

**Strangers in the archaeal world: osmopressure-responsive biosynthesis of ectoine and hydroxyectoine
by the marine thaumarchaeon *Nitrosopumilus maritimus***

Nils Widderich^{1,☞}, Laura Czech^{1,☞}, Felix J. Elling², Martin Könneke², Nadine Stöveken^{1,3}, Marco Pittelkow¹, Ramona Riclea^{4,5}, Jeroen S. Dickschat^{4,5}, Johann Heider^{1,3} and Erhard Bremer^{1,3*}

¹Laboratory for Molecular Microbiology, Department of Biology, Philipps-University, Karl-von-Frisch Str. 8, D-35043 Marburg, Germany

²Organic Geochemistry Group, MARUM – Center for Marine Environmental Sciences, University of Bremen, PO Box 330 440, D-28334 Bremen, Germany

³LOEWE-Center for Synthetic Microbiology, Philipps-University Marburg, Hans-Meerwein-Str. 6, D-35043 Marburg, Germany

⁴Kekulé-Institut for Organic Chemistry and Biochemistry, Friedrich-Wilhelms-University Bonn, Gerhard-Domagk Str. 1, D-53121 Bonn, Germany

⁵Institute of Organic Chemistry, TU Braunschweig, Hagenring 30, D-38106 Braunschweig, Germany

Running title: Ectoine and hydroxyectoine biosynthesis in *Archaea*

☞These authors contributed equally

*Corresponding author. Mailing address: Philipps-University Marburg, Dept. of Biology, Laboratory for Microbiology, Karl-von-Frisch-Str. 8, D-35032 Marburg, Germany. Phone: (+49)-6421-2821529. Fax: (+49)-6421-2828979. E-Mail: bremer@staff.uni-marburg.de

Supplemental Figure S1

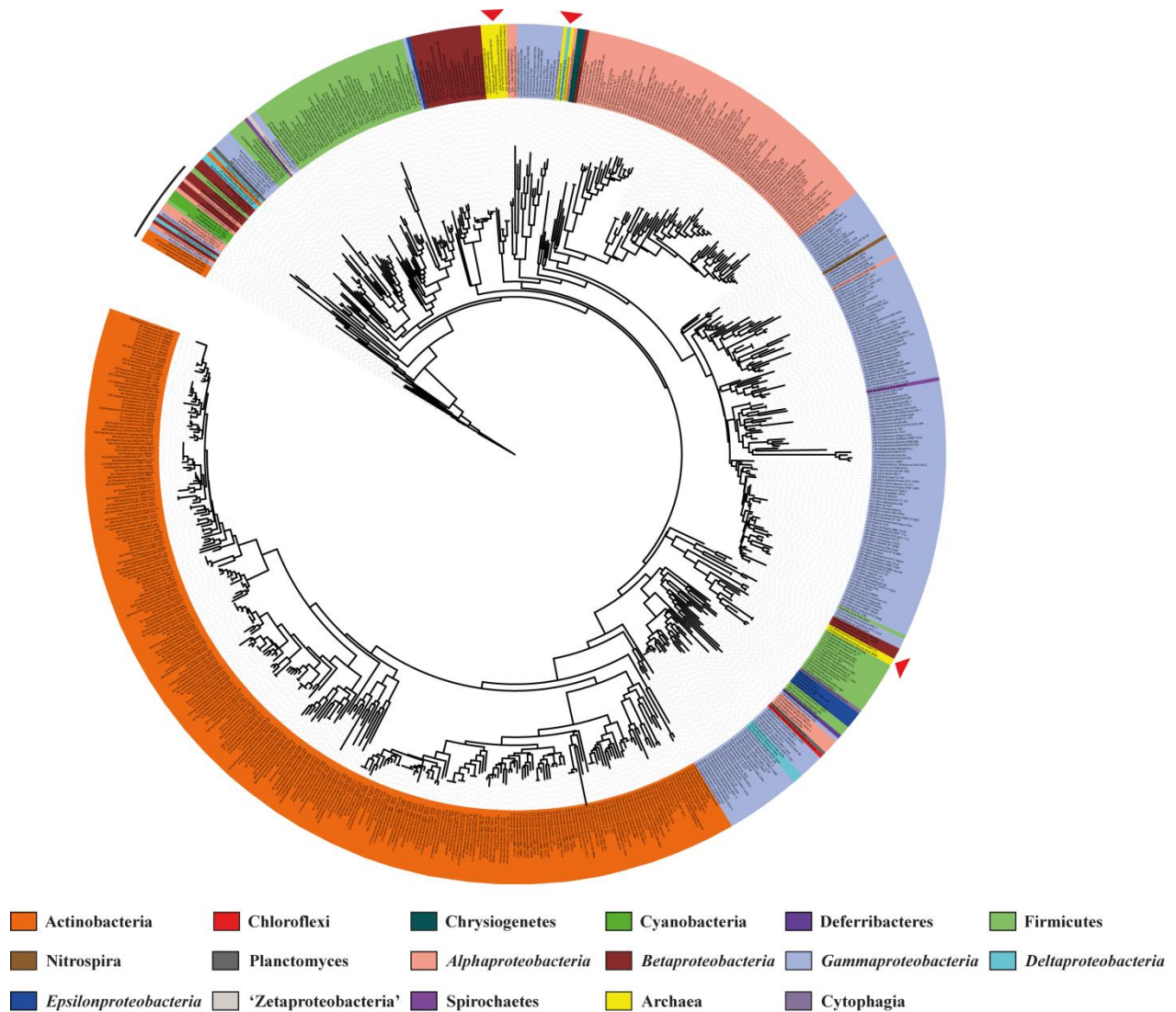


Fig. S1. Phylogenetic tree of EctC-type proteins. Based on an amino acid sequence alignment of 723 EctC-type proteins, a rooted phylogenetic tree was constructed with the iTOL program (Letunic and Bork, 2011). The three regions in the phylogenetic tree populated by archaeal EctC proteins are highlighted by red arrow-heads. The position of those 24 EctC-type proteins that originate from bacteria that lack identifiable *ectAB* genes are marked above the phylogenetic tree by a black bar.

Supplemental Figure S2

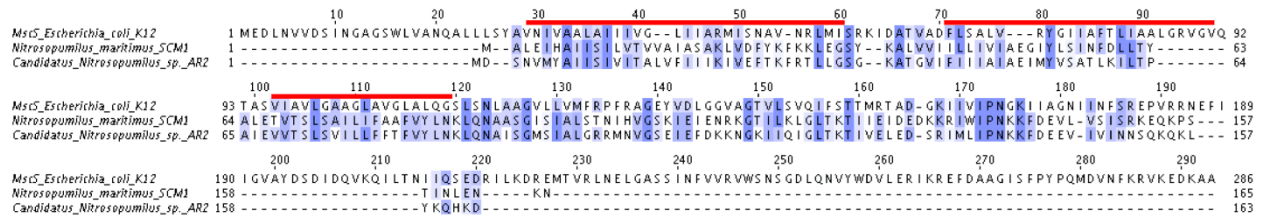


Fig. S2. Alignment of the amino acid sequence of the *E. coli* MscS protein with those MscS-type proteins whose corresponding genes are located next to ectoine/hydroxyectoine biosynthetic gene clusters in the genomes of “*Ca. Nitrosopumilus maritimus*” SCM 1 and “*Ca. Nitrosopumilus*” sp. AR2 (see Fig. 1B in the main text). Red bars indicate the membrane-spanning segments of the *E. coli* MscS protein.

Supplemental Table S1

Table S1. Amino acid sequence relatedness of the ectoine biosynthetic proteins from "*Ca. Nitrosopumilus*" strains to other ectoine biosynthetic enzymes.

EctA	" <i>Ca. Nitrosopumilus maritimus</i> " SCM1	" <i>Ca. Nitrosopumilus</i> " sp. SJ	" <i>Ca. Nitrosopumilus koreensis</i> " ARI	" <i>Ca. Nitrosopumilus</i> " sp. AR2
1	<i>Nitrosopumilus maritimus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)
2	Multispecies <i>Nitrosopumilus</i> (82%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (83%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (83%)	<i>Nitrosopumilus maritimus</i> (82%)
3	Marine Group I thaumarchaeote SCGC AAA799-N04 (80%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (83%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (83%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (78%)
4	Marine Group I thaumarchaeote SCGC AAA799-P11 (80%)	Multispecies <i>Nitrosopumilus</i> (73%)	Multispecies <i>Nitrosopumilus</i> (73%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (78%)
5	Multispecies <i>Nitrosopumilus</i> (74%)	<i>Nitrosopumilus maritimus</i> (74%)	<i>Nitrosopumilus maritimus</i> (74%; 6e-84)	Multispecies <i>Nitrosopumilus</i> (73%)
6	Marine Group I thaumarchaeote SCGC AAA799-E16 (79%)	Marine Group I thaumarchaeote SCGC AAA799-E16 (81%)	Marine Group I thaumarchaeote SCGC AAA799-E16 (81%)	Marine Group I thaumarchaeote SCGC AAA799-E16 (78%)
7	<i>Bacillus thermotolerans</i> (41%)	<i>Glycomyces arizonensis</i> (42%)	<i>Glycomyces arizonensis</i> (42%)	<i>Dehalogenimonas lykanthroporepellens</i> (46%)
8	<i>Alkalibacillus haloalkaliphilus</i> (44%)	<i>Nocardia brasiliensis</i> (40%)	<i>Nocardia brasiliensis</i> (40%)	<i>Paenibacillus daejeonensis</i> (41%)
9	<i>Oceanobacillus picturæ</i> (47%)	<i>Nocardia</i> sp. W9851 (40%)	<i>Nocardia</i> sp. W9851 (40%)	<i>Methylobacter marinus</i> (39%)
10	<i>Oceanobacillus picturæ</i> (47%)	<i>Marinospirillum minutulum</i> (43%)	<i>Marinospirillum minutulum</i> (43%)	zeta proteobacterium SCGC AB-137-C09 (42%)
EctB	" <i>Ca. Nitrosopumilus maritimus</i> " SCM1	" <i>Ca. Nitrosopumilus</i> " sp. SJ	" <i>Ca. Nitrosopumilus koreensis</i> " ARI	" <i>Ca. Nitrosopumilus</i> " sp. AR2
1	<i>Nitrosopumilus maritimus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (99%)	Multispecies <i>Nitrosopumilus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)
2	Multispecies <i>Nitrosopumilus</i> (100%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (86%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (86%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (80%)
3	Marine Group I thaumarchaeote SCGC AAA799-N04 (83%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (86%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (86%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (79%)
4	Marine Group I thaumarchaeote SCGC AAA799-P11 (82%)	<i>Nitrosopumilus maritimus</i> (82%)	<i>Nitrosopumilus maritimus</i> (82%)	<i>Nitrosopumilus maritimus</i> (79%)
5	Multispecies <i>Nitrosopumilus</i> (79%)	Multispecies <i>Nitrosopumilus</i> (77%)	Multispecies <i>Nitrosopumilus</i> (78%)	Multispecies <i>Nitrosopumilus</i> (78%)
6	<i>Neptuniibacter caesariensis</i> (59%)	Marine Group I thaumarchaeote SCGC AAA799-E16 (85%)	Marine Group I thaumarchaeote SCGC AAA799-E16 (85%)	<i>Neptuniibacter caesariensis</i> (59%)
7	<i>Marinomonas</i> sp. D104 (59%)	<i>Marinomonas</i> sp. MWYLI (59%)	<i>Marinomonas</i> sp. MWYLI (59%)	<i>Marinomonas</i> sp. MWYLI (58%)
8	Marine Group I thaumarchaeote SCGC AAA799-E16 (84%)	gamma proteobacterium BDW918 (58%)	gamma proteobacterium BDW918 (58%)	<i>Methylomicrobium kenyense</i> (59%)
9	<i>Enterovibrio norvegicus</i> (60%)	<i>Marinomonas ushuaiensis</i> (59%)	<i>Marinomonas ushuaiensis</i> (59%)	<i>Amphritea japonica</i> (58%)
10	<i>Enterovibrio calviensis</i> (59%)	<i>Marinomonas</i> sp. D104 (58%)	<i>Marinomonas</i> sp. D104 (59%)	<i>Endozoicomonas elysicola</i> (58%)
EctC	" <i>Ca. Nitrosopumilus maritimus</i> " SCM1	" <i>Ca. Nitrosopumilus</i> " sp. SJ	" <i>Ca. Nitrosopumilus koreensis</i> " ARI	" <i>Ca. Nitrosopumilus</i> " sp. AR2
1	<i>Nitrosopumilus maritimus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)
2	Multispecies <i>Nitrosopumilus</i> (82%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (86%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (86%)	Multispecies <i>Nitrosopumilus</i> (72%)
3	Marine Group I thaumarchaeote SCGC AAA799-P11 (83%)	<i>Nitrosopumilus maritimus</i> (82%)	<i>Nitrosopumilus maritimus</i> (82%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (72%)
4	Marine Group I thaumarchaeote SCGC AAA799-N04 (83%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (85%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (85%)	<i>Nitrosopumilus maritimus</i> (71%)
5	Multispecies <i>Nitrosopumilus</i> (71%)	Multispecies <i>Nitrosopumilus</i> (72%)	Multispecies <i>Nitrosopumilus</i> (72%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (71%)
6	<i>Sneathiella glossodoripedis</i> (60%)	<i>Bacillus fordii</i> (66%)	<i>Bacillus fordii</i> (66%)	<i>Hahella chejuensis</i> (63%)
7	<i>Desulfarculus baarsii</i> (59%)	<i>Hahella chejuensis</i> (60%)	<i>Hahella chejuensis</i> (60%)	<i>Hahella ganghwensis</i> (63%)
8	<i>Hahella chejuensis</i> (60%)	<i>Sneathiella glossodoripedis</i> (60%)	<i>Sneathiella glossodoripedis</i> (60%)	gamma proteobacterium BDW918 (61%)
9	<i>Nitricola</i> sp. AK23 (60%)	<i>Paenibacillus pinihumi</i> (60%)	<i>Paenibacillus pinihumi</i> (60%)	<i>Desulfarculus baarsii</i> (57%)
10	<i>Vibrio genomsp. F10</i> (61%)	<i>Nitrococcus mobilis</i> (62%)	<i>Nitrococcus mobilis</i> (62%)	<i>Sneathiella glossodoripedis</i> (57%)
EctD	" <i>Ca. Nitrosopumilus maritimus</i> " SCM1	" <i>Ca. Nitrosopumilus</i> " sp. SJ	" <i>Ca. Nitrosopumilus koreensis</i> " ARI	" <i>Ca. Nitrosopumilus</i> " sp. AR2
1	<i>Nitrosopumilus maritimus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)	Multispecies <i>Nitrosopumilus</i> (100%)
2	Marine Group I thaumarchaeote SCGC AAA799-P11 (82%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (88%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (88%)	Multispecies <i>Nitrosopumilus</i> (77%)
3	Multispecies <i>Nitrosopumilus</i> (81%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (87%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (87%)	Marine Group I thaumarchaeote SCGC AAA799-P11 (66%)
4	Marine Group I thaumarchaeote SCGC AAA799-N04 (82%)	<i>Nitrosopumilus maritimus</i> (81%)	<i>Nitrosopumilus maritimus</i> (81%)	Marine Group I thaumarchaeote SCGC AAA799-N04 (77%)
5	Multispecies <i>Nitrosopumilus</i> (76%)	Multispecies <i>Nitrosopumilus</i> (77%)	Multispecies <i>Nitrosopumilus</i> (87%)	<i>Nitrosopumilus maritimus</i> (76%)
6	<i>Nitrosococcus halophilus</i> (59%)	<i>Nitrosococcus halophilus</i> (58%)	<i>Nitrosococcus halophilus</i> (58%)	<i>Nitrosococcus halophilus</i> (59%)
7	<i>Nitrosococcus oceani</i> (57%)	<i>Nitrosococcus watsonii</i> (56%)	<i>Nitrosococcus watsonii</i> (56%)	<i>Nitrosococcus watsonii</i> (57%)
8	<i>Paenibacillus senegalensis</i> (58%)	<i>Nitrosococcus oceani</i> (56%)	<i>Nitrosococcus oceani</i> (56%)	<i>Nitrosococcus oceani</i> (57%)
9	<i>Nitrosococcus watsonii</i> (56%)	<i>Oxalobacteraceae bacterium IMCC9480</i> (58%)	<i>Oxalobacteraceae bacterium IMCC9480</i> (58%)	<i>Paenibacillus senegalensis</i> (57%)
10	<i>Pusillimonas</i> sp. I7-7 (57%)	<i>Janthinobacterium</i> sp. Marseille (56%)	<i>Janthinobacterium</i> sp. Marseille (56%)	<i>Bacillus</i> sp. L1(2012) (57%)

Amino acid sequence similarity of the ectoine biosynthetic enzymes from the four potential ectoine-producing "*Ca. Nitrosopumilus*" strains to other ectoine biosynthetic enzymes were assessed bioinformatically. The amino acid sequences of the EctA, EctB, EctC and EctD proteins of each strain were used as a query in BLAST searches. The top ten hits from each BLAST search are shown and the amino acid similarities are indicated in brackets.

Supplemental Table S2

Table S2. Distribution of mechanosensitive channels in potential ectoine-producing *Archaea*.

Organism (Accession no.)	MscS-type protein	MscL-type protein
"<i>Ca. Nitrosopumilus maritimus</i>" SCM1 (NC_010085.1)	ABX12867; ABX13231; ABX13669; ABX13233; ABX13238	not present
"<i>Ca. Nitrosopumilus</i>" sp. SJ (GCA_000328945.1)	WP_014963370; WP_014964111; WP_014963465	not present
"<i>Ca. Nitrosopumilus koreensis</i>" AR1 (GCA_000299365.1)	AFS80985; AFS81736; AFS81081	not present
"<i>Ca. Nitrosopumilus</i>" sp. AR2 (GCA_000299395.1)	WP_014965293; WP_014966054; WP_014964462	not present
Marine Group I thaumarchaeote SCGC AAA799-N04 (GCA_000722915.1)	WP_048069171; WP_048079437	not present
Marine Group I thaumarchaeote SCGC AAA799-P11 (GCA_000746685.1)	KFM18867; WP_048070662	not present
Marine Group I thaumarchaeote SCGC RSA3 (GCA_000746745.1)	WP_048069171; WP_048083714	not present
Marine Group I thaumarchaeote SCGC AAA799-E16 (GCA_000724145.1)	WP_048088428	not present
<i>Methanotherx harundinacea</i> 6Ac	WP_014587606; WP_014587910; WP_014586447; WP_014586460	AET64001; WP_048144789
<i>Methanotherx concilii</i> GP-6	WP_013718893; AEB69207; WP_052297576; WP_013720440	WP_013718432
<i>Methanobacterium formicicum</i> DSM1535	WP_048084718; WP_048072178; WP_004032010	not present
<i>Methanobacterium formicicum</i> BRM9	WP_048084718; WP_048072178; WP_004032010	not present

The presence of potential mechanosensitive channels in ectoine-producing *Archaea* was analyzed bioinformatically. The amino acid sequences of MscS and MscL proteins from *Escherichia coli* were used as the search queries. Retrieved hits by the BLAST search of potential mechanosensitive channels are indicated by their NCBI accession numbers.