figure out the conserved genes and newly added genes at each phylogenetic level of *Archaea*, COG (clusters of orthologous groups of proteins) algorithm was applied. The number of conserved genes within 9 species of *Archaea* was 340 and that of 8 species of Euryarchaeota was 388. Many of conserved 265 COGs, which are specific to *Archaea* and absent in *Bacteria* and *S. cerevisiae*, were concerned with 'information storage and processing' (94 COG, 35.5%) and 'metabolism' (82 COG, 30.9%). COGs related to these functions were assumed as highly conserved and permit peculiar life form to *Archaea*. It seemed that there was some difference in 'nucleotide transport and metabolism' and there was little difference in 'information storage and processing' between Euryarchaeota and Crenarchaeota. Marine-origin Euryarchaeota showed different conserved COGs with terrestrial Euryarchaeota. Conserved COGs, related to carbohydrate transport and metabolism and others, were different between marine- and terrestrial-origin Euryarchaeota. Hence it was assumed that their physiology might be different. This study may help to understand the origin and conserved genes at each phylogenetic level of marine-origin Euryarchaeota and may help in the mining of useful genes in marine *Archaea* as Manco *et al.* (*Arch. Biochem. Biophys.* **373**, 182 (2000)).

Key words – clusters of orthologous groups of proteins (COG), conserved gene, *Archaea*, Euryarchaeota

*To whom all correspondence should be addressed
Tel : 051-999-5748, Fax : 051-999-5636
E-mail : jhalee@silla.ac.kr

서론

지구상에는 다양한 지구 환경에 적응하며 많은 종류의 미생물들이 살고 있는데 크게 원핵생물 (prokaryote)과 진핵생물 (eukaryote)로 나눌 수 있다. 한편 16S rRNA 유전자 분석을 통하여 고세균 (Archaea) 그룹이 대두되어 원핵생물은 다시 고세균 (Archaea)과 세균 (Bacteria)으로 나눌 수 있다[36].

해양미생물은 육지에서 공급되는 강물 등에 의해 육상에서 관찰되는 미생물과 호염성세균 등 극한미생물이라 할 수 있는 해양특이적인 미생물들로 구성되어 있다. 심해수에서 해양세균 18종을 분리하였는데 그 들은 육상에서도 발견되는 *Flavobacterium*, *Micrococcus*, *Pseudomonas* 등이 포함되어 있었다[20]. 한편 해양미생물은 육지에 존재하는 미생물과 여러 가지 면에서 생태적, 생리적 특징이 달라 독특한 생리활성을 나타내는 것들이 많다[2,27]. 해양 유래의 고세균은 생태적 특이성에 기인하여, 세균과는 다른 성질의 효소[23]를 생산하는 등 응용가능성이 높아 산업적 가치도 높다[31].

Orthologous는 공통의 조상으로부터 종분화되어 서로 다른 종에 있는 유전자들의 집합으로 정의하며, paralogous는 한 유전체내에서 복사 (duplication)로 생성된 유전자들을 총칭하는 용어이다[33]. COG (Clusters of Orthologous Groups of protein)는 orthologous들에서 유래된 단백질의 집합으로 대개 유사한 구조와 기능을 갖는 것으로 알려져 있다[4,11,32]. 각 COG는 적어도 3가지 이상의 계보 (lineage)에서 유래된 paralog 그룹 혹은 개별의 단백질들로 구성되어 있어 하나의 공통조상유전자 (ancient conserved domain)에 해당하는 것으로 간주할 수 있다[33].

계통발생학적 계보 (phylogenetic lineage)에 있어 생명체들이 나타내는 각 분류단위의 독특한 생명현상과 모든 생명체가 공통적으로 나타내는 필수기능 (housekeeping function)에 대한 이해는 생명 자체의 이해에 대한 핵심이라 할 것이다[18,33].

TIGR (The Institute for Genomic Research)에서 *Haemophilus influenzae* 계놈의 염기 서열을 밝힌 이후 최근에는 원핵생물인 고세균, 세균과 진핵생물인 *Saccharomyces cerevisiae*의 염기서열이 보고되면서 여러 미생물 간의 진화학적인 그리고 비교유전학적인 분석이 가능하게 되었다[5,15,21,25]. 고세균은 복제, 전사, 번역에 관련된

기구들과 rRNA 구조에 따라서 다른 생물그룹과 대별되는데[36] 형태학적인 특성은 Gram 양성세균과 다른 것이 없으며[9] 단백질 서열에서 보존적인 서열을 분석하면 Gram 양성구균에서 고세균이 진화한 것으로 보고하였다[10]. 한편 translation elongation factor와 H⁺-ATPase는 진핵생물의 것과 유사하다고 알려져 왔다[8]. 고세균에서 예상되는 ORF (open reading frame)의 절반 정도는 아직 기능을 알 수 없어 이러한 잠재적 유전자 (potential gene)에 대한 기능을 밝히는 것은 생물학에서 중요한 역할을 한다고 할 것이다[26,35]. 실제로 COG를 이용하여 심해저 미생물인 *Archaeoglobus fulgidus*에서 유용한 유전자를 찾은 사례도 있다[22].

Makarova[21] 등은 4개의 고세균을 분석하여 Euryarchaeota의 유전자 중 일부는 bacterial homolog와 유사성이 높은 것을 발견하고 이들은 미생물간에 유동적인 것으로 간주하고 수평적 유전자 이동 (horizontal gene transfer)과 계보특이적 결실 (lineage-specific gene loss)이 Euryarchaeota의 진화에 중요하다고 하였다. 하지만 분석 개수가 4개이고 Wolf[37] 등은 10종의 고세균을 분석하였지만 계통수에 집중하여 유전체를 비교하는 것과는 거리가 있다고 할 것이다.

본 연구에서는 생물정보학적인 접근방법으로 염기서열이 알려진 미생물 중 해양성미생물에 속하는 고세균 9종이 가진 유전체를 비교하여 공통적으로 유지되고 있는 보존적 유전자의 존재를 확인하고자 하였다. COG 알고리즘 [15, 18]을 이용하여 각 분류단계별로 추가되는 보존적 유전자를 분석하였고 해양성 Euryarchaeota와 육지성 Euryarchaeota의 유전적 조성을 비교하였다.

재료 및 방법

재료

분석에 이용된 유전체는 NCBI (National Center for Biotechnology Information)의 공개 서버로부터 추출하였다[6]. 미생물 유전자의 유사성에 관한 자료는 COGs에서 정리된 자료를 이용하였는데[13] 이들은 2003년 2월 현재 43종의 미생물 유전체를 orthologous group으로 분류하여, 총 74,059개의 유전자들을 3,307개의 COG 그룹으로 분류해 놓았다[13]. 분석대상 원핵생물은 고세균이 9종, 세균

중 Firmicutes가 9종, Proteobacteria가 16종, 기타 8종 이었다 [15,18]. Table 1은 본 연구에서 분석한 9종의 고세균에 대한 COG 자료를 나타내고 있다.

현재 계놈의 염기서열이 밝혀진 9종의 고세균[10]중 *Aeropyrum pernix*[29], *Archaeoglobus fulgidus*[31], *Halobacterium sp. NRC-1*, *Methanococcus jannaschii*[14], *Pyrococcus abyssi*[12], *Pyrococcus horikoshii*[7]등은 해양성 고세균이라 할 수 있다. 그리고 *Methanobacterium thermoautotrophicum*은 폐수 슬러지에서[38], *Thermoplasma acidophilum*은 석탄 찌꺼기에서[1], *Thermoplasma volcanium*은 육지의 화산분화구 인근에서[30] 각각 분리된 것으로 확인되어 이들을 육지성 고세균으로 간주하였다.

보존적 유전자 탐색

분석 방법은 강 등[15]의 방법을 이용하였다. 여러 분석 단계 중에서 CLUSTAL 프로그램을 이용한 다중서열비교를 통해 distance value를 계산하였고[16,24], 자료의 분석과 정리에는 perl language (Practical Extraction and Report Language)를 사용하였다.

Fig. 1은 본 연구에서 사용된 미생물 유전체의 분석 작업 순서도이다. NCBI의 공개 데이터베이스로부터 전체 유전체가 공개된 75종의 미생물 자료를 수집하여[3,6,28], 이

들 유전체로만 구성된 로컬 데이터베이스를 제작하였다 [15]. 그리고 COGs 데이터베이스의 공개파일전송 (ftp) 서버로부터 원핵생물에 대한 orthologous를 확보하였다[6]. 이후 각 그룹별로 보존적 유전자의 항목을 배열화하였다. COGs 데이터베이스의 자료를 종 (species)과 orthologous에 따라 2차원으로 재정렬하여 개별 orthologous를 종을 기준으로 분류하였다. 모든 원핵생물에서 공통적으로 관찰되는 COG들을 ancestral gene에서 유래된 보존적 유전자로 간주하였다[15].

위의 보존적 유전자 탐색 과정을 42종의 원핵생물[18]과 9종의 고세균 그리고 8종의 Euryarchaeota에 대하여 적용하여 각 분류단계별로 추가되는 보존적 유전자를 탐색하였다.

결과 및 토의

세균 (Bacteria)에 속하는 단백세균 (Proteobacteria)의 분류 단계별로 추가되는 COG를 비교한 보고가 있었다[18]. 본 보고에서는 세균이 아닌 고세균에 속하는 해양성미생물인 Euryarchaeota의 유전적 조성을 분류 단계별로 COG를 이용하여 비교하였다. 비교대상으로 하는 계놈을 43종의 미생물, 42종의 원핵생물 (prokaryote), 9종의 고세균 (Archaea),

Table 1. Studied genomes derived from COGs database, number and percentage of conserved genes for 9 species of Archaea.

Phylogenetic Group	Organism	Number of protein	Number of orthologous ¹⁾	Number of conserved genes	Percentage of conserved gene (%)
Crenarchaeota	<i>Aeropyrum pernix</i>	1841	1,178	340 ²⁾	28.86
	<i>Archaeoglobus fulgidus</i>	2420	1,872		20.73
	<i>Halobacterium sp. NRC-1</i>	2605	1,701		22.81
	<i>Methanobacterium thermoautotrophicum</i>	1873	1,388		27.95
	Euryarchaeota	<i>Methanococcus jannaschii</i>	1786	1,330	386 ³⁾
<i>Pyrococcus abyssi</i>		1768	1,456		29.65
<i>Pyrococcus horikoshii</i>		1800	1,378		28.16
<i>Thermoplasma acidophilum</i>		1482	1,230		31.54
<i>Thermoplasma volcanium</i>		1499	1,243		31.21

¹⁾Number of orthologous was determined after compared 43 genomes of Archaea, Bacteria and *S. cerevisiae*.

²⁾Conserved genes in the phylogenetic level of Archaea.

³⁾Conserved genes in the phylogenetic level of Euryarchaeota.

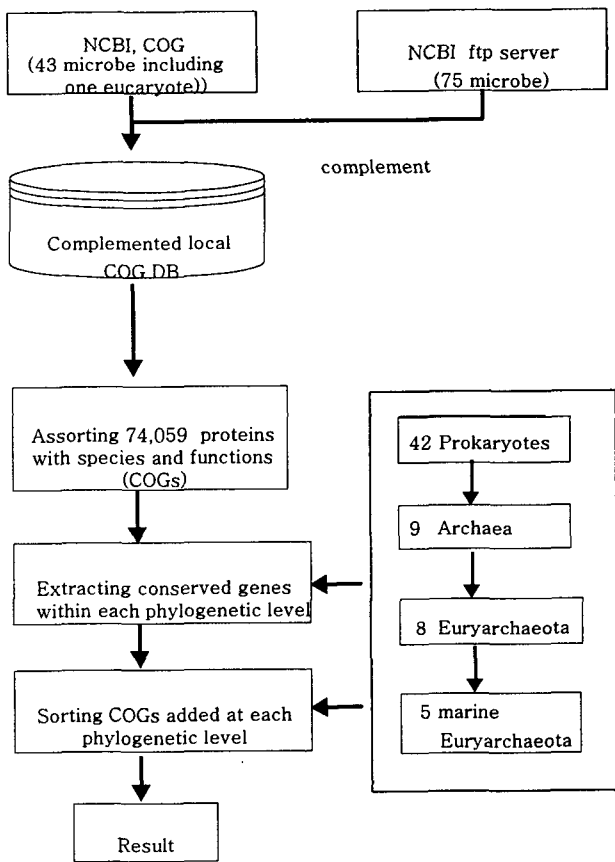


Fig. 1. Flow chart to determine conserved genes within marine Euryarchaeota using microbial genomes. Forty-three microbes cover 1 *Saccharomyces cerevisiae*, 9 archaeobacteria, 16 proteobacteria, and 17 other eubacteria (see reference 11 and 14).

8종의 Euryarchaeota, 5종의 marine Euryarchaeota 순으로 한정하면서 보존적 유전자의 수가 늘어나고 각 분류단계별로 특이적인 COG들이 추가로 보존적이라는 것을 확인할 수 있었다. 이는 비교하는 계층의 수가 줄어들면서 공통적으로 보존적인 유전자의 수가 증가될 가능성이 높다는 점과 각 분류단계별로 특이적인 COG들의 추가에 의한 결과라고 할 것이다[18].

Table 1에 9종의 고세균이 나타내는 전체 단백질의 수, orthologous 수, 보존적 유전자의 수와 비율을 나타내었다. 강[15]등은 원핵생물 42종과 진핵생물인 효모 1종 등 43종의 미생물이 공통적으로 소유하는 보존적 COG를 검색하여 총 72종의 유전자가 보존적이고, 전체 보존적 유전자 중 52개 (72.2%)가 단백질의 합성에 관련된 것을 밝혔

다. 또한 42종의 원핵생물에서는 3종의 COG가 추가되어 총 75종의 COG가 보존적인 것을 밝혔다[18]. 9종의 고세균이 공통적으로 보유하는 보존적 유전자는 총 340개로 나타났다 (Table 1). *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Archaeoglobus fulgidus*, *Pyrococcus horikoshii* 등 4종류의 고세균을 비교한 보고[21]에서 543개의 COG가 공통적이라고 보고하였다. 본 연구에서는 Crenarchaeota를 포함한 9종의 고세균이 비교되었는데 계층의 수가 늘어나면서 비교대상 계층 모두에 존재하는 공통적 COG의 개수는 줄어든 결과라고 사료되었다 [18].

Euryarchaeota에 속하는 8종의 미생물에서는 보존적 COG가 46개 증가하여 총 386개의 COG가 보존적인 것으로 파악되었고 Euryarchaeota 각 종이 보유하는 orthologous에 대한 보존적 유전자의 비율은 20.73~31.54%로 나타나 (Table 1) Makarova[21] 등이 보고 한 31%~35%와 차이를 보였다. 이러한 결과는 Crenarchaeota에 의한 영향과 전술한 비교대상 계층의 수에 따른 영향을 나타내는 것으로 사료되었다. 보존적 유전자의 비율은 공통조상유전체의 진화적 안정성으로 생각할 수 있는데[21] 단백질세균 (*Proteobacteria*) 수준에서는 총 246개의 보존적 이었고 *Rickettsia prowazekii*, *Buchnera sp.* APS를 제외하면 6.3~22.1%로 나타났다. Gupta[10]는 Gram 양성 세균에서 고세균과 Gram 음성 세균이 진화한 것으로 보고하였다. Gram 양성 세균 그룹에서는 총 468개의 COG가 보존적이었으며 16.26% (*Bacillus halodurans*)와 41.27% (*Mycobacterium leprae*) 사이의 보존적 유전자 비율을 나타내어 진화적 안정성은 고세균이 높은 것으로 사료되었다.

Fig. 2는 Euryarchaeota의 분류 단계별로 보존적인 COG의 기능과 개수를 나타내고 있다. 세균이나 진핵생물인 *S. cerevisiae*에는 없고 고세균에만 보존적인 265개의 COG를 기능분류별 (functional categories)로 보면 (Fig. 2) 유전정보의 보존과 처리에 관여하는 COG가 총 94개 (35.5%)이고 세포작용에 관여하는 COG가 총 29개 (10.9%)였으며 대사에 관여하는 COG가 총 82개 (30.9%)로 전체 보존적 COG중 유전정보와 물질대사와 관여하는 COG의 보존성이 높은 것으로 나타났다. 각각의 기능계열을 보면 전사관련 J계열 (59개), 조효소 대사관련 H계열 (23개), 기능을 알 수 없는 R계열 (38개)의 보존비율이 높은 것으로 드러났다. 이러한 결과로부터 고세균이 다른 생물군과 구

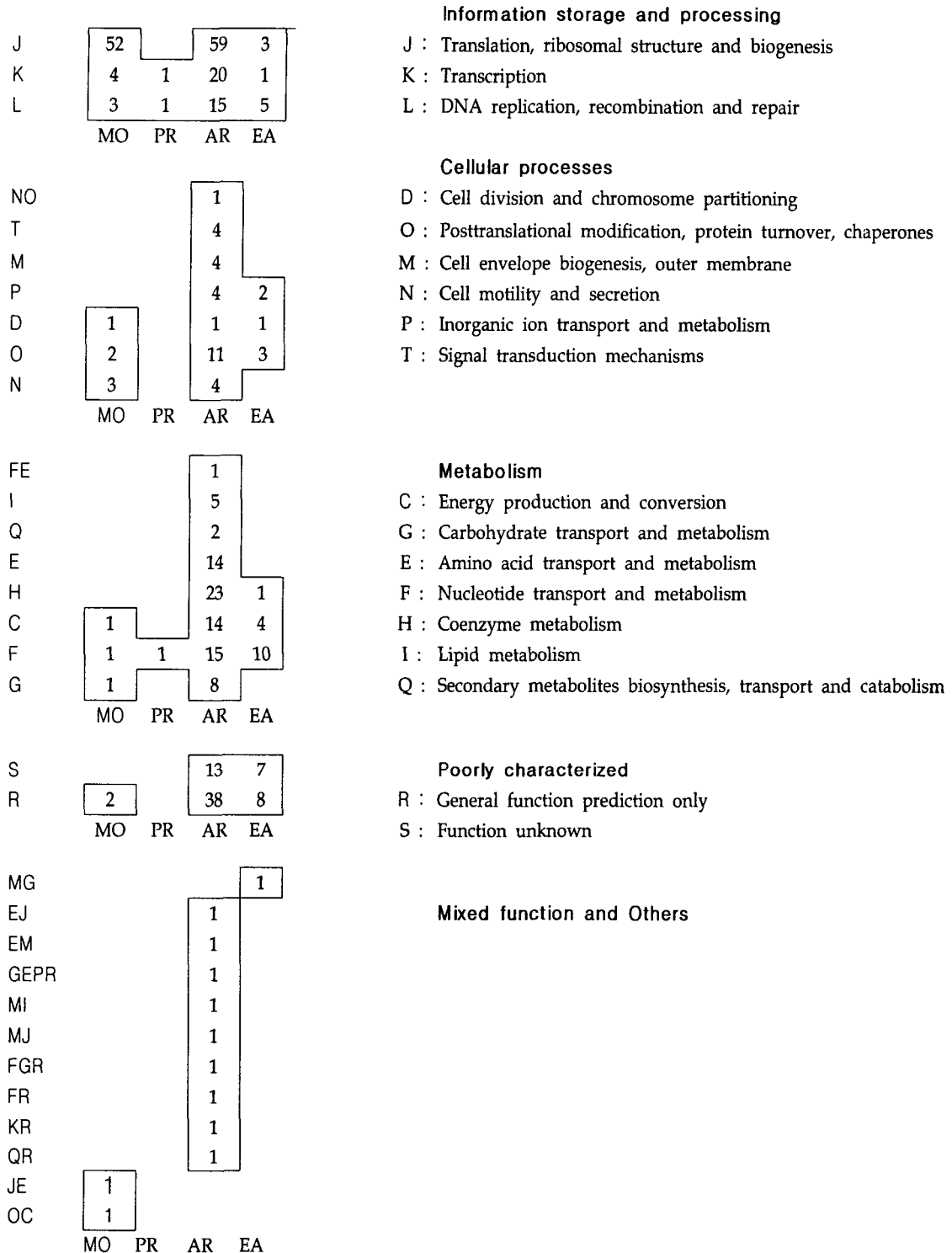


Fig. 2. Conserved functions and the number of new COGs added at each phylogenetic group. Definition of functions were followed as COGs (<http://www.ncbi.nlm.nih.gov/COG/>). Abbreviations in horizontal form represent phylogenetic group (MO; 42 prokaryotes and 1 eukaryote, PR; 42 prokaryotes, AR; 9 *Archaea*, EA; 8 *Euryarchaeota*).

별되는 독특한 생명체계를 이루고 있는 것을 사료할 수 있었다. Makarova[21] 등은 비교대상 계통에 모두 존재하는 유전자들을 evolutionary stable core로 나머지 유전자들은 variable shell로 명명하였다. 계통의 복제와 발현에 필요한 유전자들의 보존도가 높은 것[21]은 본 연구 결과와 유사하였다. R계열은 현재까지 기능이 정확하게 밝혀지지 않은 것들이다. Fig. 2에서 알 수 있듯 물질대사에서 보존적 COG는 조효소 대사관련 H계열 (23개), 핵산 대사관련 F계열 (15개) 그리고 에너지 대사관련 C계열 (14개)로 나타났다. C계열 COG를 보면 Makarova[21] 등은 중속영양생활을 하는 *Pyrococcus horikoshii*를 비교 대상에 포함하지 않으면 120개 정도의 유전자가 보존적이었고 그렇지 않은 경우는 50개 정도의 에너지 대사관련 유전자가 보존적인 것으로 밝혔다. 본 연구에서는 화학합성, 광합성, 중속영양으로 에너지 대사를 하는 것들이 포함되어 14개의 유전자만이 보존적인 것으로 나타났다. 이[19] 등은 유전자보유 계통수를 분석하면서 고세균에서 halophile인 *Halobacterium sp. NRC-1*가 다른 고세균 그룹과 독립적으로 존재하는 것은 서식지에 따른 유전적 조성의 차이로 보고하였는데 서식지에 따라 에너지 대사 기구 등이 달라 보존적 COG의 수가 작은 것으로 사료되었다.

Koonin[17] 등은 고세균에서 복제, 전사, 해독에 관련된 단백질들은 진핵생물의 것과 유사하고 물질대사, 세포분열, 세포벽합성에 관련된 단백질들은 세균의 것과 유사한 것으로 보고하였다. 하지만 이[18] 등은 원핵생물 (prokaryote) 수준에서 보존적인 유전자의 수가 많지 않은 것으로 드러났다. 이는 이[18] 등은 42개의 원핵생물 모두에 보존적인 COG를 추출한 결과로 사료되며, 본 연구에서도 비교대상 원핵생물의 수를 줄이면 보존적 COG의 수가 증가하는 것을 확인할 수 있었다 (데이터 생략).

Fig. 2에서는 Crenarchaeota인 *Aeropyrum permix*와 Euryarchaeota가 차이를 보이는 COG들을 알 수 있는데 유전정보의 저장과 처리에 관여하는 J, K, L 계열을 보면 Euryarchaeota 수준에서는 고세균 수준에 비해 추가된 COG의 수가 적은 것으로 나타났다. 이는 유전정보의 저장과 처리에 관여하는 많은 유전자들의 보존도가 고세균 수준에서 높은 것으로 해석할 수 있었다. 무기이온의 대사 (P 계열), 세포분열과 염색체 분배 (D 계열) 그리고 전사후 수식 (posttranslational modification, O 계열) 등에 관여하는

COG가 Euryarchaeota 수준에서 일부 추가되는 것으로 나타났다. 물질대사에서는 핵산대사에 관여하는 F계열 유전자가 10개 추가되었다. 이런 결과를 바탕으로 Euryarchaeota를 Crenarchaeota와 비교하면 핵산대사에서는 상당한 차이를 보이며 유전정보의 저장과 처리에서는 큰 차이가 없는 것으로 판단되어졌다. 에너지 대사관련 C계열을 보면 고세균 수준에서는 Pyruvate:ferredoxin oxidoreductase 관련 alpha, beta, gamma subunit (COG0674,1014,1015)와 Archaeal/vacuolar-type H⁺-ATPase의 각 subunit A, B, D, E, I (COG1155,1156,1394,1390,1269) 등이 보존적이었고 Euryarchaeota 수준에서는 Archaeal/vacuolar-type H⁺-ATPase의 subunit C, F, H (COG1527,1436,2811)와 Ferredoxin 3 (COG1146)이 보존적이었다. 생명현상에 필수적인 ATP 생성에 있어 필요한 유전자의 일부를 Crenarchaeota와 공통적으로 갖고 Euryarchaeota 특이적인 유전자도 보유하고 있음을 확인할 수 있었다.

Fig. 3은 Euryarchaeota를 5종의 해양성 고세균 (MA)과 3종의 육지성 고세균 (TA)으로 나누어서 추가되는 COG들을 비교한 결과이다. 해양성 Euryarchaeota 5종 (species)에서는 추가적으로 99개의 COG가 보존적인 것으로 확인되었다. 보존적 COG를 기능분류면에서 보면 어떤 한 기능에 관련된 COG들의 수가 높지 않은 것을 알 수 있었다. 그리고 육지성 Euryarchaeota 3종에 공통적인 COG는 총 180개가 추가 되었으며 기능분류별 종류가 해양성 Euryarchaeota와 달랐고 물질대사 관련 COG의 경우 육지성이 해양성보다 다양한 것을 알 수 있었다. 이러한 차이는 비교대상 계통의 수가 다른 것에 의한 영향도 있을 것으로 사료되었다. 기능별로 추가되는 COG의 수도 H계열 (조효소 대사관련)이 23개, E계열 (아미노산 대사관련)이 19개, C계열 (핵산 대사관련)이 17개 등 육지성 Euryarchaeota들은 서로간의 보존적 COG의 수가 많은 것을 알 수 있었다.

탄수화물의 전달과 대사에 관여하는 G계열 COG는 43종의 미생물에서 9종의 Euryarchaeota 수준까지는 보존적인 것이 하나도 없었지만 육지성과 해양성 Euryarchaeota로 구분되면서 각각 6개와 4개의 COG가 보존적인 것으로 나타났다. 그리고 복합적인 기능을 보이는 것으로 사료되는 보존적 COG들이 육지성과 해양성에서 서로 다른 것을 확인할 수 있었다. 이것으로 보아 육지성과 해양성 Euryarchaeota는 탄수화물대사 등을 비롯한 생리적 측면에서

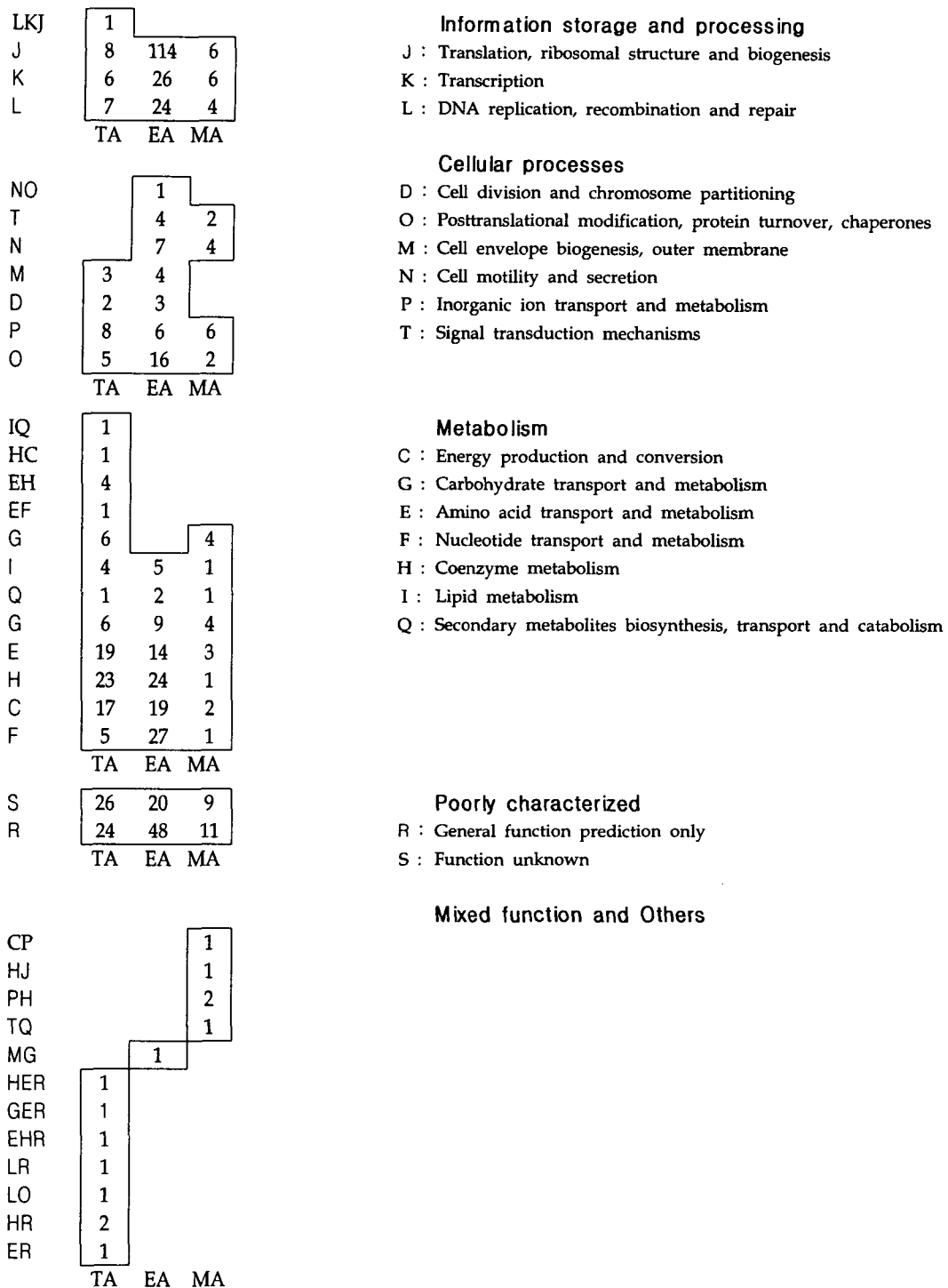


Fig. 3. Conserved functions and the number of new COGs added at Euryarchaeota, marine Euryarchaeota, and terrestrial Euryarchaeota. Definition of functions were followed as COGs (<http://www.ncbi.nlm.nih.gov/COG/>). Abbreviations in horizontal form represent 8 Euryarchaeota (EA), 5 marine Euryarchaeota (MA), and 3 terrestrial Euryarchaeota (TA). COGs of mixed function and others added at microorganism and *Archaea* level were omitted. They are EJ, EM, FGR, FR, GEPR, JE, KR, MI, MG, MJ, OC, and QR (see Fig. 2).

서로 차이가 있을 가능성이 높을 것으로 사료되었다 (Fig. 3).

본 연구를 통하여 해양 극한미생물인 고세균의 기원과 진화적 단계에 따른 유전적 조성을 분석하고 해양성 Euryarchaeota와 육지성 Euryarchaeota의 유전적 조성을 비교하였다. 또한 COG의 응용 장점은 기존의 자료와 유전자 서열만으로 미지의 유전자에 대한 기능추측이 가능하다는 것으로 향후 해양미생물에 대한 적용과 응용에 이용될 가능성을 높이는 계기가 되었다고 사료되었다.

요 약

고세균 (Archaea)의 보존적 유전자를 파악하고 각 분류 단계별로 추가되는 보존적 유전자를 밝히기 위해 그리고 해양성 Euryarchaeota와 육지성 Euryarchaeota의 유전자 조성을 비교하기 위해 COG (clusters of orthologous groups of proteins) 알고리즘을 이용하였다. 총 9종의 고세균이 공통적으로 보유하는 보존적 유전자는 340개로 나타났고 8종의 Euryarchaeota는 388개의 유전자가 보존적이었다. Euryarchaeota 각 종이 보유하는 orthologous에 대한 보존적 유전자의 비율은 20.73~31.54%로 나타났다. 세균과 *S. cerevisiae*에는 없고 고세균 수준에서만 공통적인 265개 COG의 조성은 유전정보의 보존과 처리에 관여하는 COG가 94개 (35.5%)이고 대사에 관여하는 COG가 82개 (30.9%)로 유전정보와 물질대사와 관여하는 COG의 보존성이 높은 것으로 나타나 고세균이 독특한 생명체계를 이루고 있는 것으로 사료되었다. Euryarchaeota를 Crenarchaeota와 비교하면 핵산대사에서는 상당한 차이를 보이며 유전정보의 저장과 처리에서는 큰 차이가 없는 것으로 판단되었다. 해양성 Euryarchaeota의 보존적 COG는 기능분류별 종류가 육지성 Euryarchaeota와 달랐고 물질대사 관련 COG의 경우 육지성이 해양성보다 다양한 것을 알 수 있었다. 그리고 육지성과 해양성 Euryarchaeota는 탄수화물대사 등을 비롯한 생리적 측면에서 서로 차이가 있을 가능성이 높을 것으로 사료되었다. 본 연구는 해양 극한미생물인 해양성 Euryarchaeota의 기원과 분류단계에 따른 보존적 유전자를 파악하는데 도움을 줄뿐만 아니라 향후 해양미생물 등의 유용유전자 탐색 등에서도 Manco (*Arch. Biochem.*

Biophys. 373, 182 (2000)) 등의 보고와 같이 충분한 연구가치가 있는 것으로 사료되었다.

참 고 문 헌

1. Darland, G., T. D. Brock, W. Samsonoff and S. F. Conti. 1970. Athermophilic acidophilic mycoplasma isolated from a coal refuse pile. *Science* 170, 1416-1418.
2. Dib R., J. M. Chobert, M. Dalgalarondo, G. Barbier and T. Haertle. 1998. Purification, molecular properties and specificity of a thermoactive and thermostable proteinase from *Pyrococcus abyssi*, strain st 549, hyperthermophilic archaea from deep-sea hydrothermal ecosystem. *FEBS Lett.* 431, 279-284.
3. Durbin, R. S., R. Eddy, A. Krogh and G. Mitchison. 1998. Biological sequence analysis (Probabilistic models of proteins and nucleic acids, pp. 134-159 Cambridge University Press. London.
4. Eisen, J. A. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis, *Genome Res.* 8, 163-187.
5. Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. M. Fritchman, J. F. Weidman, J. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, O. M. Saudek, T. A. Phillips, J. M. Merrick, J.-F. Tomb, D. A. Dougherty, K. F. Bott, P. -C. Hu, T. S. Lucier, S. M. Peterson, H. O. Smith, C. A. Hutchison III and J. C. Venter. 1995. The minimal gene complement of *mycoplasma genitalium*. *Science* 270, 397-403.
6. <ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria>
7. Gonzalez, J. M., Y. Masuchi, F. T. Robb, J. W. Ammerman, D. L. Maeder, M. Yanagibayashi, J. Tamaoka and C. Kato. 1998. *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophile* 2, 123-130.
8. Gribaldo, S. and P. Cammarano. 1998. The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery *J. Mol. Evol.* 47, 508-516.
9. Gupta, R. S. 1998. Life's third domain (Archaea): An established fact or an endangered paradigm?. *Theor.*

- Popul. Biol.* **54**, 91-104.
10. Gupta R. S. 1998. What are archaeobacteria: life's third domain or monoderm prokaryotes related to gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Molecular Biol.* **29**, 695-707.
 11. Henikoff, S., E. A. Greene, B. S. Pietrokovski, T. K. Attwood and L. Hood. 1997. Gene Families: The Taxonomy of Protein Paralogs and Chimeras, *Science* **278**, 609-614.
 12. <http://www.genoscope.cns.fr/Pab/>
 13. <http://www.ncbi.nlm.nih.gov/COG/>
 14. <http://www.tigr.org/CMR2/BackGround/arg.html>
 15. Kang, H.-Y., C.-J. Shin, B.-C. Kang, J.-H. Park, D.-H. Shin, J.-H. Choi, H.-G. Cho, J.-H. Cha, D.-G. Lee, J.-H. Lee, H.-K. Park and C.-M. Kim. 2002. Investigation of Conserved Gene in Microbial Genomes using *in silico* Analysis, *Korean Journal of Life Sci.* **5**, 610-621.
 16. Kimura, M. 1983. The neutral theory of molecular evolution, pp. 65-97, *Cambridge University Press*.
 17. Koonin, E. V., A. R. Mushegian, M. Y. Galperin and D. R. Walker. 1997. Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol. Microbiol.* **25**, 619-637.
 18. Lee, D.-G., H.-Y. Kang, J.-H. Lee and C.-M. Kim. 2003. Detection of Conserved Genes in *Proteobacteria* by using a COG Algorithm. *Korean J. Biotechnol. Bioeng.* **17**, 560-565.
 19. Lee, D.-G., H.-Y. Kang, S. Kim, S.-H. Lee, C.-M. Kim, S.-J. Kim and J.-H. Lee. 2003. Classification of Archaeobacteria and Bacteria using a gene content tree approach. *Korean J. Biotechnol. Bioeng.* **18**, 39-44.
 20. Lee, W.-J. and K. Ohwada. 1995. Studies on the ecological characteristics of marine bacteria isolated from deep sea. *J. Korean Fish. Soc.* **28**, 401-411.
 21. Makarova, K. S., L. Aravind, M. Y. Galperin, N. V. Grishin, R. L. Tatusov, Y. I. Wolf and E. V. Koonin. 1999. Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res.* **9**, 608-28.
 22. Manco, G., E. Giosue, S. D'Auria, P. Herman, G. Carrea and M. Rossi. 2000. Cloning, Overexpression, and Properties of a New Thermophilic and Thermostable Esterase with Sequence Similarity to Hormone-Sensitive Lipase Subfamily from the Archaeon *Archaeoglobus fulgidus*. *Arch. Biochem. Biophys.* **373**, 182-192.
 23. Mullis, K. B. 1990. Process for amplifying nucleic acid sequences. *Science* **6**, 16-25.
 24. Mushegian, A. 1999, The minimal genome concept, *Curr. Opin. Genet.* **9**, 709-714.
 25. Mushegian, A. and E. V. Koonin. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes, *Proc. Natl. Acad. Sci. USA.* **93**, 10268-10273.
 26. Nierman, W., J. A. Eisen and C. M. Fraser. 2000. Microbial genome sequencing 2000: new insights into phylogeny, evolution and expression analysis. *Res. Microbiol.* **151**, 79-84.
 27. Pisano, M. A., M. K. Sommer and M. M. Lopez. 1989. Isolation of bioactive actinomycetes from marine sediment using rifampicin. *Appl. Microbiol. Biotech.* **31**, 609-612.
 28. Saitou, N. and M. Nei. 1987. The neighbor-joining method, a new method for reconstructing method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* **4**, 406-425.
 29. Sako, Y., N. Nomura, A. Uchida, Y. Ishida, H. Morii, Y. Koga, T. Hoaki and T. Maruyama. 1996. *Aeropyrum pernix* gen. nov., sp. nov., a novel aerobic hyperthermophilic archaeon growing at temperatures up to 100 degrees C. *Int. J. Syst. Bacteriol.* **46**, 1070-1077.
 30. Segerer, A., T. A. Langworthy and K. O. Stetter. 1988. *Thermoplasma acidophilum* and *Thermoplasma volcanium* sp. nov. from solfatara fields. *Syst Appl Microbiol.* **10**, 161-171.
 31. Stetter, K. O. 1988. *Archaeoglobus fulgidus* gen. nov., sp. nov.: a new taxon of extremely thermophilic archaeobacteria. *Syst. Appl. Microbiol.* **10**, 172-173.
 32. Tatusov, R. L., M. Y. Galperin, D. A. Natale and E. V. Koonin. 2000. The COG database, a tool for genome-scale analysis of protein functions and evolution, *Nucleic Acids Res.* **28**, 33-36.
 33. Tatusov, R. L., E. V. Koonin and D. L. Lipman. 1997. A genomic perspective on protein families, *Science* **278**, 631-637.
 34. Vielle, C. and G. J. Zeikus. 2001. Hyperthermophilic

- enzymes: Sources, uses, and molecular mechanism for thermostability. *Microbiol. Molecul. Biol. Rev.* **65**, 1-43.
35. Vothknecht, U. C. and D. L. Tumbula. 1999. Archaea: from genomics to physiology and the origin of life. *Trends Cell Biol.* **9**, 159-161.
36. Woese, C. R., O. Kandler and M. L. Wheelis. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* **87**, 4576-4579.
37. Wolf, Y. I., I. B. Rogozin, N. V. Grishin, R. L. Tatusov and E. V. Koonin. 2001. Genome trees constructed using five different approaches suggest new major acterial clades. *BMC Evolutionary Biol.* **1**, 8.
38. Zeikus, J. G., and R. S. Wolfe. 1972. *Methanobacterium thermoautotrophicus* sp. n., an anaerobic, autotrophic, extreme thermophile. *J. Bacteriol.* **109**, 707-713.

(Received March 3, 2003; Accepted June 11, 2003)