

보 문

## *Nitrosomonadales* 목의 핵심유전체(core genome)와 범유전체(pan-genome)의 비교유전체학적 연구

이진환<sup>1</sup> · 김경호<sup>2\*</sup>

<sup>1</sup>국립수산과학원 양식관리과, <sup>2</sup>부경대학교 미생물학과

### Comparative analysis of core and pan-genomes of order *Nitrosomonadales*

Jinhwan Lee<sup>1</sup> and Kyoung-Ho Kim<sup>2\*</sup>

<sup>1</sup>Aquaculture Division, National Institute of Fisheries Science, Busan 46083, Republic of Korea

<sup>2</sup>Department of Microbiology, Pukyong National University, Busan 48513, Republic of Korea

(Received November 30, 2015; Accepted December 14, 2015)

**ABSTRACT:** All known genomes (N=10) in the order *Nitrosomonadales* were analyzed to contain 9,808 and 908 gene clusters in their pan-genome and core genome, respectively. Analyses with reference genomes belonging to other orders in *Betaproteobacteria* revealed that sizes of pan-genome and core genome were dependent on the number of genomes compared and the differences of genomes within a group. The sizes of pan-genomes of the genera *Nitrosomonas* and *Nitrospira* were 7,180 and 4,586 and core genomes, 1,092 and 1,600, respectively, which implied that similarity of genomes in *Nitrospira* were higher than *Nitrosomonas*. The genomes of *Nitrosomonas* contributed mostly to the size of the pan-genome and core genomes of *Nitrosomonadales*. COG analysis of gene clusters showed that the J (translation, ribosomal structure and biogenesis) category occupied the biggest proportions (9.7–21.0%) among COG categories in core genomes and its proportion increased in the group which genetic distances among members were high. The unclassified category [-] occupied very high proportions (34–51%) in pan-genomes. Ninety seven gene clusters existed only in *Nitrosomonadales* and not in reference genomes. The gene clusters contained ammonia monooxygenase (*amoA* and *amoB*) and -related genes (*amoE* and *amoD*) which were typical genes characterizing the order *Nitrosomonadales* while they contained significant amount (16–45%) of unclassified genes. Thus, these exclusively-conserved gene clusters might play an important role to reveal genetic specificity of the order *Nitrosomonadales*.

**Key words:** *Nitrosomonadales*, comparative genomics, core genome, pan-genome

*Nitrosomonadales* 목은 *Betaproteobacteria*에 속해 있으며 암모니아 산화 세균을 포함하고 있는 대표적인 분류군이다. 암모니아 산화 세균은 질산화과정(nitrification)의 첫 번째 단계를 담당하고 있기 때문에 자연환경의 질소 순환에서 매우 중요하다. *Nitrosomonadales* 목은 분류학적으로 *Nitrosomonadaceae* 과 하나로 이루어져 있으며 이 과는 다시 *Nitrosomonas* (*Nm.*) 와 *Nitrospira* (*Ns.*)의 두 개의 속으로 구성되어 있다(<http://www.ezbiocloud.net>, <http://www.bacterio.net>). 이전의 *Nitrosolobus*와 *Nitrosovibrio*는 *Nitrospira*와 합하여 *Nitrospira*

로 재 분류되었다(Head *et al.*, 1993). *Nitrosomonas*와 *Nitrospira*의 균주들은 토양, 담수, 폐수, 해양환경 등에서 분리가 되고 있다(Head *et al.*, 1993; Utáker *et al.*, 1995). 특히 순환여과양식장(recirculating aquaculture system)과 같은 환경에서 세균 군집의 일정 부분을 차지하고 있으며 중요한 역할을 하는 것으로 알려져 있다(Itoi *et al.*, 2006; Foesel *et al.*, 2008). *Nitrosomonadaceae* 과에서 배양된 균들은 모두 암모니아 산화 독립 영양 세균으로 알려져 있다(Teske *et al.*, 1994). 반면에 암모니아 산화 세균을 포함하고 있는 *Gammaproteobacteria*의 *Chromatiaceae* 과의 경우에는 많은 수의 비 암모니아 산화 세균들을 포함하고 있는 것으로 알려져 있다(Klotz and Stein,

\*For correspondence. E-mail: kimkh@pknu.ac.kr;  
Tel.: +82-51-629-5611; Fax: +82-51-629-5619

2011). 16S rRNA 유전자 서열 분석을 통하여 *Nitrosomondales* 목의 계통학적 연관성이 분석되었으며 이에 따라 7-8개의 계통으로 나뉘는 것으로 알려져 있다(Purkhold *et al.*, 2003).

범유전체(pan-genome)라는 것은 주어진 계통학적 분기군(phylogenetic clade)에 속해있는 유전자들의 전체집합으로 구성원들의 모든 생활방식을 모아놓은 것이라고 할 수 있다(Vernikos *et al.*, 2015). 핵심유전체(core genome)는 한 분기군 내의 모든 균주가 포함하고 있는 유전자들의 집합으로 분기군의 가장 기본적인 기능과 주요 표현형들과 관련된 유전자들로 이루어져 있다(Tettelin *et al.*, 2005). 부가유전체(dispensable genome)의 경우에는 한 분기군 내의 부분집단이 공유하고 있는 유전자들의 모음으로 필수적이지는 않으나 선택적 이득을 주는 부가적인 생화학적 대사과정을 암호화하고 있는 유전자들을 나타낸다(Medini *et al.*, 2005). 범유전체, 핵심유전체, 부가유전체의 분석을 통하여 주어진 집단의 공통 유전자들과 다양성을 살펴볼 수 있다.

현재 50,000개 가량의 세균의 유전체 분석이 진행되었거나 진행되고 있다(<http://www.genomesonlines.org>). *Nitrosomondales* 목의 경우에도 *Nm. europaea* ATCC 19718 (Chain *et al.*, 2003)와 *Nm. eutropha* C-91 (Stein *et al.*, 2007)와 *Ns. multiformis* ATCC 25196 (Norton *et al.*, 2008)와 *Nm. sp.* Is79 (Bollmann *et al.*, 2013)의 유전체 서열 분석 결과가 보고되었다. *Nm. cryotolerans* ATCC 49181와 *Ns. briensis* C-128와 *Ns. sp.* AL212의 경우에는 Joint Genome Institute에서 유전체를 분석하여 공개하였으며, *Nm. communis* Nm2, *Ns. sp.* NpAV, *Ns. sp.* APG3의 유전체가 NCBI에 공개되어 있다. 따라서 현재 *Nitrosomonas* 속 6균주와 *Nitrospira* 속 4균주를 합하여 총 10개의 유전체 서열을 이용할 수 있다.

본 연구에서는 *Nitrosomonas* 속, *Nitrospira* 속, *Nitrosomonadales* 목의 범유전체와 핵심유전체를 분석하고 *Beta-proteobacteria*의 다른 목의 유전체들을 참조균으로 하여 그 특성을 알아보려 한다. 본 연구는 *Nitrosomonadales* 목의 최초의 범유전체와 핵심유전체 분석 보고라 할 수 있다.

## 재료 및 방법

### 서열 정보

*Nitrosomonadales* 목에 속하는 10개의 유전체 서열과 *Beta-proteobacteria*의 다른 목에 속하는 6개의 서열을 NCBI FTP (<ftp://ftp.ncbi.nih.gov/genome/Bacteria>)에서 수집하여 범유전체 분석에 사용하였다. 각각의 RefSeq assembly accession

number는 GCF\_001007935.1 (*Nm. communis* Nm2), GCF\_000620405.1 (*Nm. cryotolerans* ATCC 49181), GCF\_000009145.1 (*Nm. europaea* ATCC 19718), GCF\_000014765.1 (*Nm. eutropha* C91), GCF\_000175095.2 (*Nm. sp.* AL212), GCF\_000219585.1 (*Nm. sp.* Is79A3), GCF\_000619905.2 (*Ns. briensis* C-128), GCF\_000196355.1 (*Ns. multiformis* ATCC 25196<sup>T</sup>), GCF\_000832065.1 (*Ns. sp.* NpAV), GCF\_000355765.1 (*Ns. sp.* APG3)이었다. *Beta-proteobacteria* 내의 다른 목에 속하고 *Nitrosomonadales*와 가장 가까운 균들 중 유전체 서열의 품질이 높고 유전체의 크기가 비슷한 균들의 유전체들로 총 6개를 참조균으로 선정하였다. 참조균의 6개 서열은 GCF\_000297055.2 (*Sulfuricella denitrificans* skB26), GCF\_000026125.1 (*Herminiimonas arsenicoxydans* ULPAs1), GCF\_000013625.1 (*Janthinobacterium sp.* Marseille), GCF\_000025705.1 (*Sideroxydans lithotrophicus* ES-1T), GCF\_000429665.1 (*Azovibrio restrictus* DSM 23866), GCF\_000478565.1 (*Methyloversatilis universalis* Fam500)와 같다(Table 1). 이들은 *Nitrosomonas europaea* ATCC 19718와 89-90% 정도의 16S rRNA 유전자 상동성을 보였다.

### 유전체 비교

PGAP (pan-genomes analysis pipeline) ver 1.12를 사용하여 범유전체 분석을 수행하였다(Zhao *et al.*, 2012). GF (gene family) 방식을 통하여 ortholog clusters를 결정하였으며 프로그램에서 제공하는 기본값(E-value: 1e-10, score: 40, identity: 0.5, coverage: 0.5)를 사용하였다(Zhao *et al.*, 2012). 추가적인 분석은 R-project (<http://www.R-project.org>)와 python(<http://www.python.org>)을 이용하여 분석하였다.

### 계통 분석

16S rRNA 유전자는 각각의 유전체에 있는 것을 추출하여 이용하였다. MEGA6 프로그램을 이용하여 neighbor joining tree를 그렸다(Tamura *et al.*, 2013). 거리 행렬(distance matrix)을 구하기 위하여 Kimura 2 parameter 방법을 이용하였다(Kimura, 1980). 16S rRNA 유전자를 비교하기 위한 참고서열로는 NC\_002696 (*Caulobacter crescentus* CB15)을 이용하였다. 공통 유전자의 수를 이용한 계통수는 PGAP ver 1.12를 이용하여 작성하였다(Zhao *et al.*, 2012).

### 단백질의 특성 분석

단백질의 유전자 약자와 관련된 정보는 EMBL-EBI (Squizzato

**Table 1.** Genome sequences used in this study

ID	Strain	Order	Assembly Accn*	Chr*	Pla*	Con*	Size (bps)	GC (%)
Nmcom2	<i>Nitrosomonas communis</i> Nm2	<i>Nitrosomonadales</i>	GCF_001007935.1	1	0	1	4,067,838	44.7
Nmcr49	<i>Nitrosomonas cryotolerans</i> ATCC 49181	<i>Nitrosomonadales</i>	GCF_000620405.1	ND*	ND	91	2,767,948	43.5
Nmeu19	<i>Nitrosomonas europaea</i> ATCC 19718	<i>Nitrosomonadales</i>	GCF_000009145.1	1	0	1	2,812,094	50.7
Nmeu91	<i>Nitrosomonas eutropha</i> C91	<i>Nitrosomonadales</i>	GCF_000014765.1	1	2	3	2,781,824	48.5
Nmsp21	<i>Nitrosomonas</i> sp. AL212	<i>Nitrosomonadales</i>	GCF_000175095.2	1	2	3	3,337,023	44.8
Nmsp79	<i>Nitrosomonas</i> sp. Is79A3	<i>Nitrosomonadales</i>	GCF_000219585.1	1	0	1	3,783,444	45.4
Nsbr12	<i>Nitrospira briensis</i> C-128	<i>Nitrosomonadales</i>	GCF_000619905.2	1	0	1	3,187,395	53.3
Nsmu25	<i>Nitrospira multiformis</i> ATCC 25196 <sup>T</sup>	<i>Nitrosomonadales</i>	GCF_000196355.1	1	3	4	3,234,309	53.9
NsspAV	<i>Nitrospira</i> sp. NpAV	<i>Nitrosomonadales</i>	GCF_000832065.1	1	0	317	3,450,004	53.1
NsspG3	<i>Nitrospira</i> sp. APG3	<i>Nitrosomonadales</i>	GCF_000355765.1	ND	ND	84	3,107,181	53.6
Sude26	<i>Sulfuricella denitrificans</i> skB26	<i>Sulfuricellales</i>	GCF_000297055.2	1	1	2	3,217,213	56.1
Hears1	<i>Herminiimonas arsenicoxydans</i> ULPAs1	<i>Burkholderiales</i>	GCF_000026125.1	1	0	1	3,424,307	54.3
JaspMa	<i>Janthinobacterium</i> sp. Marseille	<i>Burkholderiales</i>	GCF_000013625.1	1	0	1	4,110,251	54.2
SiliS1	<i>Sideroxydans lithotrophicus</i> ES-1T	<i>Gallionellales</i>	GCF_000025705.1	1	0	1	3,003,656	57.5
Azre23	<i>Azovibrio restrictus</i> DSM 23866	<i>Rhodocyclales</i>	GCF_000429665.1	ND	ND	56	4,025,691	65.4
Meun50	<i>Methyloversatilis universalis</i> Fam500	Unclassified	GCF_000478565.1	1	0	1	4,023,710	67.0

\* Accn, accession number; Chr, number of chromosomes; Pla, number of plasmids; Con, number of contigs; ND, not determined.

*et al.*, 2015), InterPro (Mitchell *et al.*, 2015), UniProt (Apweiler *et al.*, 2004)에서 검색하여 정리하였다.

## 결과 및 고찰

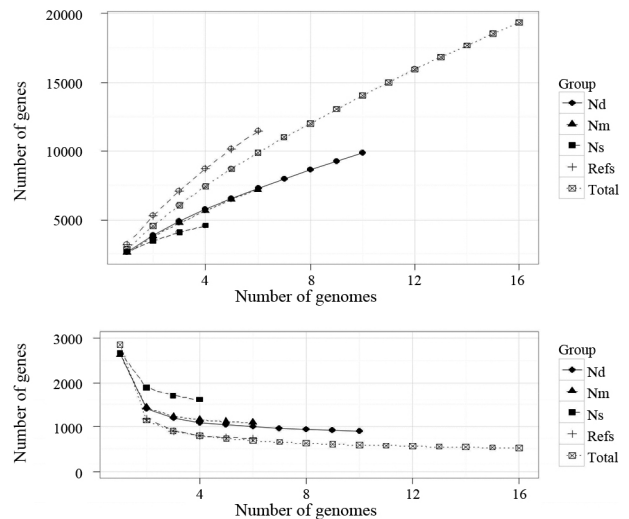
### 연구에 사용된 유전체

총 16개의 유전체는 각각의 *Nitrosomonas* 속, *Nitrospira* 속, 그 외 참조군의 세 개의 집단으로 나눌 수 있다. 모든 유전체의 크기가 2.8-4.1 Mb의 범주에 속하며 집단 내에서도 골고루 존재하고 있다(Table 1). 단 *Nitrosomonas* 속의 3종은 유전체의 크기가 2.7-2.8 Mb로 작은 모습을 보였으며, 이에 비해 *Nm. communis* Nm2의 경우에는 4.1 Mb의 비교적 큰 유전자를 가지고 있다. G+C 함량의 경우에는 집단 별로 차이가 있으며 *Nitrosomonas*, *Nitrospira*, 참조군으로 갈수록 커지는 경향을 보여주고 있다. 12개의 유전체는 완전한 서열이 밝혀졌으나 4개의 서열은 수십에서 수백 개의 서열조각(contigs)으로 이루어져 있다.

### 범유전체와 핵심유전체 분석

*Nitrosomonadales* 목(Nd)의 10개의 유전체의 총 27,841개의 유전자의 분석 결과, 범유전체는 9,808개, 핵심유전체는

908개의 유전자 클러스터로 이루어져 있었다(Fig. 1). 각각 *Nitrosomonas* 속(Nm)과 *Nitrospira* 속(Ns)의 경우에는 범유전체는 7,180개와 4,586개, 핵심유전체는 1,600개와 1,092개의 유전자 클러스터로 각각 이루어져 있다. 비교된 유전체의 증가에 따른 범유전체와 핵심유전체의 크기의 변화를 각각



**Fig. 1.** Number of gene clusters with pangenome (the upper plot) and core genome (the lower plot) vs number of genomes compared according to group. A spot indicates a mean value of number of gene clusters at given number of genomes. Nd, *Nitrosomonadales* group; Nm, *Nitrosomonas* group; Ns, *Nitrospira* group; Refs, reference group; Total, total genomes.

의 집단에 따라 비교하였다(Fig. 1). Ns의 경우에는 범유전체 크기의 증가와 핵심유전체 크기의 감소 속도가 가장 낮은 것을 볼 수 있으며 이는 가장 동질한 유전체를 가지고 있다는 것을 의미한다. *Nitrosomonas* 속은 *Nitrospira* 속에 비해 좀 더 유전체의 동질성이 적은 것으로 볼 수 있다. 또한 Nm과 Nd의 그래프가 비슷하게 변하는 것을 볼 때 *Nitrosomonadales* 목의 범유전체와 핵심유전체의 크기는 주로 *Nitrosomonas* 속에 의해서 좌우되는 것을 알 수 있다. 서로 다른 목들로 구성되어 있는 참조군 집단(Refs)의 경우에는 범유전체의 증가 속도가 가장 빠른 것을 볼 수 있으며 핵심유전체의 감소도 가장 빠르게 일어나고 그 값도 가장 낮은 것을 볼 수 있었다.

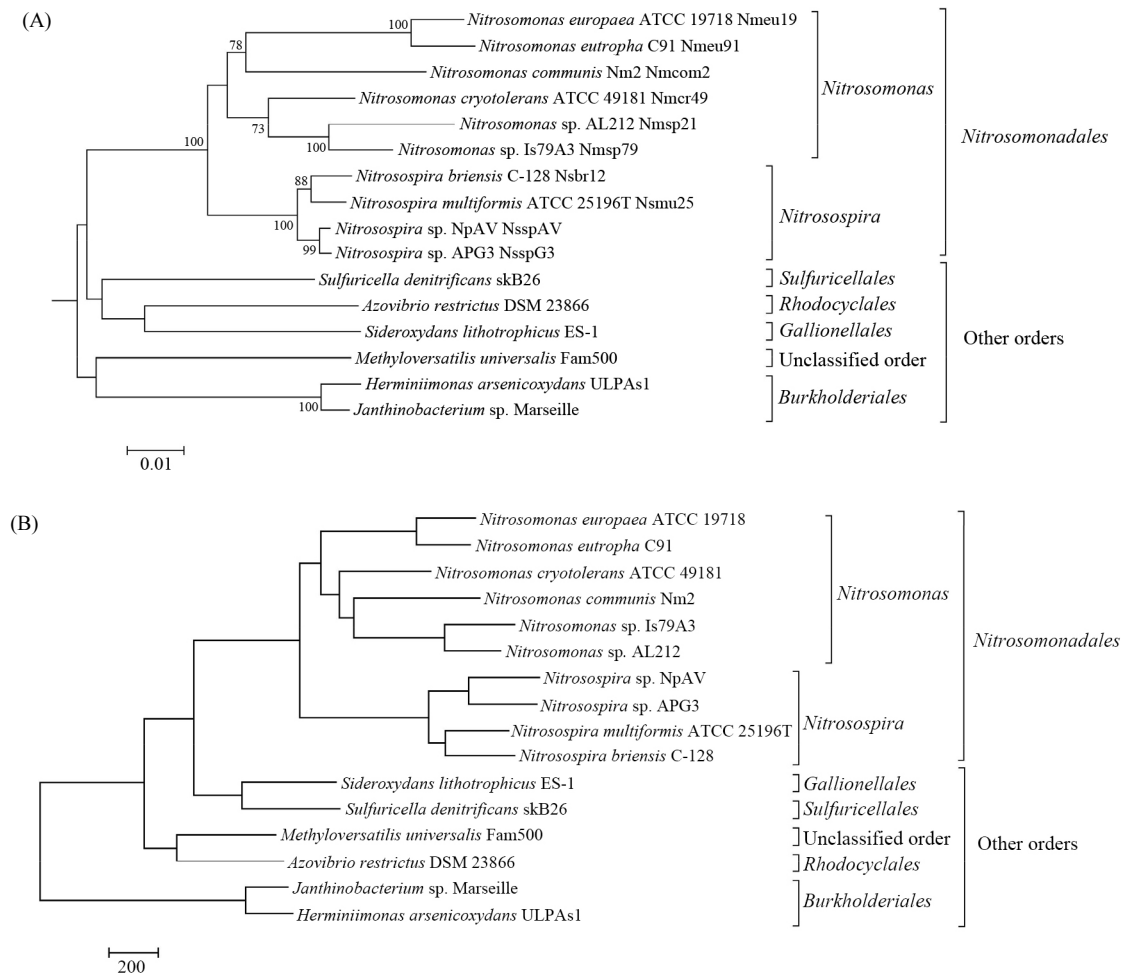
Fig. 1을 보면 비교하는 유전체의 개수에 따라서 범유전체의 크기가 증가하는 것을 볼 수 있다. 따라서 포화되기 이전에는 집단간의 비교를 위해서는 같은 수의 유전체의 결과를 비교하는 것이 바람직하다고 할 수 있다. 본 연구에서는 최소의

유전체의 수인 4개에서 비교한 결과를 보면 상대적인 비교를 할 수 있다.

또한 각각의 균주가 가지고 있는 유전자의 수가 다르기 때문에 범유전체의 증가속도 또한 다를 수 있다. 예를 들어서 *Streptomyces*의 경우에는 17개 균주의 비교에서 34,592개의 범유전체의 크기를 보고하였다(Kim *et al.*, 2015). 이는 같은 속 내의 비교이고 본 연구는 목 수준의 비교이지만 같은 속의 범유전체의 크기가 오히려 1.5배 이상 큰 것을 볼 수 있다. 이는 *Nitrosomonadales* 목과 그 관련 균들은 유전체 크기의 평균이 3.4 Mb인데 반하여 비교된 *Streptomyces* 속의 평균은 8.9 Mb로 훨씬 더 크기 때문이다.

### 16S rRNA 유전자와 공통 유전자의 수를 이용한 계통수 분석

16S rRNA 유전자를 이용한 서열 비교와 공통 유전자 클러스터의 수를 이용한 계통수 분석을 수행하였다(Fig. 2). 두 계



**Fig. 2.** Neighbor joining tree based on 16S rRNA genes. (A) The bar indicates distance of 16S rRNA genes. (B) Neighbor-joining tree based on absence and presence of gene clusters in each strain from pan-genome profiles. The bar indicates difference of gene cluster numbers.



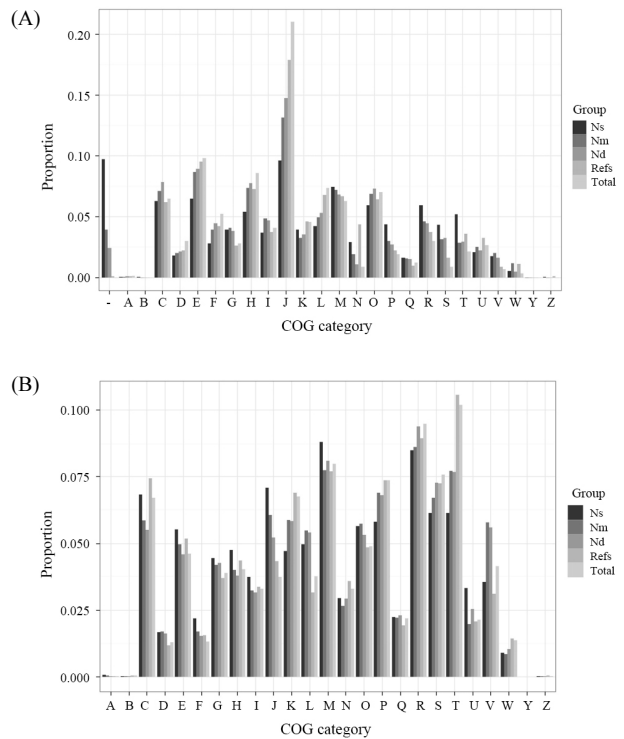
통수의 형태는 어느 정도 일치하는 것을 볼 수 있었다. *Nitrospira*의 경우에는 양 계통수 모두에서 근연관계에 있는 단계 통적 분기군(monophyletic clade)을 이루는 것을 볼 수 있었다. *Nitrosomonas*의 경우에는 *Nm. communis*와 *Nm. cryotolerans*의 위치가 양 계통수에서 조금씩 차이가 있는 것을 볼 수 있었다. 목 수준에서는 두 계통수가 분기되는 모습이 다른 모습을 보여주고 있으나 16S rRNA 유전자의 계통수의 bootstrap value가 낮은 것을 미루어 보아 계통학적 위치가 확실치 않기 때문일 것으로 사료된다. 공통유전자의 수를 이용한 계통 분석에서도 Ns의 경우에는 매우 가까운 관계에 있으며 Refs의 경우에는 서로 관련성이 떨어지는 집단인 것을 확인할 수 있었다.

### 유전자 클러스터의 기능적 비교

PGAP 프로그램(Zhao *et al.*, 2012)을 통하여 COGs (Tatusov *et al.*, 1997)와 관련성이 있는 유전자 클러스터들을 개개의 유전체에만 존재하는 지, 여러 유전체에 존재하는지, 모든 유전체에 존재하는 지에 따라서 특이(specific), 부가(dispensable), 핵심(core) 유전자 클러스터로 나누었다. 핵심유전자 클러스터는 핵심유전체와 관련성이 있으며 핵심, 부가, 특이유전자 클러스터의 합은 범유전체와 관련성이 높기 때문에 핵심유전자 클러스터와 세 유전자 클러스터의 합에서 각각의 COG 범주가 차지하는 비율을 살펴보았다(Fig. 3). 이를 통하여 핵심유전체와 범유전체에서 각각의 기능에 속하는 유전자 클러스터들의 비율을 비교하며, 집단에 따라 어떻게 달라지는지 살펴볼 수 있다.

핵심유전자의 경우 COG 범주의 비율이 집단에 따라 달라지는 것을 볼 수 있다. 전체적으로 가장 많은 비율을 차지하고 있는 J (translation, ribosomal structure and biogenesis) 범주의 경우에는 유전적 다양성이 높아질수록 그 비율이 크게 증가하는 것을 볼 수 있으며 이는 유전적 안정성이 높은 것을 의미한다(Fig 3A). E (amino acid transport and metabolism), F (nucleotide transport and metabolism), L (replication, recombination and repair) 등의 경우에도 상대적으로 미약하지만 같은 경향을 보이고 있다. 반면에 “-” (unclassified)의 경우에는 반대의 경향을 보이며 참조군이나 전체에서 비율은 거의 0에 가까운 것을 볼 수 있다. S (function unknown), R (general function prediction only), P (inorganic ion transport and metabolism)의 경우에도 조금 덜하지만 비슷한 경향이 나타나고 있다. 유전적으로 안정적이지 못한 유전자들을 포함하고 있는 것을 짐작할 수 있다.

핵심, 부가, 특이유전자 클러스터를 합하여 이를 COG 범주에 따라 그 비율을 구하였으며 이는 범유전체를 구성하는 유



**Fig. 3.** Proportion of genes in each category for (A) core genes and (B) sum of core, dispensable, and specific genes according to group: -, Unclassified; A, RNA processing and modification; B, Chromatin structure and dynamics; C, Energy production and conversion; D, Cell cycle control, cell division, chromosome partitioning; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; G, Carbohydrate transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; J, Translation, ribosomal structure and biogenesis; K, Transcription; L, Replication, recombination and repair; M, Cell wall/membrane/envelope biogenesis; N, Cell motility; O, Posttranslational modification, protein turnover, chaperones; P, Inorganic ion transport and metabolism; Q, Secondary metabolites biosynthesis, transport and catabolism; R, General function prediction only; S, Function unknown; T, Signal transduction mechanisms; U, Intracellular trafficking, secretion, and vesicular transport; V, Defense mechanisms; W, Extracellular structures; Y, Nuclear structure; Z, Cytoskeleton. For figure B, the number of gene clusters in the unclassified (-) category was not included in proportion calculation.

전자 클러스터들에서 COG 범주의 비율을 의미한다. 그 결과 “-” (unclassified) 범주가 34-51%를 차지하며 범유전체에 새로운 유전자 클러스터들을 제공하는데 가장 중요한 역할을 하며 이는 가장 덜 보존되어 있기 때문이라고 할 수 있다. 나머지 범주는 대부분 5% 미만으로 존재하고 있는 것을 볼 수 있다. “-” 범주만을 제외하고 다른 범주들만 모아 상대적인 비율을 구하여 그래프로 나타내었다(Fig. 3B). Ns와 Nm을 비교한 결과 일반적으로 Ns에서 C, E, F, H, I, J, M 등의 비율이 높으며, Nm에서는 K (transcription), P (inorganic ion transport and metabolism), T (signal transduction mechanisms)와 V (defense

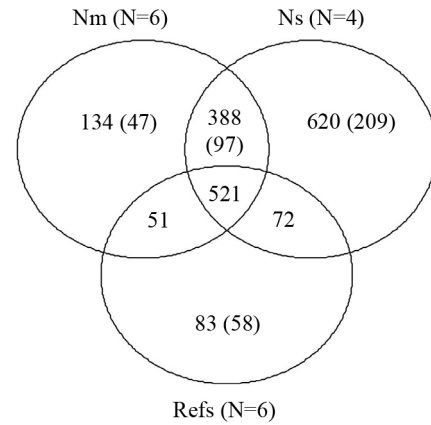
mechanisms) 범주에서 비율이 더 높은 것을 볼 수 있다. 일반적으로 Ns에서 높은 비율을 가진 것들은 기본적인 house keeping 능력과 관련된 범주들이며, Nm에서 높은 비율을 가진 것들은 대사능력과 관련된 범주들이다. 이는 *Nitrosomonas*의 경우에는 좀 더 다양한 대사과정을 포함하고 있다는 것을 의미하며, 이러한 차이는 *Nitrospira*와 *Nitrosomonas*에 공통적인 질산화 및 이산화탄소고정 대사과정 외에 존재하는 생태학적 또는 생리적 차이를 의미할 수 있다.

### *Nitrosomonadales*의 중요 유전자

Fig. 4는 각각의 집단에 100% 존재하는 유전자 클러스터의 개수를 비교하는 Venn diagram이다. 세 집단에 모두 존재하는 유전자 클러스터의 갯수는 핵심 유전체의 갯수인 521과 같은 것을 볼 수 있다. 이중 Ns, Nm, Nd에만 존재하는 유전자 클러스터는 각각 209, 47, 97개이다. 이들 중-(unclassified) 범주에 속하는 것이 각각 45, 28, 16%를 차지하는 것으로 확인되었다. 이는 각각의 속이나 목에만 특이적으로 존재하지만 아직까지 그 특성이 제대로 밝혀지지 않은 것들이 많이 존재한다는 것

을 의미한다.

*Nitrosomonadales*에 존재하는 유전자 중 중요한 유전자의 특성을 정리하였다(Table 2). Ammonia monooxygenase 유전



**Fig. 4.** Venn diagram showing distribution of gene clusters which existed in the all members of each group. The numbers 47, 209, and 97 in parentheses indicate that the number of gene clusters exist exclusively in Nm, Ns, and Nd group, respectively.

**Table 2.** Important gene clusters in the order *Nitrosomonadales*

Cluster ID	Gene	Description	Remark	Protein id*	COG no	COG category
1142	<i>amoA</i>	ammonia monooxygenase	first enzyme from ammonia to nitrite	WP_041357108	-	-
1179	<i>amoB</i>	ammonia monooxygenase	first enzyme from ammonia to nitrite	WP_011111552	-	-
17	<i>amoC</i>	ammonia monooxygenase	first enzyme from ammonia to nitrite not found in NsspG3	WP_011111979	-	-
25	<i>amoE (orf4)</i>	hypothetical protein	conserved in betaproteobacterial ammonia oxidizing bacteria	WP_011111551	-	-
762	<i>amoD (orf5)</i>	hypothetical protein	conserved in betaproteobacterial ammonia oxidizing bacteria	WP_011111550	-	-
1060	<i>hao</i>	hydroxylamine reductase	not found in NsspG3	WP_011111571	COG3303	P
1239	<i>cycA</i>	cytochrome C554	involved in ammonia oxidation, cytochrome c554, not found in NsspG3 and Nmcr49	WP_011111569	-	-
1090	<i>cycB</i>	cytochrome C	probable tetraheme cytochrome, not found in NsspG3 and Nmcr49	WP_011111568	COG3005	C
1228	<i>haoB</i>	hypothetical protein	not found in NsspG3, Nmcr49	WP_011111570	-	-
1567	<i>amt</i>	ammonium transporter	ammonium transporter, found in 7 <i>Nitrosomonadales</i> members	WP_011111083	COG0004	P
1897	<i>amt</i>	ammonium transporter	ammonium transporter, found in all Refs members, not in <i>Nitrosomonadales</i> members	WP_009204881	COG0004	P
47	<i>cbbL</i>	ribulose biphosphate carboxylase	Rubisco large subunit	WP_011112458	COG1850	G
77	<i>cbbE</i>	ribulose-phosphate 3-epimerase	found in all strains	WP_011112648	COG0036	G
1195	<i>cbbS</i>	ribulose biphosphate carboxylase	Rubisco small subunit	WP_011112457	COG4451	G

\* Protein id, representative protein id found in *Nitrosomonas europaea* ATCC 19718 or *Sulfuricella denitrificans* skB26.

자는 암모니아를 아질산으로 바꾸는 과정에서 첫 번째로 사용되는 효소로 오페론을 이루고 있다(Arp *et al.*, 2002). 3개의 유전자로 구성된 *amoCAB* 오페론은 유전체 안에 두 개 내지 세 개의 복사본을 가지고 있는 것이 보고되었다(Stein *et al.*, 2007; Bollmann *et al.*, 2013). *amoA*, *amoB*, *amoC*, 세 개의 유전자 모두 암모니아 산화세균에 필수적이고 대표적인 유전자로 본 분석에서 *Nitrosomonadales* 목에만 배타적으로 존재하였다. *amoA*와 *amoB*는 모든 *Nitrosomonadales* 목의 유전체에 존재하나 *amoC*의 경우에는 *Ns. sp. APG3*에서만 발견되지 않았다. 이는 이 유전체가 83개의 서열조각으로 이루어져 누락된 부분이 있기 때문인 것으로 보인다. 실제로 *amoA*와 *amoB*는 6,400 bp 길이의 서열조각에 역순으로 존재하고 *amoC*가 있어야 할 그 앞 부분은 결실되어 있는 것을 확인하였다. *amoCAB* 오페론 아래에 있는 두 개의 유전자 *amoE* (*orf4*)와 *amoD* (*orf5*)는 *Betaproteobacteria* 암모니아 산화세균에만 특이적으로 존재하는 것으로 알려져 있다(Stein *et al.*, 2007). 본 연구에서도 두 유전자는 모든 *Nitrosomonadales* 목의 유전체에서 발견되었고 참조군에서는 발견되지 않았다.

수산화아민 산화환원효소(hydroxylamine oxidoreductase)는 암모니아 산화과정에서 두 번째 효소로 수산화아민을 아질산으로 바꾸는 역할을 한다(Arp *et al.*, 2002). *Betaproteobacteria* 암모니아 산화세균들은 세 개의 *hao* 오페론을 가지고 있으며 octaheme cytochrome c 효소의 소단위(subunit)를 암호화하는 *haoA*와 알려지지 않은 유전자 산물을 생산하는 *haoB*, cytochrome c<sub>554</sub>의 *cycA*, cytochrome c<sub>M552</sub>의 *cycB*로 이루어져 있다(El Sheikh *et al.*, 2008). 이 중 *hao*의 경우에는 *Ns. sp. APG3*에서만 발견되지 않았다. 그리고 *haoB*, *cycA*, *cycB*의 경우에는 *Ns. sp. APG3*와 함께 *Nm. cryotolerans* ATCC 49181에서도 발견되지 않았다. *Ns. sp. APG3*와 마찬가지로 *Nm. cryotolerans* ATCC 49181의 경우에는 유전체가 91개의 서열조각으로 이루어져 있으며 *haoA*가 존재하는 위치가 44 kb 크기 서열조각의 가장 마지막인 것으로 볼 때 유전체 서열이 완벽히 얻어지지 않았기 때문이라 할 수 있다. Ammonium transporter는 암모니아를 동화적으로 이용하기 위해서 필요한 유전자이다. 흥미롭게도 각각 *Nitrosomonadales* 목(N=7)과 참조군에 특이적으로 존재하는 두 종류의 ammonium transporter 유전자 클러스터가 발견되었다. 이는 암모니아 산화세균의 경우에 다른 균과 다른 동화적 암모니아 이용 능력을 가지고 있을 가능성을 의미한다(Bollmann *et al.*, 2013). 본 연구에서 분석한 모든 암모니아 산화세균은 독립영양세균으로서 이산화탄소를 고정화하는 주요 효소인 RuBisCo (Ribulose-1,5-bisphosphate carboxylase oxygenase)의 큰 소단위(large

subunit)를 모두 가지고 있으며 참조군에서도 3개의 유전체에서 발견된다. 이는 참조군도 *Sulfuricella denitrificans*, *Sideroxydans lithotrophicus*, *Methyloversatilis universalis* 등과 같은 독립영양세균을 포함하고 있기 때문이다. 참조군으로 분석된 *Methyloversatilis universalis*와 같은 메틸영양세균으로는 *Betaproteobacteria* 내의 다른 목인 *Methylophilaceae*의 *Methylophilus* 속(Xia *et al.*, 2015)이나 *Methylovorus* 속(Xiong *et al.*, 2011) 등이 잘 알려져 있다. 또한 본 연구에 사용된 *Bukholderiales* 목의 *Herminiimonas arsenicoxydans* ULPA1와 *Janthinobacterium sp. Marseille*와 같이 3-4 Mb의 유전체를 가지고 있는 것뿐만 아니라 7-8 Mb 정도의 더 큰 유전체를 가진 *Bukholderia* 속의 주요 균들의 분석을 통하여 *Betaproteobacteria*에 대한 전체적인 범유전체 및 핵심 유전체 분석 연구가 차후에 진행되어야 할 것이다.

## 결론

*Nitrosomonadales* 목은 암모니아 산화세균 중 매우 중요한 분류군이며 알려진 모든 균주가 암모니아를 산화시키는 독립영양세균으로 알려져 있듯이 계통적으로나 생리적으로 매우 동질성을 가지고 있는 집단이다. 범유전체 분석이 10여 년 전에 최초로 수행된 이래로 아직까지 *Nitrosomonadales* 목과 관련된 연구가 수행된 적이 없다(Vernikos *et al.*, 2015). 본 연구의 범유전체와 핵심유전체 분석을 통하여 이들의 유전적 특성을 좀 더 잘 이해하고 이를 이용하여 질산화 과정에 대한 이해를 더욱 높일 수 있을 것으로 사료된다.

## 적요

*Nitrosomonadales* 목에서 속하는 균주 중 현재 유전체 서열이 알려진 모든 유전체(N=10)를 이용하여 범유전체 및 핵심 유전체 분석을 수행한 결과, 각각 9,808개와 908개 유전자 클러스터를 포함하는 것을 확인하였다. *Betaproteobacteria*의 다른 목의 참조군들과 비교를 통하여 범유전체와 핵심유전체의 크기에 유전체의 수와 집단 내의 유전체들의 차이가 영향을 미치는 것을 확인하였다. *Nitrosomonas* 속과 *Nitrospira* 속의 범유전체는 7,180개와 4,586개, 핵심유전체는 1,092개와 1,600개로 각각 측정되어 *Nitrospira* 속의 동질성이 더 높은 것을 확인하였다. *Nitrosomonadales* 목의 범유전체와 핵심 유전체의 크기에 *Nitrosomonas* 속이 대부분의 영향을 미치는

것을 확인하였다. COG 분석을 통하여 핵심유전체의 크기에 J (translation, ribosomal structure and biogenesis) 범주가 가장 큰 비율(9.7-21.0%)을 차지하며, 유전체 사이의 유전적 거리가 먼 집단일수록 그 비율이 높아지는 것을 확인하였다. 범유전체의 크기에는 “-” (unclassified) 범주가 34-51%의 높은 비율을 차지하고 있을 정도로 큰 영향을 미치는 것을 확인하였다. 총 97 개의 유전자 클러스터가 참조군에는 없고 *Nitrosomonadales*에만 존재하는 것을 확인하였다. 이들 클러스터들은 *Nitrosomonadales*을 특징 지우는 유전자들인 ammonia monooxygenase의 유전자인 *amoA*와 *amoB*와 그와 관련 있는 *amoE*와 *amoD*들을 포함하는 반면에 unclassified 유전자들도 상당량(16-45%)을 포함하고 있다. 이러한 유전자 클러스터는 *Nitrosomonadales*의 유전적 특이성을 밝히는 데 중요한 역할을 할 것이다.

## 감사의 말

이 논문은 2014학년도 부경대학교 자율창의연구비의 지원을 받아 수행된 연구임.

## References

- Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. 2004. UniProt: the Universal Protein knowledge base. *Nucleic Acids Res.* **32**, D115-119.
- Arp, D.J., Sayavedra-Soto, L.A., and Hommes, N.G. 2002. Molecular biology and biochemistry of ammonia oxidation by *Nitrosomonas europaea*. *Arch. Microbiol.* **178**, 250-255.
- Bollmann, A., Sedlacek, C.J., Norton, J., Laanbroek, H.J., Suwa, Y., Stein, L.Y., Klotz, M.G., Arp, D., Sayavedra-Soto, L., Lu, M., et al. 2013. Complete genome sequence of *Nitrosomonas* sp. Is79, an ammonia oxidizing bacterium adapted to low ammonium concentrations. *Stand. Genomic Sci.* **7**, 469-482.
- Chain, P., Lamertin, J., Larimer, F., Regala, W., Lao, V., Land, M., Hauser, L., Hooper, A., Klotz, M., Norton, J., et al. 2003. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J. Bacteriol.* **185**, 2759-2773.
- El Sheikh, A.F., Poret-Peterson, A.T., and Klotz, M.G. 2008. Characterization of two new genes, *amoR* and *amoD*, in the *amo* operon of the marine ammonia oxidizer *Nitrosococcus oceanus* ATCC 19707. *Appl. Environ. Microbiol.* **74**, 312-318.
- Foesel, B.U., Gieseke, A., Schwermer, C., Stief, P., Koch, L., Cytryn, E., de la Torre, J.R., van Rijn, J., Minz, D., Drake, H.L., et al. 2008. *Nitrosomonas* Nm143-like ammonia oxidizers and *Nitrosopira marina*-like nitrite oxidizers dominate the nitrifier community in a marine aquaculture biofilm. *FEMS Microbiol. Ecol.* **63**, 192-204.
- Head, I.M., Hiorns, W.D., Embley, T.M., McCarthy, A.J., and Saunders, J.R. 1993. The phylogeny of autotrophic ammonia-oxidizing bacteria as determined by analysis of 16S ribosomal RNA gene sequences. *J. Gen. Microbiol.* **139 Pt 6**, 1147-1153.
- Ito, S., Niki, A., and Sugita, H. 2006. Changes in microbial communities associated with the conditioning of filter material in recirculating aquaculture systems of the pufferfish *Takifugu rubripes*. *Aquaculture* **256**, 287-295.
- Kim, J.N., Kim, Y., Jeong, Y., Roe, J.H., Kim, B.G., and Cho, B.K. 2015. Comparative genomics reveals the core and accessory genomes of *Streptomyces* species. *J. Microbiol. Biotechnol.* **25**, 1599-1605.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111-120.
- Klotz, M.G. and Stein, L.Y. 2011. Genomics of ammonia-oxidizing bacteria and insights into their evolution. Nitrification, American Society of Microbiology.
- Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. 2005. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589-594.
- Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., et al. 2015. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213-221.
- Norton, J.M., Klotz, M.G., Stein, L.Y., Arp, D.J., Bottomley, P.J., Chain, P.S., Hauser, L.J., Land, M.L., Larimer, F.W., Shin, M.W., et al. 2008. Complete genome sequence of *Nitrosospira multiformis*, an ammonia-oxidizing bacterium from the soil environment. *Appl. Environ. Microbiol.* **74**, 3559-3572.
- Purkhold, U., Wagner, M., Timmermann, G., Pommerening-Roser, A., and Koops, H.P. 2003. 16S rRNA and *amoA*-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: extension of the dataset and proposal of a new lineage within the nitrosomonads. *Int. J. Syst. Evol. Microbiol.* **53**, 1485-1494.
- Squizzato, S., Park, Y.M., Buso, N., Gur, T., Cowley, A., Li, W., Uludag, M., Pundir, S., Cham, J.A., McWilliam, H., et al. 2015. The EBI Search engine: providing search and retrieval functionality for biological data from EMBL-EBI. *Nucleic Acids Res.* **43**, W585-588.
- Stein, L.Y., Arp, D.J., Berube, P.M., Chain, P.S., Hauser, L., Jetten, M.S., Klotz, M.G., Larimer, F.W., Norton, J.M., Op den Camp, H.J., et al. 2007. Whole-genome analysis of the ammonia-oxidizing bacterium, *Nitrosomonas eutropha* C91: implications for niche adaptation. *Environ. Microbiol.* **9**, 2993-3007.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729.



- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* **278**, 631–637.
- Teske, A., Alm, E., Regan, J.M., Toze, S., Rittmann, B.E., and Stahl, D.A. 1994. Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria. *J. Bacteriol.* **176**, 6623–6630.
- Tettelin, H., Maignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Anguoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., *et al.* 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. USA* **102**, 13950–13955.
- Utåker, J.B., Bakken, L., Jiang, Q.Q., and Nes, I.F. 1995. Phylogenetic analysis of seven new isolates of ammonia-oxidizing bacteria based on 16S rRNA gene sequences. *Syst. Appl. Microbiol.* **18**, 549–559.
- Vernikos, G., Medini, D., Riley, D.R., and Tettelin, H. 2015. Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* **23**, 148–154.
- Xia, F., Zou, B., Shen, C., Zhu, T., Gao, X.H., and Quan, Z.X. 2015. Complete genome sequence of *Methylophilus* sp. TWE2 isolated from methane oxidation enrichment culture of tap-water. *J. Biotechnol.* **211**, 121–122.
- Xiong, X.H., Zhi, J.J., Yang, L., Wang, J.H., Zhao, Y., Wang, X., Cui, Y.J., Dong, F., Li, M.X., Yang, Y.X., *et al.* 2011. Complete genome sequence of the bacterium *Methylovorus* sp. strain MP688, a high-level producer of pyrroloquinolone quinone. *J. Bacteriol.* **193**, 1012–1013.
- Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., and Yu, J. 2012. PGAP: pan-genomes analysis pipeline. *Bioinformatics* **28**, 416–418.