

Additional File 1 Index

Table S1 : *Main features and taxonomic assignment of Oosacac minuta's metagenome assembly*

Table S2 : *Taxonomic assignment of non-Eukaryotic and non-Archaeal Taxa*

Figure S1 : *Phylogenetic position of the main microbial species associated with O. minuta*

Table S3 : *Genome features of Candidatus Cenarchaeum massiliensis compared to other Thaumarchaea.*

Table S4 : *Annotation of the predicted protein-coding genes of Ca. C. massiliensis*

Figure S2 : *Characterization of the genome of Ca. C. massiliensis*

Figure S3 : *The cobalamin synthesis genes in Thaumarchaeota.*

Figure S4 : *Localization of Ca. C. massiliensis in the tissues of O. minuta by Card-FISH*

Table S1: Main features and taxonomic assignment of *Oopsacas minuta*'s metagenome assembly.

Scaffolds (>1kb) were retained and submitted to MetaGeneMark (Zhu et al, 2010)

followed by a blastp search (Altschul et al, 1990) strategy against NR (best hits with evalue < 10⁻⁵). A LCA-like method computed over all annotated genes was applied on each scaffold:

the taxonomy assigned to each scaffold corresponds to that found for at least half of the genes composing the scaffold.

Scaffolds were then classified into 5 groups: Eukaryota, Bacteria, Archaea, virus and Not Assigned taxa.

Taxa	Mapped Reads	% of total reads	Mean Coverage	Total length (bp)	% of total assembly	Contigs > 1kb	polished scaffolds	Polished scaffold mean coverage	Reference ID
Eukaryota									
Porifera nuclear genome	111941409	77.95	177	61170773	95.91	936	365	186	JAKMXF010000000
Mitochondria	89460	0.06	455	19518	0.03	1	1	455	NC_027419.1
Archae									
mainly Thaumarcheota	26189162	18.24	1537	1658688	2.60	50	11	1206	JAJIZT010000000
Bacteria									
Gamma proteobacteria	3762639	2.62	657	555205	0.87	44			
other-Proteobacteria	454494	0.32	759	54400	0.09	13			
Cyanobacteria	79313	0.06	104	75563	0.12	17			
Firmicutes	33314	0.02	92	36020	0.06	10			
other-bacteria	71121	0.05	127	55808	0.09	14			
NA*-bacteria	344951	0.24	468	71964	0.11	9			
Virus	633133	0.44	778	81222	0.13	2			
Ambiguous-taxa	2037432	1.42	1124	164832	0.26	33			
NA*-Taxa	3888087	2.71	241	1525204	2.39	630			
TOTAL	143598996		220**	63779161		1096			

All statistics were collected before final scaffolding and polishing and only for contigs > 1 kbp

the total number of raw reads is: 395 851 746 (197 925 873 pairs)

* NA= Not Assigned

Table S2: Taxonomic assignment of non-Eukaryotic and non-Archaeal Taxa

Scaffolds >1kb have been submitted to MetaGeneMark followed by a blastp search against NR (best hits with evalue < 10⁻⁵) A LCA-like method computed over all annotated genes was applied on each scaffold: the taxonomy assigned to each scaffold corresponds to that found for at least half of the genes composing the scaffold. Scaffolds were then classified into 5 groups: Eukaryota, Bacteria, Archaea, virus and Not Assigned taxa.

number of scaffolds		Taxonomic assignment				
1	Bacteria Aquificae	Aquificae	Aquificales	Hydrogenothermaceae	Persephonella	Persephonella sp.
1	Bacteria Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	Flavobacterium sp.
1	Bacteria Bacteroidetes	Cytophagia	Cytophagales	Amoebophilaceae	Candidatus Amoebophilus	Candidatus Amoebophilus asiaticus
1	Bacteria Bacteroidetes					
2	Bacteria Candidatus Dadabacteria	Candidatus Dadabacteria	Candidatus Dadabacteria bacterium	CSP1-2		
1	Bacteria Candidatus Tectomicrobia	Candidatus Entotheonella	Candidatus Entotheonella sp.	TSY2		
1	Bacteria Chloroflexi	Chloroflexia	Chloroflexiales	Chloroflexineae	Oscillochloridaceae	Oscillochloris Oscillochloris trichoides
13	Bacteria Cyanobacteria	Oscillatoriophycideae	Chroococcales	Microcystaceae	Microcystis	Microcystis aeruginosa
1	Bacteria Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Oscillatoriaceae	Limnoraphis	Limnoraphis robusta
1	Bacteria Cyanobacteria	Oscillatoriophycideae	Chroococcales	Microcystaceae	Microcystis	Microcystis sp.
1	Bacteria Cyanobacteria	Nostocales	Nostocaceae	Anabaena	Anabaena sp.	
1	Bacteria Cyanobacteria	cyanobacterium				
3	Bacteria Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Aneurinibacillus	Aneurinibacillus soli
1	Bacteria Firmicutes	Negativicutes	Selenomonadales	Selenomonadaceae	Selenomonas	Selenomonas ruminantium
1	Bacteria Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacterales	Family III.Incertae Sedis	Caldicellulosiruptor Caldicellulosiruptor kristjanssonii
1	Bacteria Firmicutes	Clostridia	Halanaerobiales	Halobacteroidaceae	Halobacteroidaceae	bacterium
1	Bacteria Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus	Enterococcus hirae
1	Bacteria Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	Bacillus humi
1	Bacteria Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	Bacillus fastidiosus
1	Bacteria Firmicutes	Bacilli	Bacillales			
2	Bacteria Nitrospirae	Nitrospirae	bacterium			
1	Bacteria Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Candidatus Magnetoovum	Candidatus Magnetoovum chiemensis
1	Bacteria Parcubacteria	group bacterium GW2011_GWC1_45_13				
1	Bacteria Parcubacteria	group bacterium GW2011_GWC1_38_22				
34	Bacteria Proteobacteria	Gammaproteobacteria				
5	Bacteria Proteobacteria					
2	Bacteria Proteobacteria	Gammaproteobacteria	endosymbiont of Ridgeia piscesae			
1	Bacteria Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Vibrio	Vibrio crassostreae
1	Bacteria Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oleiphilaceae	Oleiphilus	Oleiphilus sp. HI0086
1	Bacteria Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	Oleispira	Oleispira antarctica
1	Bacteria Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	Marinobacterium	Marinobacterium jannaschii
1	Bacteria Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae	Haliea	Haliea salexigens
1	Bacteria Proteobacteria	Gammaproteobacteria	Cellvibrionales			
1	Bacteria Proteobacteria	Gammaproteobacteria	Alteromonadales	Moritellaceae	Moritella	Moritella sp.
1	Bacteria Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Marinobacter	Marinobacter lipolyticus
1	Bacteria Proteobacteria	delta/epsilon subdivisions	Epsilonproteobacteria	Sulfurovum	Sulfurovum sp.	
1	Bacteria Proteobacteria	delta/epsilon subdivisions	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	Sulfurospirillum Sulfurospirillum sp.
1	Bacteria Proteobacteria	delta/epsilon subdivisions	Deltaproteobacteria	Desulfuromonadales	Desulfuromonadaceae	Pelobacter Pelobacter carbinolicus
1	Bacteria Proteobacteria	delta/epsilon subdivisions	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio Desulfovibrio longus
1	Bacteria Proteobacteria	delta/epsilon subdivisions				
1	Bacteria Proteobacteria	Betaproteobacteria	Candidatus Accumulibacter	Candidatus Accumulibacter sp.		
1	Bacteria Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium	Sphingobium sp.
1	Bacteria Proteobacteria	Alphaproteobacteria				
1	Bacteria Spirochaetes	Spirochaetia	Leptospiraceae	Leptospira	Leptospira	Leptospira kmetyi
9	Bacteria					
1	Viruses Circoviridae	Circovirus	Bat	circovirus		
1	Viruses					

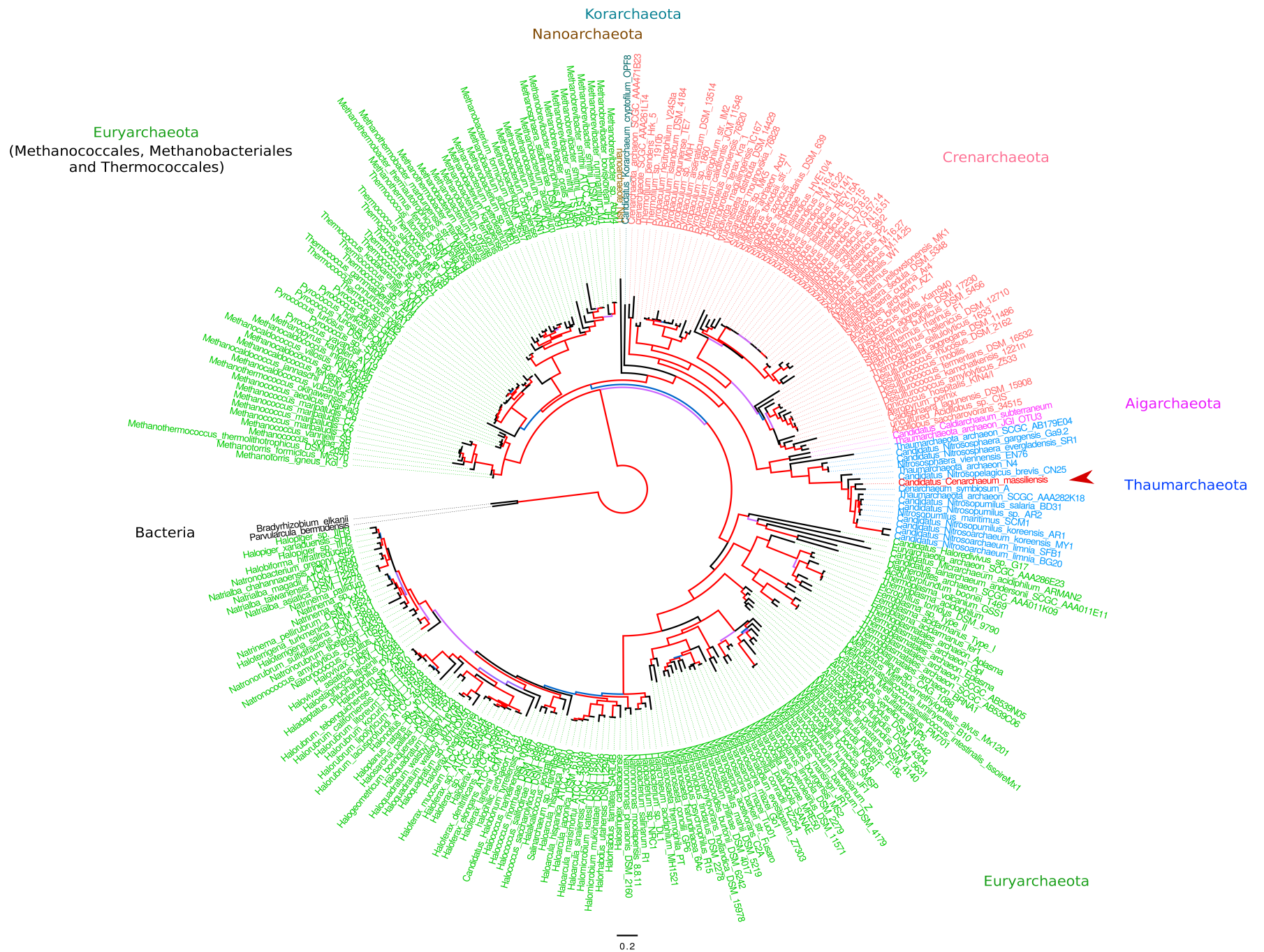


Figure S1: Phylogenetic position of the main microbial species (in red) associated to *O. minuta* (inferred by maximum-likelihood on ribosomal RNA genes, see material and method section for details). Main archaeal phyla are indicated and branch supports (approximate likelihood ratio test statistics with a Shimodaira-Hasegawa-like procedure) are indicated with a color code (>0.9, red; >0.7, purple; >0.5, blue and <0.5, black). Scale bar represents the number of substitutions per site.

Table S3: Genome features of *Candidatus Cenarchaeum massiliensis* compared to other Thaumarchaea.

All statistics were obtained from the current version of the genome and the tRNA were predicted using tRNA-scan-SE (Chan & Lowe 2019).

Species	Size (Mb)	GC content	CDS content	Coding genes	tRNA number	Source
<i>Cenarchaeum massiliensis</i>	1.63	37.30%	83.90%	1675	48	
<i>Cenarchaeum symbiosum</i>	2.05	57.70%	91.40%	2017	45	http://bacteria.ensembl.org/cenarchaeum_symbiosum_a/Info/Index
<i>Nitrosopumilus maritimus</i>	1.65	34.20%	90.30%	1795	44	http://bacteria.ensembl.org/nitrosopumilus_maritimus_scm1/Info/Index
<i>Nitrosopumilus koreensis</i>	1.64	34.20%	87.20%	1890	41	http://bacteria.ensembl.org/candidatus_nitrosopumilus_koreensis_ar1/Info/Index
<i>Nitrosoarchaeum koreensis</i>	1.61	32.70%	89%	1945	42	http://bacteria.ensembl.org/candidatus_nitrosoarchaeum_koreensis_my1/Info/Index
<i>Nitrosoarchaeum limnia</i>	1.77	32.50%	83.40%	2038	45	http://bacteria.ensembl.org/candidatus_nitrosoarchaeum_limnia_sfb1/Info/Index
<i>Nitrosopelagicus brevis</i>	1.23	33.20%	94.60%	1445	42	https://www.ncbi.nlm.nih.gov/nuccore/CP007026.1
<i>Nitrososphaera viennensis</i>	2.53	52.70%	87%	3123	39	http://www.ncbi.nlm.nih.gov/genome/?term=Nitrososphaera+viennensis
<i>Nitrososphaera gargensis</i>	2.83	48.30%	81.60%	3565	34	https://www.ncbi.nlm.nih.gov/genome/?term=Candidatus+Nitrososphaera%20gargensis
<i>Nitrososphaera evergladensis</i>	2.95	50.10%	83.40%	3499	39	http://www.ncbi.nlm.nih.gov/nuccore/CP007174.1

Mb: megabase

CDS: coding sequence.

Candidatus Cenarchaeum massiliensis K8823_1073	382 Redoxin domain-containing protein	IPR008866;IPR012336	OG5_208929	Misc	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1140	257 inositol-3-γ-phosphatase synthase	K01609	OG5_125960	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_12	405 Cysteine sulfinate desulfurase/cysteine desulfurase	K04487	OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1460	335 pyridoxal-phosphate-dependent protein	K01738	OG5_126662	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1471	76 hypothetical protein		nmay YP_001582526	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1475	151 hypothetical protein		OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1598	382 aminotransferase class V	K04487	OG5_138561	Misc	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_164	565 hypothetical membrane protein	K07333	OG5_141276	Misc	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1658	272 6-pyruvoyl-tetrahydropterin synthase		OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1663	384 cysteine desulfurase	K04487	OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1670	384 cysteine desulfurase	K04487	OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_242	131 hypothetical protein	K09736	OG5_145374	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_267	695 regulatory protein ArsR	K00527	OG5_138803	Nucleotide metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_349	79 hypothetical protein		nmay YP_001582401	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_380	1019 DEAD/DEAH box helicase	K01153	OG5_135105	Misc	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_381	405 restriction endonuclease S subunit		OG5_132451	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_382	529 N-6 DNA methylase		OG5_135747	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_540	840 helicase	K03427	OG5_130740	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_549	225 class I glutamine amidotransferase	K06877	OG5_136506	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_559	418 hypothetical protein		OG5_241572	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_576	330 hypothetical protein		OG5_128763	Carbohydrate metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_581	606 potassium transporter Trk		OG5_133741	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_641	253 Nucleotidyl transferase	K00966	OG5_131193	Carbohydrate metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_666	371 carbamoyl-phosphate synthase small subunit		OG5_128943	Nucleotide metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_61	327 Mgt and Co2 transporter	K01956	OG5_156594	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_747	74 hypothetical protein	K03284	nmay YP_001581547	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_773	389 aminotransferase class V		OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_70	334 tRNA (S-methyl aminomethyl-2-thiouridyate)-methyltransferase	K04487	OG5_158947	Translation	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_828	246 uracopyrimin-III C-methyltransferase	K00566	OG5_128100	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_842	91 hypothetical protein	K02303	OG5_149721	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_273	159 Integrase protein family		OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_337	502 MBL fold metallo-hydrolase		OG5_131049	Misc	Cenarchaeum Nitrosoarchaeum Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1333	3274 hypothetical protein	K12574	OG5_131049	Misc	Cenarchaeum Nitrosoarchaeum Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1684	84 hypothetical protein		OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_682	436 hypothetical protein		OG5_126959	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_205	102 hypothetical protein		OG5_146378	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_422	317 aerotolerance regulator Bata		OG5_143376	Stress & defense	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_423	295 hypothetical protein	K07114	OG5_149721	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_424	330 ATPase	K03924	OG5_130615	Misc	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_565	414 iron transporter		OG5_177650	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_598	43 hypothetical protein		nmay YP_001582507	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1179	253 DNA (cytosine-5)-methyltransferase	K00558	OG5_129430	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1398	165 hypothetical membrane protein		OG5_234685	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1619	462 DNA cytosine methyltransferase	K00558	OG5_129430	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_322	48 Ribosomal protein L37E		OG5_126918	Translation	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_518	324 DNA-cytosine methyltransferase		OG5_129430	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_655	332 cytosine-specific methyltransferase	K00558	OG5_129430	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_763	353 DNA (cytosine-5)-methyltransferase	K00558	OG5_129430	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1536	156 methyltransferase		OG5_129618	Transcription	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1573	124 hypothetical protein		nmay YP_001582496	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1603	320 hypothetical protein		OG5_229125	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1650	356 Phosphonate ABC transporter, periplasmic phosphonate-binding protein	K02044	OG5_143155	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1651	250 phosphate ABC transporter ATPase	K02071	OG5_152077	Nucleotide metabolism	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1652	260 phosphonate ABC transporter inner membrane subunit (pH _{in})	K02042	OG5_136945	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_421	624 hypothetical protein		OG5_136945	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_521	77 zinc finger CDGSH-type domain protein		OG5_131688	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_602	416 ABC-type branched-chain amino acid transport system, periplasmic component	K01999	OG5_130525	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_604	550 ABC-type branched-chain amino acid transport system, periplasmic component	K01999	OG5_130525	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_605	243 ABC-type branched-chain amino acid transport system, periplasmic component	K01999	OG5_130525	Transport	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_627	90 zinc finger UBP-type protein		OG5_132669	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_783	320 hypothetical protein		OG5_229125	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_904	231 Cobyrinic acid a-c-diamide synthase	K03496	OG5_130386	Metabolism of cofactors and vitamins	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1322	271 parB-like partition protein		OG5_132436	Transcription	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1566	275 Gp37Gp68 family protein		OG5_165182	Misc	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_398	346 CRISPR-associated protein Cas 1		OG5_136066	Stress & defense	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_389	85 hypothetical protein		OG5_130386	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_890	244 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_902	99 Winged helix-turn-helix containing protein		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_903	244 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_904	79 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_933	334 transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_937	100 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_939	82 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_940	203 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_940	101 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_8	116 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_88	58 hypothetical protein		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_969	92 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_993	314 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1004	347 RelA/SpoT protein		OG5_149795	Signaling	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1034	72 hypothetical membrane protein		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1036	148 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1079	233 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1080	79 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1088	1137 restriction endonuclease		OG5_145232	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1098	1088 DEAD-like helicase-domain-containing protein	K07317	OG5_133337	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1174	371 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1187	44 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1188	119 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1194	187 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1194	145 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1235	323 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1255	506 abortive infection protein		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_119	196 DNA-3-methyladenine glycosylase II	K01247	OG5_136684	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1275	92 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1276	179 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1306	144 death-on-curing family protein		OG5_144670	Signaling	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera
Candidatus Cenarchaeum massiliensis K8823_1362	274 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum Nitrosopumilus Nitrososphaera

Candidatus Cenarchaeum massiliensis K8823_1383	79 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1392	109 hypothetical protein		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1393	212 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1411	96 hypothetical protein		STG_0103	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1507	118 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1571	323 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1574	304 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1575	106 Transposase		OG5_136743	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1577	304 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1579	332 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1600	77 Transposase	IPR002559	STG_0032	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1608	183 Transposase	IPR002559	STG_0032	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1609	179 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1610	134 Homeodomain-like domain-containing protein		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1611	112 Transposase	IPR003346	OG5_127931	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1615	248 Transposase		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1616	79 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1628	125 Homeodomain-like domain-containing protein		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1629	179 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1633	150 hypothetical protein	IPR007527	STG_0289	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_163	430 hypothetical membrane protein		STG_0143	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1661	79 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_175	604 AAA-like domain-containing protein		OG5_179911	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_176	480 Type II/IV secretion system protein	K07332	OG5_175628	Transport	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_18	369 Transposase		OG5_127931	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_229	135 hypothetical protein		STG_0203	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_234	75 hypothetical protein		STG_0101	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_1	431 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_25	230 SMF family protein	K04096	OG5_133325	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_274	185 Transposase	IPR007529	STG_0289	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_27	193 Transposase	IPR003346	OG5_127931	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_31	115 hypothetical protein		STG_0047	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_360	311 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_368	427 DNA methylase domain-containing protein	IPR002941	OG5_181907	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_370	369 Transposase	IPR003346;IPR002525	OG5_127931	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_374	179 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_384	240 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_394	761 CRISPR-associated helicase Cas3 family protein	K07012	OG5_136526	Stress & defense	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_396	297 type I-C CRISPR-associated protein Cas7/Csd2		OG5_228703	Stress & defense	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_397	211 CRISPR-associated protein Cas4	K07464	OG5_139298	Stress & defense	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_399	107 CRISPR-associated endonuclease Cas2		OG5_189005	Stress & defense	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_414	72 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_416	73 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_418	427 DNA methylase	IPR002941	OG5_181907	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_420	286 Transposase	IPR025959	OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_447	98 Transposase		STG_0101	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_472	1353 N6-adenine-specific DNA methyltransferase	IPR002052	OG5_145232	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_473	198 helicase	IPR025202	OG5_163337	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_474	893 helicase	IPR016530;IPR014001;IPR003330;IPR01650;IPR014001;IPR01650	OG5_163337	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_40	148 hypothetical protein		OG5_141713	Amino acid metabolism	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_41	230 SMF family protein	K04096	OG5_133325	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_3	320 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_651	315 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_657	331 Transposase	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_660	331 Transposase DDE domain protein	IPR002559	OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_662	76 Transposase DDE domain protein		OG5_136743	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_663	179 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_693	94 hypothetical protein		STG_0071	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_697	148 Transposase		STG_0047	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_719	38 Transposase	IPR004291	STG_0047	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_725	103 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_726	180 hypothetical protein		OG5_185721	DNA modification	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_69	136 Transposase-like protein		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_790	207 Transposase		OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_794	1029 ATPase	IPR003593	OG5_234716	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_372	286 Transposase	IPR025959	OG5_127031	Unknown function	Cenarchaeum Nitrosoarchaeum
Candidatus Cenarchaeum massiliensis K8823_77	162 hypothetical protein	IPR025272		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_888	163 hypothetical protein	IPR025272		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_910	334 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_934	68 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_942	251 hypothetical protein	IPR007421		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_944	123 DNA sulfur modification protein DndE	IPR014969		DNA modification	Species specific
Candidatus Cenarchaeum massiliensis K8823_945	664 DNA sulfur modification protein DndD	IPR017599		Transport	Species specific
Candidatus Cenarchaeum massiliensis K8823_950	123 Transposase	IPR004291		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_954	158 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_955	79 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_956	84 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_958	100 Transposase	IPR025959		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_959	151 Transposase			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_966	148 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_968	130 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_970	113 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_971	65 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_972	84 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_984	284 HNH endonuclease	IPR003615		Misc	Species specific
Candidatus Cenarchaeum massiliensis K8823_987	510 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_988	198 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_991	117 hypothetical protein	IPR002525		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_995	94 Transposase			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_997	62 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_999	123 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1000	80 Transposase	IPR004291		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1005	390 hypothetical protein	IPR014942		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1006	627 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1018	60 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1020	81 Transposase	IPR002525		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1022	126 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1024	95 hypothetical protein	IPR004291		Unknown function	Species specific

Candidatus Cenarchaeum massiliensis K8823_1564	258 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1565	318 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1569	65 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1570	95 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1576	108 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1578	95 Winged helix-turn helix domain-containing protein	IPR025959		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1592	137 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1597	103 DNA sulfur modification protein DnSE	IPR014969		DNA modification	Species specific
Candidatus Cenarchaeum massiliensis K8823_1601	98 Transposase	IPR002525		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1602	102 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1613	136 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1614	226 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_160	885 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1617	359 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1618	49 transcriptional regulator AbrB family	IPR007159;IPR007159		Transcription	Species specific
Candidatus Cenarchaeum massiliensis K8823_161	245 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1620	127 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1623	164 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1626	91 Transposase	IPR003346		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1627	73 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_162	121 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1631	229 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_165	152 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1654	100 hypothetical protein	IPR025959		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_166	219 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1686	125 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1680	77 Transposase	IPR002559		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1681	128 Transposase	IPR003346		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1682	327 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1683	77 Transposase	IPR002559		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_1695	162 Isocitrate/isopropylmalate dehydrogenase	K00052	IPR024084;IPR001804	Amino acid metabolism	Species specific
Candidatus Cenarchaeum massiliensis K8823_1696	186 2-isopropylmalate synthase	K01649	IPR013709;IPR013709	Misc	Species specific
Candidatus Cenarchaeum massiliensis K8823_173	143 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_174	137 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_182	44 cation-binding protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_190	101 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_191	250 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_21	259 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_233	114 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_23	92 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_243	105 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_24	62 Transposase	IPR002525		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_26	113 Transposase	IPR002525		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_305	129 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_29	109 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_30	590 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_346	205 hydrolase	IPR003615;IPR002711		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_366	141 nitrous oxide reductase	IPR028096		Oxidation-reduction	Species specific
Candidatus Cenarchaeum massiliensis K8823_369	163 restriction endonuclease	IPR019042		Misc	Species specific
Candidatus Cenarchaeum massiliensis K8823_371	145 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_372	63 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_373	49 hypothetical protein	IPR025959		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_375	95 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_376	107 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_377	152 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_383	310 hypothetical protein	IPR002793		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_32	100 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_385	86 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_415	77 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_419	128 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_440	62 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_442	94 Transposase	IPR002525		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_443	78 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_2	127 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_464	109 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_468	463 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_42	191 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_43	217 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_558	191 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_44	192 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_568	105 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_45	355 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_572	124 transposase	IPR002525		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_573	62 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_46	288 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_582	34 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_4	142 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_651	177 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_632	103 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_652	971 LVIVD repeat-containing protein	IPR013211		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_654	225 Eco47II family restriction endonuclease	IPR018057		DNA modification	Species specific
Candidatus Cenarchaeum massiliensis K8823_656	2904 LVIVD repeat-containing protein	IPR013211		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_659	147 Transposase DDE domain protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_664	95 hypothetical protein	IPR025959		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_691	98 hypothetical protein	IPR004291		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_692	110 hypothetical membrane protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_695	110 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_698	131 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_700	122 Transposase	IPR004291		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_712	90 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_713	177 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_716	263 hypothetical protein	IPR021398		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_717	100 hypothetical protein	IPR004291		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_720	75 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_749	31 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_752	184 hypothetical protein	IPR025959		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_758	72 Transposase	IPR003346		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_759	105 Transposase	IPR003346		Unknown function	Species specific

Candidatus Cenarchaeum massiliensis K8823_762	39 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_764	239 restriction endonuclease			Misc	Species specific
Candidatus Cenarchaeum massiliensis K8823_765	126 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_766	75 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_767	95 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_772	114 hypothetical protein	IPR014969		DNA modification	Species specific
Candidatus Cenarchaeum massiliensis K8823_775	65 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_782	785 SIR2 family protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_785	158 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_787	147 Transposase	IPR025225		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_791	62 hypothetical protein	IPR025959		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_796	105 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_797	170 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_808	103 hypothetical protein	IPR002559		Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_809	118 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_810	282 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_861	106 hypothetical protein			Unknown function	Species specific
Candidatus Cenarchaeum massiliensis K8823_866	137 hypothetical protein			Unknown function	Species specific

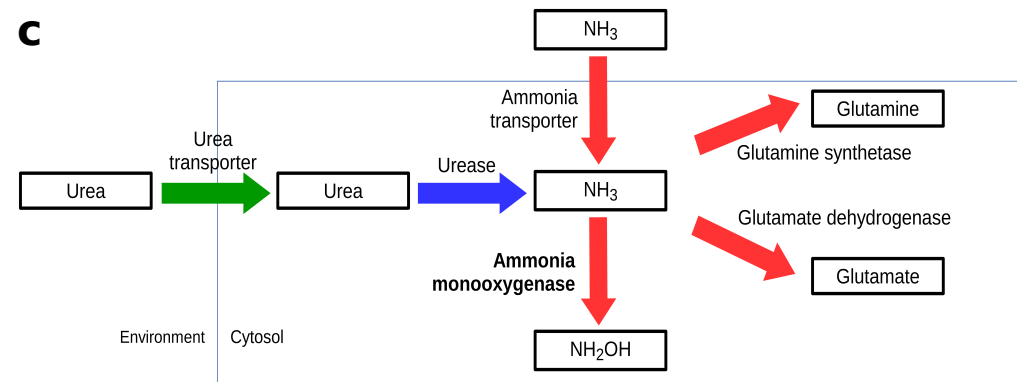
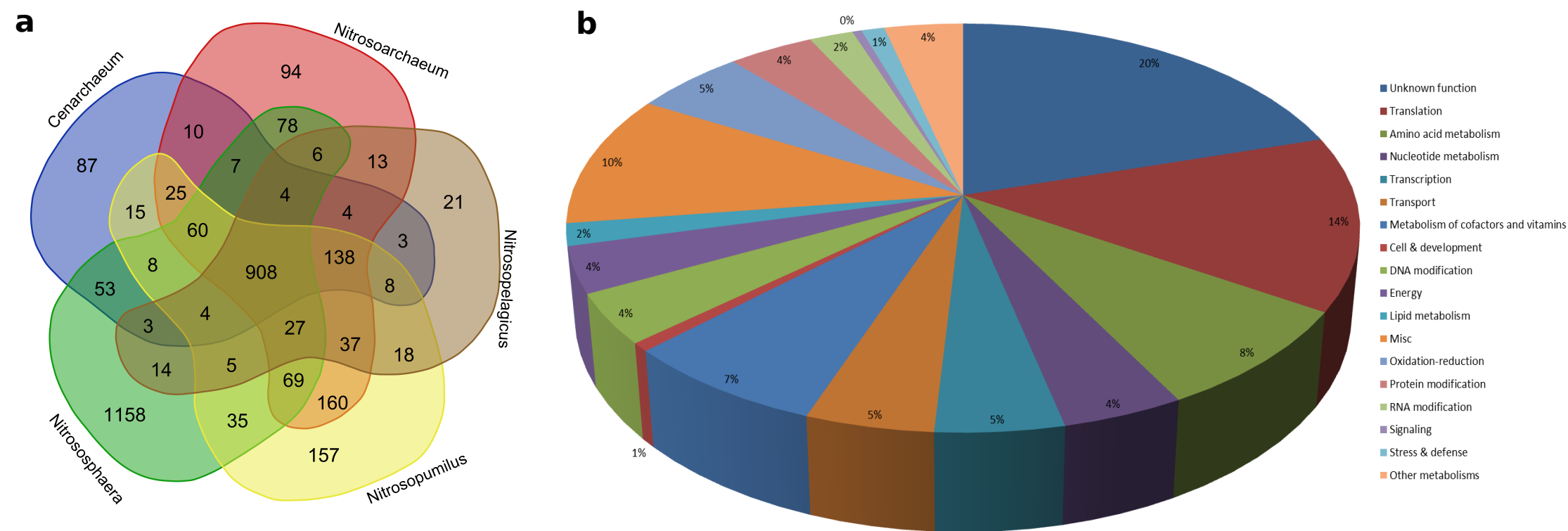


Figure S2: Characterization of the genome of *Candidatus Cenarchaeum massiliensis*, the main symbiont of *O. minuta*, compared to other Thaumarchaeota species. **a :Five-way-Venn-diagram illustrating orthologous genes clusters generated using OrthoMCL analysis on predicted proteins of *Cenarchaeum symbiosum* and *Cenarchaeum massiliensis* (blue circle), *Nitrosoarchaeum koreensis* and *Nitrosoarchaeum limnia*(red circle), *Nitrosopelagicus brevis* (brown circle), *Nitrosopumilus maritimus* and *Nitrosopumilus koreensis* (yellow circle) and three soil Thaumarchaeota (*Nitrososphaera viennensis*, *Nitrososphaera gargensis* and *Nitrososphaera evergladensis*; green circle). Venn-diagram is generated using a webtool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). **b**: Functional classes of the 908 Thaumarchaeota conserved gene families. **c**: The nitrogen metabolism in Thaumarchaeota. Conserved steps in all Thaumarchaeota (red arrows), conserved step only in soil Thaumarchaeota and *Cenarchaeum symbiosum* (blue arrow) and conserved step only in soil Thaumarchaeota (green arrow) are shown.**

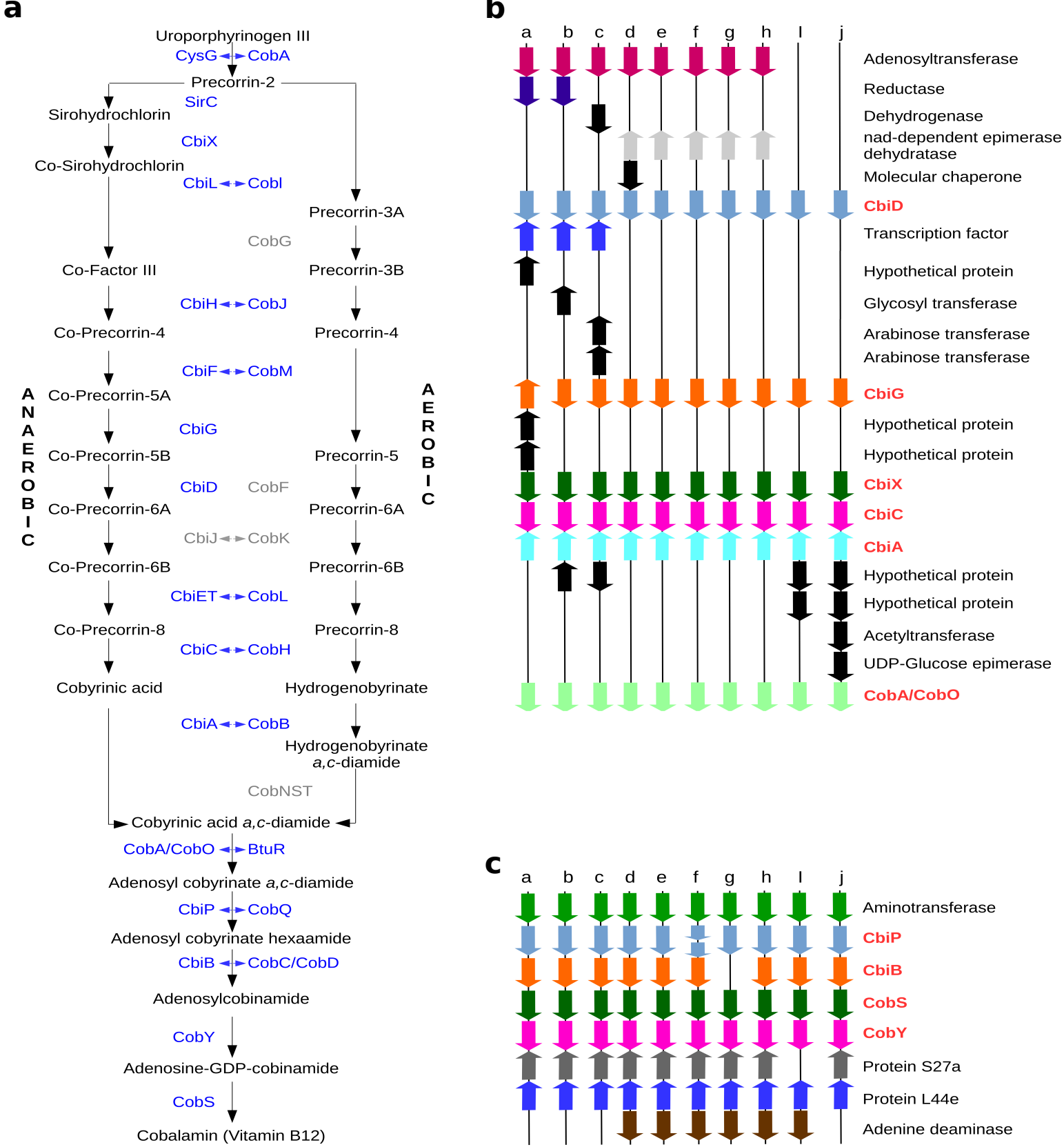


Figure S3: The cobalamin synthesis genes in Thaumarchaeota.
a: Pathway of the cobalamin synthesis adapted from Doxey et al., 2015. Blue arrows indicate shared enzymatic steps between aerobic and anaerobic pathways. All enzymes of the anaerobic pathway are retrieved (indicated in blue) except CbiJ (indicated in grey). **b:** Microsynteny analysis of Thaumarchaeota genomic segments bearing six genes coding for early enzymatic steps of cobalamin synthesis. **c:** Microsynteny analysis of Thaumarchaeota genomic segments bearing four genes coding for late enzymatic steps of cobalamin synthesis. Segments "a" to "j" correspond to genomic segments from *Nitrososphaera gargensis*, *Nitrososphaera viennensis*, *Nitrososphaera evergladensis*, *Nitrosopelagicus brevis*, *Nitrosoarchaeum koreensis*, *Nitrosoarchaeum limnia*, *Nitrosopumilus koreensis*, *Nitrosopumilus maritimus*, *Cenarchaeum symbiosum* and *Candidatus Cenarchaeum massiliensis*, respectively. Arrows of the same color indicate orthologous genes and black arrows represent genes without orthology relationship in these regions from OrthoMCL analysis.

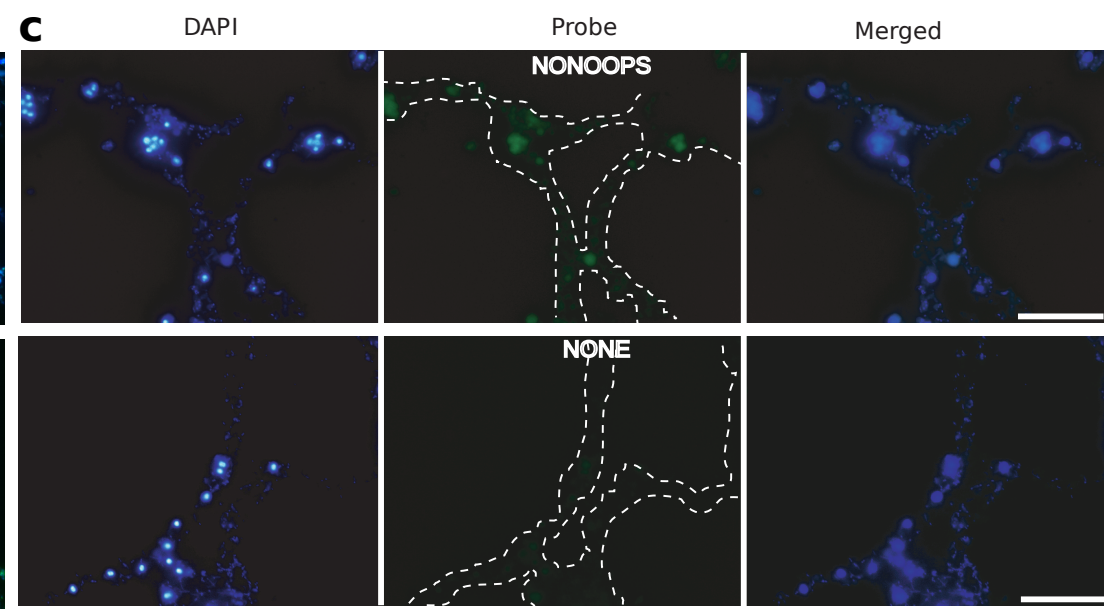
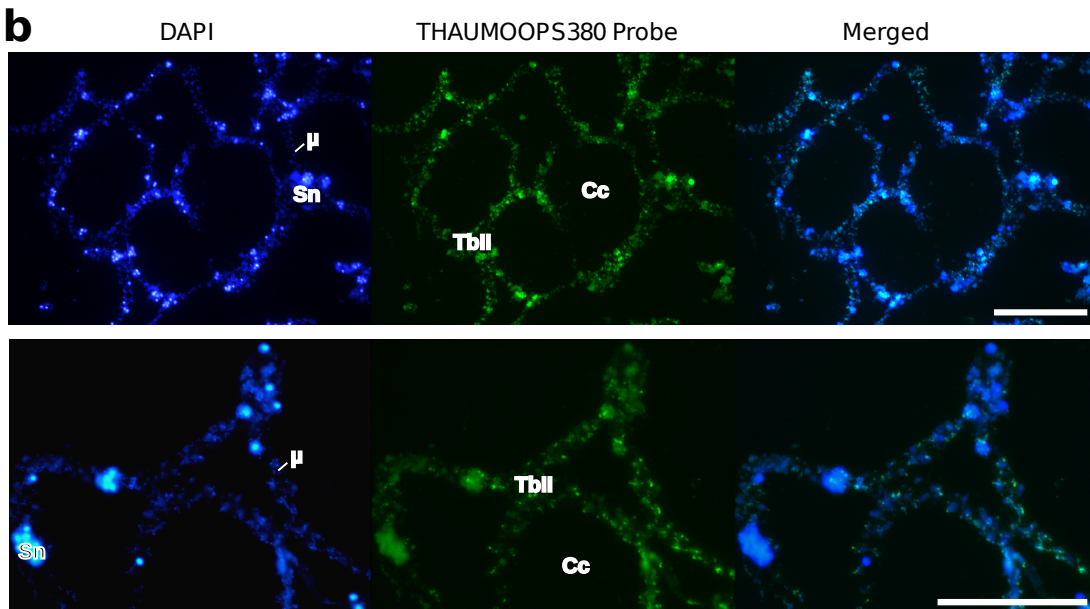
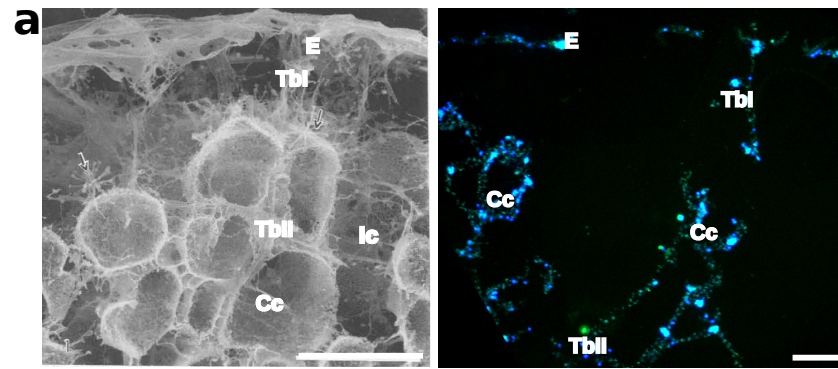


Figure S4: Localization by Card-FISH of the thaumarchaeote *Candidatus Cenarchaeum massiliens* in the tissues of *Opsacas minuta*. **a:** Confocal picture (on the right) showing its localization in the different parts of the trabecular syncytium (Tb) : the Ectosome (E) or dermal layer, the Choanochambers (CC) or flagellated chambers, the inhalant canals (Ic), recognized by comparison with SEM pictures (on the left) (Boury-Esnault and Vacelet, 1994) scale bars = 150 μ m. **b:** Localization in the trabecular syncytium shown at higher magnification in two different individuals with the THAUMOOPS380-HRP probe (green) and counterstained with DAPI (blue), μ : microorganisms DNA, Sn: Sponge nuclei, scale bars = 70 μ m. **c:** Card-FISH negative controls: with a random probe with no 100% match in the metagenome (NONOOPS-HRP), or no probe, evidencing that sponge cells have an endogenous peroxidase activity uncompletely abolished by 0.3% H₂O₂.

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
2. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 2010;38(12):e132.
3. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol.* 2019;1962:1–14.
4. Stieglmeier M, Klingl A, Alves RJE, Rittmann SKMR, Melcher M, Leisch N, et al. *Nitrososphaera viennensis* gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon from soil and a member of the archaeal phylum Thaumarchaeota. *Int J Syst Evol Microbiol.* 2014;64(Pt 8):2738–52.
5. Spang A, Poehlein A, Offre P, Zumbrägel S, Haider S, Rychlik N, et al. The genome of the ammonia-oxidizing *Candidatus Nitrososphaera gargensis*: insights into metabolic versatility and environmental adaptations. *Environ Microbiol.* 2012;14(12):3122–45.
6. Zhalnina KV, Dias R, Leonard MT, Dorr de Quadros P, Camargo FAO, Drew JC, et al. Genome sequence of *Candidatus Nitrososphaera evergladensis* from group I.1b enriched from Everglades soil reveals novel genomic features of the ammonia-oxidizing archaea. *PLoS One.* 2014;9(7):e101648.
7. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35:W182-185.
8. Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, et al. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics.* 2011;Chapter 6:6.12.1-6.12.19.
9. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21(18):3674–6.
10. Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and proteomic characterization of “*Candidatus Nitrosopelagicus brevis*”: an ammonia-oxidizing archaeon from the open ocean. *Proc Natl Acad Sci U S A.* 2015;112(4):1173–8.
11. Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y ichi, Sugahara J, et al. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci U S A.* 2006;103(48):18296–301.
12. Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, et al. *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci U S A.* 2010;107(19):8818–23.
13. Park SJ, Kim JG, Jung MY, Kim SJ, Cha IT, Kwon K, et al. Draft genome sequence of an ammonia-oxidizing archaeon, “*Candidatus Nitrosopumilus koreensis*” AR1, from marine sediment. *J Bacteriol.* 2012;194(24):6940–1.
14. Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS One.* 2011;6(2):e16626.

15. Kim BK, Jung MY, Yu DS, Park SJ, Oh TK, Rhee SK, et al. Genome sequence of an ammonia-oxidizing soil archaeon, "*Candidatus Nitrosoarchaeum koreensis*" MY1. *J Bacteriol.* 2011;193(19):5539-40.
16. Doxey AC, Kurtz DA, Lynch MDJ, Sauder LA, Neufeld JD. Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J.* 2015;9(2):461-71.