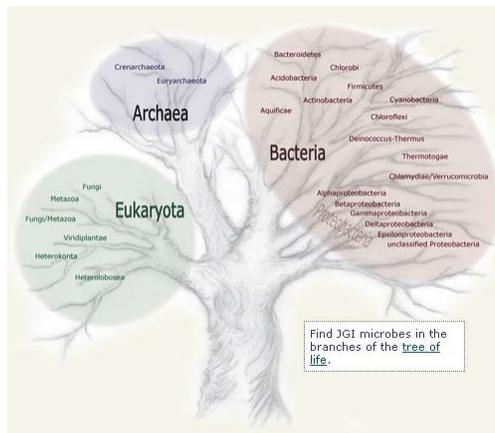


# Archaea Genome Analysis at JGI

Iain Anderson  
Genome Biology Program  
January 8, 2008

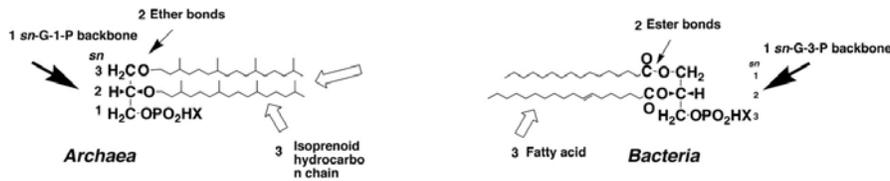
## Third domain of life



Discovered in 1977 by Carl Woese

## Common characteristics

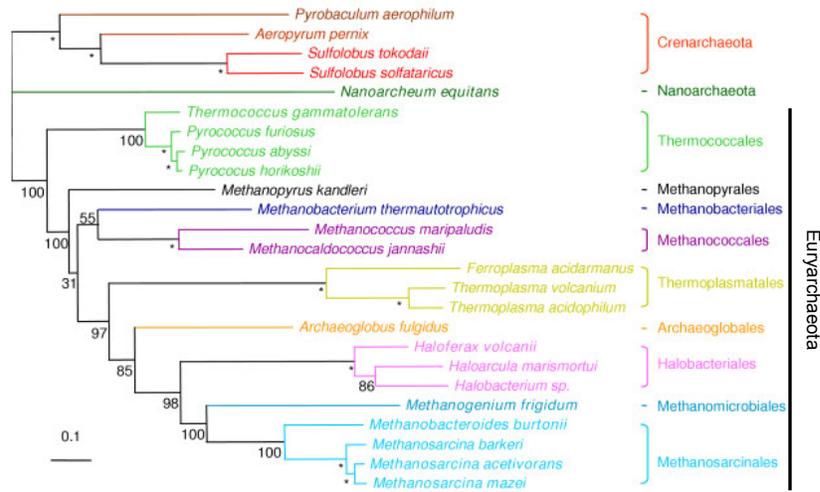
- Transcription/translation similar to eukaryotes, not bacteria (TFIID, TFIIB, no sigma factors)
- Ether-linked branched membrane lipids
- Lipid backbone is glycerol-1P, not glycerol-3P
- No peptidoglycan (some have pseudomurein)
- Many are extremophiles – salt, temperature, pH



From Koga and Morii, 2007

## Phylogenetic tree of Archaea

from Brochier et al., 2005



Archaeal "Tree of Life" project to fill in gaps – 11 archaea

# Methanogens

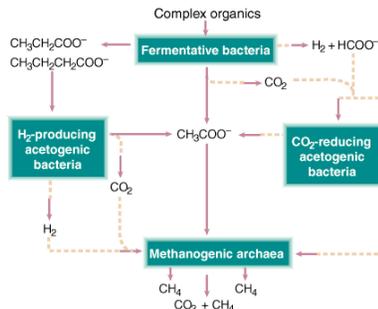


Figure from Ferry, 1997 Science

- Catalyze final steps in anaerobic degradation of organic matter
- Produce 400,000,000 metric tons of methane each year
- Methane sources include wetlands, rice paddies, cows/sheep, biomass burning, automobiles

Methane importance : biofuel and greenhouse gas (21X more potent than CO<sub>2</sub>)

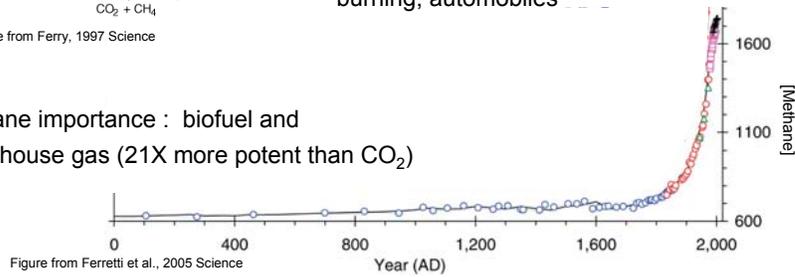


Figure from Ferretti et al., 2005 Science

## Four methanogens in Archaea tree of life project

### *Methanocorpusculum labreanum*, *Methanoculleus marisnigri*

- involved in biomethanation processes (anaerobic wastewater reactors, landfills, etc.)
- Methanomicrobiales – no genomes yet

### *Methanobacterium formicicum*

- mesophile, important for comparison with thermophilic relatives to determine protein adaptations to high temperature

### *Methanohalobium evestigatum*

- extreme halophilic methanogen, important for comparison with other methanogens to determine protein adaptations to high salt

## Hyperthermophiles

Importance: production of heat-stable enzymes for industry (glycosidases, proteases)

- *Thermophilum pendens* – cellulase, amylase
- *Staphylothermus marinus*
- *Desulfurococcus fermentans* – cellulase, glycosidases



## Thermoacidophiles

Importance: bioleaching agents for sulfide ores

- *Acidianus* sp. JP7 – grows at pH as low as 0.3



## Halophiles

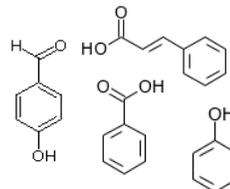
Importance: production of salt-stable enzymes for Industry, bioremediation of saline environments

- *Halorubrum lacusprofundi* – Antarctic lake, first psychrotolerant halophile to be sequenced
- “*Halococcoides aestuarii*” – low-salt-adapted halophile, important for comparative studies



## Archaeoglobales

- *Ferroplasma placidus* – anaerobically oxidizes aromatic compounds (first archaeon found to do this)



## Crenarchaeota physiology

Aerobes, facultative, anaerobes

Heterotrophs – grow on peptides, complex media

many reduce sulfur producing H<sub>2</sub>S

(similar to *Pyrococcus* sp. from Euryarchaeota)

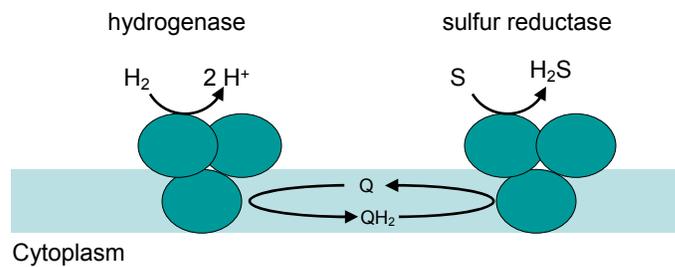
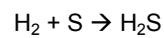
Autotrophs – hydrogen/sulfur, hydrogen/nitrate

Carbon fixation - reverse TCA cycle, hydroxypropionate cycle, other?

(same phenotype, different genotype)

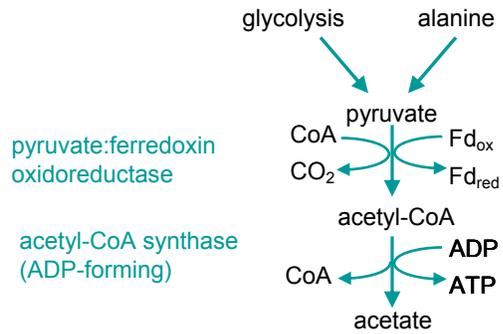
Much is unknown, because Crenarchaeota have not been studied as much as other archaea.

## Hydrogen-sulfur autotrophy



*Acidianus ambivalens*, *Pyrodicticum abyssi*, *Ignicoccus hospitalis*  
*Wolinella succinogenes*

## Heterotrophy



Characterized in *Pyrococcus furiosus* (Euryarchaeota)

Similar pathways for degradation of branched-chain and aromatic amino acids and glutamate

## Published Crenarchaeal genomes

### Thermoproteales

*Pyrobaculum aerophilum*  
*Thermofilum pendens*

### Sulfolobales

*Sulfolobus acidocaldarius*  
*Sulfolobus solfataricus*  
*Sulfolobus tokodaii*  
*Acidianus* sp. JP7

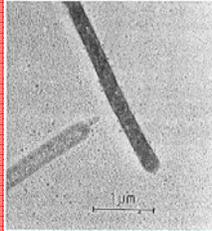
### Desulfurococcales

*Aeropyrum pernix*  
*Hyperthermus butylicus*  
*Staphylothermus marinus*

### Cenarchaeales

*Cenarchaeum symbiosum*

## *Thermofilum pendens*



Optimal growth at 85-90°C, pH 5.0

Sulfur-dependent anaerobe (*P. aerophilum* is facultative, inhibited by sulfur)

Cells can be longer than 100 μm

No septum observed, “golf club” shaped cells

Requires lipid extract of *Thermoproteus tenax* for growth

Grows on peptide mixtures

## Genes found in *T. pendens* but not in other Crenarchaeota using IMG phylogenetic profiler

Find Genes In*	With Homologs In	Without Homologs In	Ignoring	Taxon Name
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Archaea
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Crenarchaeota
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Aeropyrum
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Aeropyrum pernix K1</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Hyperthermus
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Hyperthermus butylicus DSM 5456</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Staphylothermus
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Staphylothermus marinus F1</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Metallosphaera
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Metallosphaera sedula DSM 5348</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sulfolobus
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Sulfolobus acidocaldarius DSM 639</a> [F]
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Sulfolobus solfatarius P2</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<a href="#">Sulfolobus tokodaii 7</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Thermofilum
<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<a href="#">Thermofilum pendens Hrk 5</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Pyrobaculum
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<a href="#">Pyrobaculum aerophilum IM2</a> [F]
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Pyrobaculum islandicum DSM 4184</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Caldivirga
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Caldivirga maquilingensis IC-167</a> [D]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Pyrobaculum
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Pyrobaculum arsenaticum DSM 13514</a> [F]
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Pyrobaculum caldifontis JCM 11548</a> [F]

Phylogenetic Profiler Results

590 gene(s) retrieved

Summary Statistics

Processing 12 comparison(s)  
 1883 genes found for genome (bin) of interest, *Thermofilum pendens* Hrk 5  
 1120 genes remaining after subtracting genes with homologs in *Aeropyrum pernix* K1  
 856 genes remaining after subtracting genes with homologs in *Pyrobaculum aerophilum* IM2  
 803 genes remaining after subtracting genes with homologs in *Sulfolobus acidocaldarius* DSM 639  
 762 genes remaining after subtracting genes with homologs in *Sulfolobus solfataricus* P2  
 745 genes remaining after subtracting genes with homologs in *Sulfolobus tokodaii* 7  
 697 genes remaining after subtracting genes with homologs in *Pyrobaculum islandicum* DSM 4184  
 658 genes remaining after subtracting genes with homologs in *Hyperthermus butylicus* DSM 5456  
 641 genes remaining after subtracting genes with homologs in *Pyrobaculum caldifontis* JCM 11548  
 611 genes remaining after subtracting genes with homologs in *Staphylothermus marinus* F1  
 604 genes remaining after subtracting genes with homologs in *Metallosphaera sedula* DSM 5348  
 599 genes remaining after subtracting genes with homologs in *Pyrobaculum arsenaticum* DSM 13514  
 590 genes remaining after subtracting genes with homologs in *Ignicoccus hospitalis* KIN4/1

	Number	% of Total
<b>Genes total number</b>	590	100.00%
COG	151	25.59%
Enzyme	4	0.68%
Pfam	108	18.31%
InterPro	192	32.54%
No Functional Hit	365	61.86%
Unique In IMG	53	8.98%

Add Selected to Gene Cart    Select All    Clear All

Select	Result Row	Gene Object ID	Locus tag	Gene Name	Length	COG	Enzyme	Pfam	InterPro	Unique In IMG
<input type="checkbox"/>	1	639772438	Tpen_1840	hypothetical protein	555aa	-	-	-	-	No
<input type="checkbox"/>	2	639772439	Tpen_1841	hypothetical protein	98aa	-	-	-	-	No
<input type="checkbox"/>	3	639772440	Tpen_1842	Absorptive infection protein	314aa	-	-	pfam02517	IPR003675	No
<input type="checkbox"/>	4	639772441	Tpen_1843	hypothetical protein	186aa	-	-	-	-	No
<input type="checkbox"/>	5	639772442	Tpen_1844	hypothetical protein	72aa	-	-	-	-	No
<input type="checkbox"/>	6	639772443	Tpen_1845	hypothetical protein	122aa	-	-	-	-	No
<input type="checkbox"/>	7	639772444	Tpen_1846	hypothetical protein	123aa	-	-	-	-	No
<input type="checkbox"/>	8	639772445	Tpen_1847	hypothetical protein	128aa	-	-	-	-	No
<input type="checkbox"/>	9	639772446	Tpen_1848	hypothetical protein	132aa	-	-	-	-	No
<input type="checkbox"/>	10	639772447	Tpen_1849	hypothetical protein	77aa	-	-	-	-	No
<input type="checkbox"/>	11	639772448	Tpen_1850	hypothetical protein	118aa	-	-	-	-	Yes
<input type="checkbox"/>	12	639772449	Tpen_1851	hypothetical protein	65aa	-	-	-	-	No
<input type="checkbox"/>	13	639772450	Tpen_1852	hypothetical protein	257aa	-	-	-	IPR010985	No
<input type="checkbox"/>	14	639772451	Tpen_1853	hypothetical protein	232aa	-	-	-	-	No
<input type="checkbox"/>	15	639772452	Tpen_1854	hypothetical protein	132aa	-	-	-	-	No
<input type="checkbox"/>	16	639772453	Tpen_1855	hypothetical protein	126aa	-	-	-	-	No
<input type="checkbox"/>	17	639772454	Tpen_1856	hypothetical protein	150aa	-	-	-	IPR002016	No

Domains more abundant in *T. pendens* than in other Crenarchaeota using IMG abundance profile search

Find Functions In*	More Abundant Than In	Less Abundant Than In	Ignoring	Taxon Name
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Archaea</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Crenarchaeota</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Aeropyrum</b>
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Aeropyrum pernix</i> K1 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Hyperthermus</b>
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Hyperthermus butylicus</i> DSM 5456 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Staphylothermus</b>
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Staphylothermus marinus</i> F1 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Metallosphaera</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<i>Metallosphaera sedula</i> DSM 5348 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Sulfolobus</b>
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Sulfolobus acidocaldarius</i> DSM 639 [F]
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Sulfolobus solfataricus</i> P2 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Sulfolobus tokodaii</i> 7 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Thermofilum</b>
<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Thermofilum pendens</i> Hrk 5 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Pyrobaculum</b>
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Pyrobaculum aerophilum</i> IM2 [F]
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Pyrobaculum islandicum</i> DSM 4184 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Caldivirga</b>
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<i>Caldivirga maquilingensis</i> IC-167 [D]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Pyrobaculum</b>
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Pyrobaculum arsenaticum</i> DSM 13514 [F]
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<i>Pyrobaculum caldifontis</i> JCM 11548 [F]

Abundance Profile Search Results

Add Selected to Function Cart    Select All    Clear All

Red - Less abundant.  
Green - More abundant.

Selection	COG ID	COG Name	Thermotoga parvulus T15	Halobacterium salinarum R1	Pyrobaculum aerophilum Str. 82	Sulfolobus solfataricus DSM 639	Sulfolobus solfataricus P2	Sulfolobus solfataricus Str. 2	Pyrobaculum ferroplasma DSM 4184	Hypothermus barophilus DSM 5456	Pyrobaculum caldiatum JCM 11148	Halobacterium salinarum R1	Metabolite transferase DSM 2348
<input type="checkbox"/>	COG0017	AspartyltRNA synthetase	4	1	2	1	1	2	1	2	3	3	1
<input type="checkbox"/>	COG0053	Glyceraldehyde-3-phosphate dehydrogenase	2	1	1	1	1	1	1	1	1	1	1
<input type="checkbox"/>	COG0170	Dialcohol dehydrogenase	2	0	1	0	0	0	1	1	1	0	0
<input type="checkbox"/>	COG0207	Thymidylate synthase	1	0	0	0	0	0	0	0	0	0	0
<input type="checkbox"/>	COG0212	5-formyltetrahydrofolate cycloligase	2	1	1	1	1	1	1	1	1	1	1
<input type="checkbox"/>	COG0258	5'-3' exonuclease (including N-terminal domain of PolI)	2	1	1	1	1	1	1	1	1	1	1
<input type="checkbox"/>	COG0309	Hydrogenase maturation factor	3	1	1	2	2	2	1	2	2	2	2
<input type="checkbox"/>	COG0364	Glycosidase	2	0	1	0	0	0	1	0	1	1	0
<input type="checkbox"/>	COG0395	ABC-type sugar transport system, permease component	6	2	2	0	4	1	0	2	2	4	2
<input type="checkbox"/>	COG0436	Aspartate/tyrosine/aromatic amino-transferase	5	2	2	2	2	1	2	1	2	2	1
<input type="checkbox"/>	COG0454	Acetyltransferase	5	2	4	2	2	2	2	2	2	2	2
<input type="checkbox"/>	COG0462	Phosphoenolpyruvate synthetase	2	1	1	1	1	1	1	1	1	1	1
<input type="checkbox"/>	COG0474	Calcium transport ATPase	2	0	0	0	0	0	0	0	0	0	1
<input type="checkbox"/>	COG0567	Amino acid transporter and related factors	2	0	0	0	0	0	0	0	0	1	0
<input type="checkbox"/>	COG0619	ABC-type cobalt transport system, permease component CheQ and related transporters	7	1	0	1	2	1	0	1	0	2	2
<input type="checkbox"/>	COG0680	N <sub>2</sub> F <sub>2</sub> hydrogenase maturation factor	2	0	0	0	1	0	1	1	1	1	1
<input type="checkbox"/>	COG0694	Ribose 5-phosphate isomerase FpbI	1	0	0	0	0	0	0	0	0	0	0
<input type="checkbox"/>	COG0842	ABC-type malolactate transport system, permease component	5	2	1	1	2	1	1	0	1	0	2
<input type="checkbox"/>	COG1000	NADH-dependent oxidoreductase subunit 5 (chain L) [Molecular Weight transporter, MtbA subunit]	3	1	1	1	1	1	2	1	2	1	1
<input type="checkbox"/>	COG1054	Predicted nucleotide-utilizing enzyme related to methyltransferase: histone methylase MtaA	2	1	1	1	1	1	1	1	1	0	1
<input type="checkbox"/>	COG1061	DNA or RNA helicase of superfamily II	4	1	2	2	2	2	2	1	3	2	3
<input type="checkbox"/>	COG1070	Sugar (pentose and hexose) kinase	3	1	0	1	0	0	0	0	0	0	0
<input type="checkbox"/>	COG1080	Phosphoenolpyruvate-protein kinase (PTS system EI component in bacteria)	1	0	0	0	0	0	0	0	0	0	0
<input type="checkbox"/>	COG1085	Galactose-1-phosphate uridylyltransferase	2	0	1	0	0	0	0	0	1	1	0
<input type="checkbox"/>	COG1118	ABC-type malate/oxalate transport system, ATPase component	1	0	0	0	0	0	0	0	0	0	0

CDS	COG	Function
<b>COGs not found in any other sequenced Archaea</b>		
Tpen_1241	0698	Ribose 5-phosphate isomerase rpiB
Tpen_1297	3525	Glycosyl hydrolase family 20
Tpen_1097	3444	Phosphotransferase system IIB subunit
Tpen_1100	3715	Phosphotransferase system IIC subunit
Tpen_1100	3716	Phosphotransferase system IID subunit
Tpen_1090	4821	Phosphosugar binding protein, SIS domain
<b>COGs not found in any other sequenced Crenarchaeota</b>		
Tpen_1155	1554	Glycoside hydrolase family 65
Tpen_0948	0207	Thymidylate synthase
Tpen_1467	5598	Trimethylamine:corrinoid methyltransferase
Tpen_1211		Monomethylamine:corrinoid methyltransferase
Tpen_1092	1080	Phosphoenolpyruvate-protein kinase (Enzyme I of PTS)
Tpen_1091	1925	Phosphotransferase system HPr protein
Tpen_1098	2893	Phosphotransferase system IIA component
Tpen_1491	1268	Biotin transporter bioY
Tpen_0929	3601	Riboflavin transporter
Tpen_0191	2116	Formate transporter
Tpen_1479	2060	Potassium-transporting ATPase, A chain
Tpen_1480	2216	Potassium-transporting ATPase, B chain
Tpen_1481	2156	Potassium-transporting ATPase, c chain

Genes present in all Crenarchaeota except *T. pendens*

COG	Function
0214	Pyridoxine biosynthesis enzyme (yaaD)
0311	Glutamine amidotransferase involved in pyridoxine synthesis (yaaE)
0413	Ketopantoate hydroxymethyltransferase
0452	Phosphopantothenoylcysteine synthetase/decarboxylase
0108	3,4-dihydroxy-2-butanone 4-phosphate synthase
1985	Pyrimidine reductase, riboflavin biosynthesis
0054	Riboflavin synthase beta chain
0163	3-polyprenyl-4-hydroxybenzoate decarboxylase
1635	Flavoprotein involved in thiazole biosynthesis
0112	Glycine/serine hydroxymethyltransferase
0189	Glutathione synthase/Ribosomal protein S6P modification enzyme/L-2-aminoacidate N-acetyltransferase
0105	Nucleoside diphosphate kinase
2046	Sulfate adenyltransferase
1650	Uncharacterized protein conserved in archaea
1701	Uncharacterized protein conserved in archaea

Function profile - 125  
COGs involved in  
biosynthesis

*T. pendens* – 11  
*C. trachomatis* – 27  
*R. prowazekii* - 15

*T. pendens* has  
fewer biosynthetic  
enzymes than  
obligate intracellular  
pathogens

Function Profile

hint: Mouse over genome abbreviation to see genome name  
(Cell coloring is highlighting of gene counts: white = 0, blue <= 4, yellow >= 5)  
Click on column name to sort

Function ID	Name	Asn	Arg	Sta	Met	Ser	Sul	Thr	Val	Pro	Leu	Ile	Val	Met	Pro	Val	Pro
		111	546	411	538	639	822	77	55	114	131	118					
COG0001	Glutamate-1-semialdehyde aminotransferase	2	2	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0002	Acetylglutamate semialdehyde dehydrogenase	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0007	Uroporphyrinogen-III methylase	1	1	0	2	2	2	2	0	2	2	2	2	2	2	2	2
COG0015	Adenylosuccinate lyase	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0026	Phosphoribosylaminoimidazole carboxylase (HCAIR synthetase)	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0027	Formate-dependent phosphoribosylglycinamide formyltransferase (GAR transferoylase)	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
COG0028	Thiamine pyrophosphate-requiring enzymes (acetylactate synthase, pyruvate dehydrogenase (cytochrome), glyoxylate carboxylase, phosphoenolpyruvate decarboxylase)	1	0	0	6	7	7	4	0	1	1	2	2	2	2	2	2
COG0034	Glutamate phosphoribosylpyrophosphate amidotransferase	0	0	0	2	2	2	2	0	2	2	2	2	2	2	2	2
COG0040	ATP phosphoribosyltransferase	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0041	Phosphoribosylcarboxyaminoimidazole (HCAIR) mutase	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0044	Dihydroorotate and related cyclic amidohydrolases	1	1	2	1	1	1	4	1	1	1	1	1	1	1	1	1
COG0046	Phosphoribosylformylglycinamide (FGAM) synthase, synthetase domain	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0047	Phosphoribosylformylglycinamide (FGAM) synthase, glutamine amidotransferase domain	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0054	Riboflavin synthase beta chain	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0059	Ketol-acid reductoisomerase	0	0	0	1	1	3	1	0	1	1	1	1	1	1	1	1
COG0065	3-isopropylmalate dehydratase large subunit	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0066	3-isopropylmalate dehydratase small subunit	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0077	Prephenate dehydratase	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0082	Chorismate synthase	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0083	Homoserine kinase	1	3	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0104	Adenylosuccinate synthase	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0106	Phosphoribosylformamino-5-aminimidazole carboxamide ribonucleotide (PrFAR) isomerase	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0107	Inositolglycerol phosphate synthase	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0108	3,4-dihydroxy-2-butanone 4-phosphate synthase	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0113	Delta-aminolevulinic acid dehydratase	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0117	Pyrimidine deaminase	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG0128	5-enolpyruvylshikimate-3-phosphate synthase	2	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1
COG0131	Inositolglycerol phosphate dehydratase	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1

Does not make:

Phenylalanine  
Glycine  
Histidine  
Isoleucine  
Lysine  
Leucine  
Arginine  
Threonine  
Valine  
Tryptophan  
Tyrosine

**Purines**

Riboflavin  
Coenzyme A  
Biotin  
Quinolate (for NAD)  
Heme  
Pyridoxine

## Biosynthesis

Partial pathway:

Cysteine  
Glutamate  
Methionine  
Proline  
Serine

Can probably make:

Alanine  
Aspartate  
Asparagine  
Glutamine  
  
Pyrimidines  
Deoxyribonucleotides

Obligate parasites have lower GC than relatives

What about *T. pendens*?

D	C	Genome Name	Genes	GC Perc	Bases
A	F	<a href="#">Thermofilum pendens Hrk 5</a>	1923	0.58	1813393
A	F	<a href="#">Pyrobaculum caldifontis JCM 11548</a>	2210	0.57	2009313
A	F	<a href="#">Aeropyrum pernix K1</a>	2815	0.56	1669695
A	F	<a href="#">Pyrobaculum arsenaticum DSM 13514</a>	2408	0.55	2121076
A	F	<a href="#">Hyperthermus butylicus DSM 5456</a>	1654	0.54	1667163
A	F	<a href="#">Pyrobaculum aerophilum IM2</a>	2617	0.51	2222430
A	F	<a href="#">Pyrobaculum islandicum DSM 4184</a>	2062	0.50	1826402
A	F	<a href="#">Metallosphaera sedula DSM 5348</a>	2345	0.46	2191517
A	D	<a href="#">Caldivirga maquilingensis IC-167</a>	1986	0.43	2077575
A	F	<a href="#">Sulfolobus acidocaldarius DSM 639</a>	2336	0.37	2225959
A	F	<a href="#">Sulfolobus solfataricus P2</a>	3141	0.36	2992245
A	F	<a href="#">Staphylothermus marinus F1</a>	1655	0.36	1570485
A	F	<a href="#">Sulfolobus tokodaii 7</a>	2964	0.33	2694756

*T. pendens* has highest GC of sequenced Crenarchaeota

Obligate parasites have fewer fusions than relatives

What about *T. pendens*?

D	C	Genome Name	Genes	GC Perc	Bases	Fused Genes	Fused Genes Perc
A	F	<a href="#">Thermofilum pendens Hrk 5</a>	1923	0.58	1813393	77	4.00%
A	F	<a href="#">Sulfolobus acidocaldarius DSM 639</a>	2343	0.37	2225959	88	3.76%
A	F	<a href="#">Pyrobaculum aerophilum IM2</a>	2617	0.51	2222430	88	3.36%
A	F	<a href="#">Pyrobaculum islandicum DSM 4184</a>	2062	0.50	1826402	65	3.15%
A	F	<a href="#">Sulfolobus tokodai 7</a>	2982	0.33	2694756	93	3.12%
A	F	<a href="#">Sulfolobus solfataricus P2</a>	3141	0.36	2992245	90	2.87%
A	F	<a href="#">Staphylothermus marinus F1</a>	1655	0.36	1570485	46	2.78%
A	F	<a href="#">Metallospira sedula DSM 5348</a>	2345	0.46	2191517	63	2.69%
A	F	<a href="#">Pyrobaculum caldifontis JCM 11548</a>	2210	0.57	2009313	58	2.62%
A	D	<a href="#">Caldivirga maquilensis IC-167</a>	1986	0.43	2077575	49	2.47%
A	F	<a href="#">Pyrobaculum arsenaticum DSM 13514</a>	2408	0.55	2121076	58	2.41%
A	F	<a href="#">Hyperthermus butylicus DSM 5456</a>	1668	0.54	1667163	39	2.34%
A	F	<a href="#">Aeropyrum pernix K1</a>	2817	0.56	1669695	61	2.17%
A	F	<a href="#">Ignicoccus hospitalis KIN4/I</a>	1494	0.57	1297538	28	1.87%

*T. pendens* has highest percentage of fused genes in Crenarchaeota

Obligate intracellular pathogens

*T. pendens*

reduced genome size

normal genome size

less signal transduction

more signal transduction

fewer DNA repair genes

normal DNA repair genes

high %AT

low %AT

increased noncoding DNA

less noncoding DNA

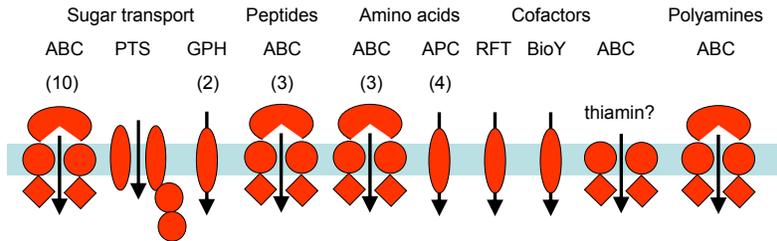
lower fusions

higher fusions

Conclusion: *T. pendens* does not show signs of becoming a parasite.

It has adapted to a nutrient-rich environment

## Transport capabilities



- More sugar ABC transporters than other archaea
- PTS, GPH, thiamin ABC rare in archaea
- RFT, BioY not in other Crenarchaeota

Lack of biosynthesis → Increased transport

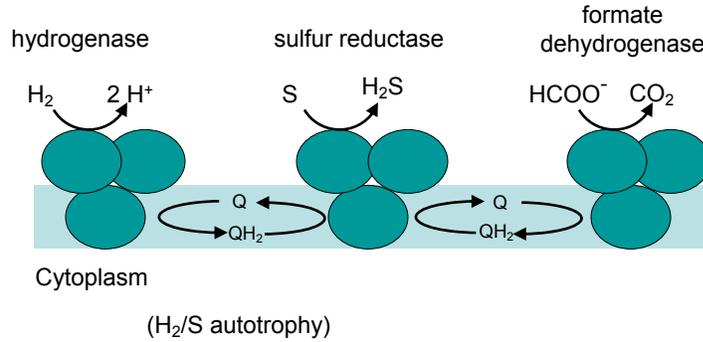
Function profile –  
all published  
archaeal  
genomes

*T. pendens*      *Haloarcula marismortui*

Function ID	Name	<i>T. pendens</i>		<i>Haloarcula marismortui</i>		Hal															
		Gen	Pro	Gen	Pro	Gen	Pro	Gen	Pro	Gen	Pro	Gen	Pro	Gen	Pro	Gen	Pro	Gen	Pro	Gen	Pro
COG1088	Phosphotransferase system (PTS) system EI component (bacteria)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1261	Phosphotransferase system IIC component, glucose/raffinose/1-acetylglucosamine-specific	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1264	Phosphotransferase system IIB component	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1290	Phosphotransferase system, fructose-specific IIC component	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1448	Phosphotransferase system cellobiose-specific component IIB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1445	Phosphotransferase system fructose-specific component IIB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1447	Phosphotransferase system cellobiose-specific component IIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1450	Phosphotransferase system cellobiose-specific component IIC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1761	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG1925	Phosphotransferase system, HPr-related proteins	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG2190	Phosphotransferase system IIA component	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG2211	Phosphotransferase system, mannitol-specific IIBC component	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG2893	Phosphotransferase system, mannose/fructose-specific component IIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG3414	Phosphotransferase system, galactitol-specific IIB component	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG3444	Phosphotransferase system, mannose/fructose/1-acetylglucosamine-specific component IIB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG3715	Phosphotransferase system, mannose/fructose/1-acetylglucosamine-specific component IIC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG3716	Phosphotransferase system, mannose/fructose/1-acetylglucosamine-specific component IID	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG3730	Phosphotransferase system sorbitol-specific component IIC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG3731	Phosphotransferase system sorbitol-specific component IIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG3732	Phosphotransferase system sorbitol-specific component IIB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG3774	Phosphotransferase system, galactitol-specific IIC component	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COG4668	Mannitol/fructose-specific phosphotransferase system, IIA domain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

## Other sources of energy?

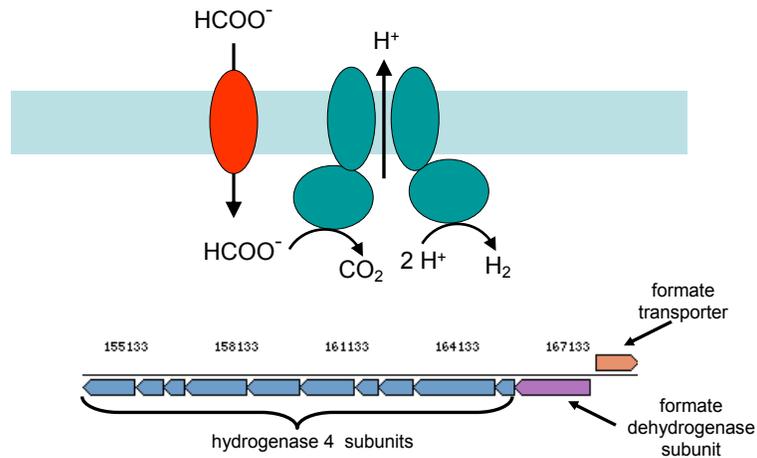
Hydrogen and formate may be prevalent



## Formate without sulfur

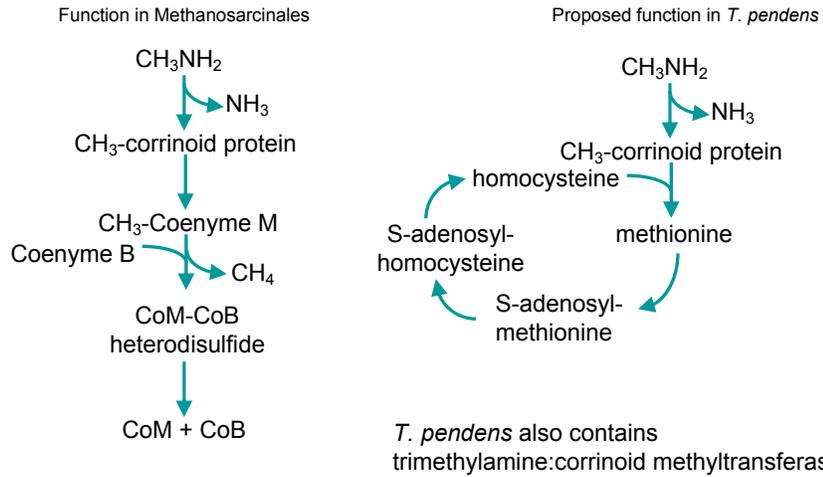
*T. pendens* is the only Crenarchaeote to have a formate transporter and formate hydrogen-lyase.

In *E. coli*, FHL is used when no electron acceptors other than protons are present.



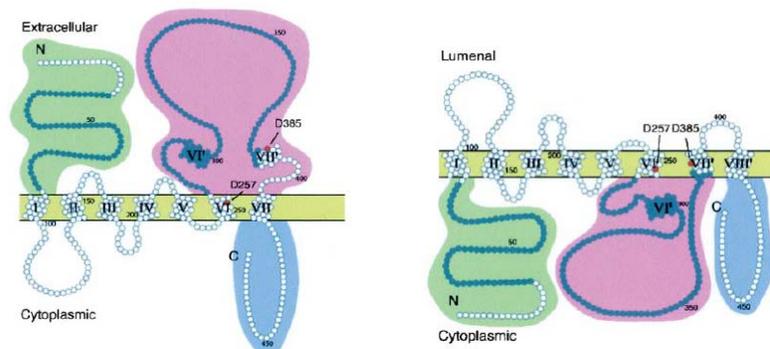
### Monomethylamine:corrinoid methyltransferase

- only found before in Methanosarcinales

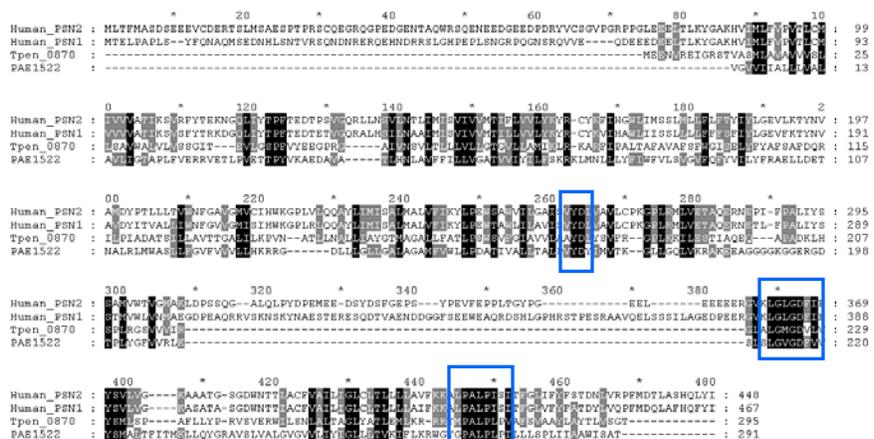


### Presenilin

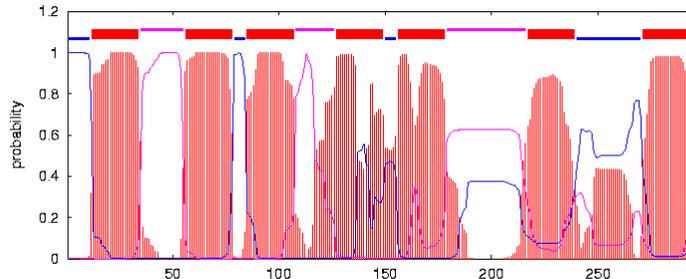
- part of gamma-secretase protease complex in animals
- implicated in Alzheimer's disease
- involved in Notch signaling
- structure is still being debated



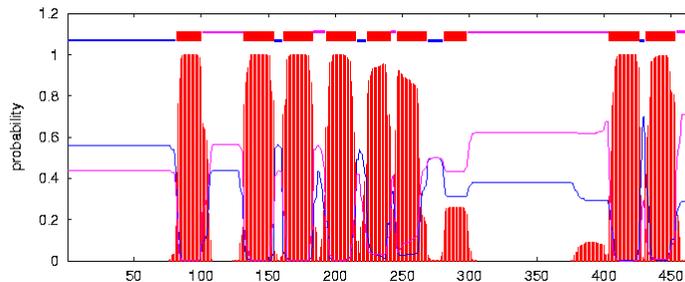
*T. pendens* and *Pyrobaculum aerophilum* have presenilin-related proteins



Tpen\_0870



human presenilin 1



### Predicted highly expressed (PHX) genes in *T. pendens*

1. Training set (Karlín et al., 2005)
2. Run EMBOSS program cusp to get codon usage table
3. Run EMBOSS program cai to get codon adaptation index
4. Top 5% of genes taken as PHX

### Carbohydrate metabolism – 7 PHX genes

- Tpen\_1511 – cytosolic alpha-glucosidase
- Tpen\_1269 – secreted glycosidase
- Tpen\_1458 – cyclomaltodextrinase
- Tpen\_1055 – ABC maltose-binding protein
- Tpen\_1208 – ABC sugar-binding protein
- Tpen\_1257 – ABC sugar-binding protein
- Tpen\_1649 - phosphofructokinase

### Peptide metabolism – 10 PHX genes

- Tpen\_1350 – peptidase
- Tpen\_1225 – peptidase
- Tpen\_1635/1636/1638 – peptide ABC transporter subunits
- Tpen\_1245/1247/1248/1249 – peptide ABC transporter subunits
- Tpen\_1579 – branched chain AA ABC transporter ATPase

### CRISPR-associated – 6+ PHX genes

- Tpen\_1263 – APE2256 family
- Tpen\_1287 – SSO1426 family (RAMP)
- Tpen\_1288 – SSO1426 family (RAMP)
- Tpen\_1316 – Crm2 family
- Tpen\_1342 – APE2256 family
- Tpen\_1356 – Csa2 family

Tpen\_1287-1293 all PHX (7 genes)

Tpen\_1289/1291/1293 RAMP family

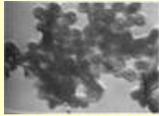
### Other stress-associated – 6 PHX genes

- Tpen\_1301 – nucleotidyltransferase
- Tpen\_0991 – nucleotidyltransferase
- Tpen\_1459 – nucleotidyltransferase
- Tpen\_1430 – peroxiredoxin
- Tpen\_1223 – drug resistance ABC membrane protein
- Tpen\_1136 – drug resistance ABC ATPase

### PHX Conclusions

- Nutrient acquisition (both peptides and carbohydrates) is a high priority for *T. pendens* because of the lack of biosynthetic capabilities
- Stress responses (to viruses, toxins, and oxygen) are another high priority
- Transcription, translation, and replication-associated proteins are not PHX, suggesting fast growth is not a priority

*Staphylothermus  
marinus*



Isolated from shallow heated sediment and from black smoker

Grows up to 98°C

Sulfur-dependent anaerobe

Forms clusters of up to 100 cells

Forms giant cells (15 µm diameter) at high nutrient concentrations

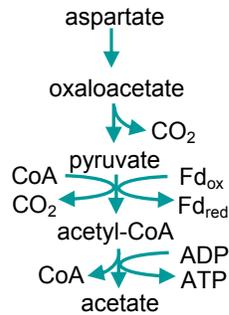
Grows on complex media, heterotroph

Domains unique among published Crenarchaeal genomes		
COG1883	Smar_1504	Na <sup>+</sup> -transporting decarboxylase, beta subunit
pfam04277	Smar_1503	Na <sup>+</sup> -transporting decarboxylase, gamma chain
COG1006	Smar_0026 Smar_0654 Smar_1069	Multisubunit Na <sup>+</sup> /H <sup>+</sup> antiporter, MnhC subunit
COG1320	Smar_0023 Smar_0651 Smar_1066	Multisubunit Na <sup>+</sup> /H <sup>+</sup> antiporter, MnhG subunit
COG1863	Smar_0027 Smar_0655 Smar_1070	Multisubunit Na <sup>+</sup> /H <sup>+</sup> antiporter, MnhE subunit
COG2111	Smar_0025 Smar_0653 Smar_1068	Multisubunit Na <sup>+</sup> /H <sup>+</sup> antiporter, MnhB subunit
COG0286	Smar_0761	Type I restriction-modification system methyltransferase
COG0610	Smar_0763	Type I restriction-modification system, R subunit
COG0732	Smar_0762	Restriction endonuclease S subunits
COG0424	Smar_0734	Protein implicated in inhibition of septum formation
pfam03169	Smar_1400	OPT oligopeptide transporter

Sodium-transporting decarboxylases – 2 possibilities  
both involved in amino acid degradation

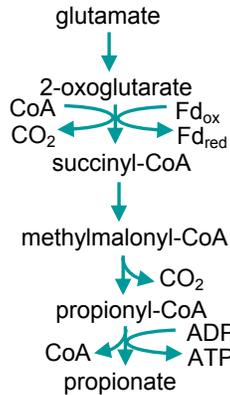
Oxaloacetate decarboxylase

Smar\_0341 - alpha  
Smar\_1504 - beta  
Smar\_1503 - gamma



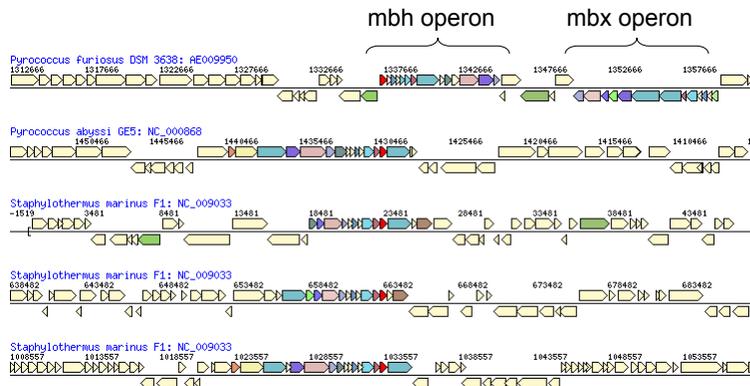
Methylmalonyl-CoA decarboxylase

Smar\_1426 - alpha  
Smar\_1504 - beta  
Smar\_1427 - gamma  
Smar\_1503 - delta

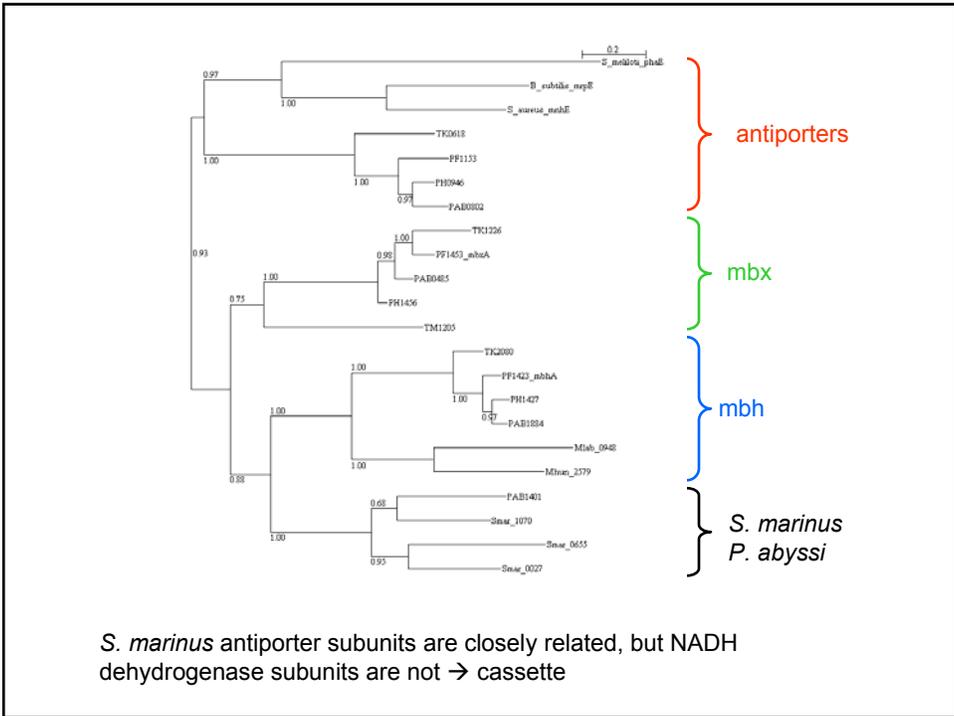


Multisubunit membrane protein complexes

- *S. marinus* has three operons with antiporter subunits and NADH dehydrogenase-related genes
- They are related to *P. furiosus* mbh and mbx operons
- Mbh = membrane bound hydrogenase
- Mbx = unknown function in electron transport/sulfur reduction



	<b>mbh</b>	<b>mbx</b>	<b>PabFHL</b>	<b>Smar1</b>	<b>Smar2</b>	<b>Smar3</b>
COG1863, MnhE	PF1423	PF1453	PAB1401	0027	0655	1070
COG2212, MnhF	PF1424	PF1452	PAB1398.1	0022	0650	1065
COG1320, MnhG	PF1425	PF1451	PAB1398	0023	0651	1066
COG1563	PF1426	PF1450	PAB1399.1	0024	0652	1067
COG2111, MnhB	PF1427 PF1428	PF1449	PAB1399	0025	0653	1068
COG1006, MnhC	PF1429	PF1448	PAB1400	0026	0654	1069
Pfam00361, MnhD/nuoLMN	PF1430	PF1447 PF1446	PAB1402 PAB1392 PAB1391	0028	0645	1057 1058 1071
mbhI-related	PF1431			0029		
Pfam01058, nuoB	PF1432	PF1444	PAB1396	0018	0646	1063
Pfam00329, nuoC	PF1433	PF1443	PAB1394	0019	0647	1061
Pfam00346, nuoD	PF1434	PF1442	PAB1394	0020	0648	1061
Pfam00146, nuoH	PF1435	PF1445	PAB1393	0030	0657	1060
COG1143, FHL6/nuoI	PF1436	PF1441	PAB1395	0021 (pseudo)	0649	1062



Do all sulfur-dependent Crenarchaeota use the same sulfur reduction pathways?

Enzyme	<i>S. marinus</i>	<i>T. pendens</i>	<i>H. butylicus</i>
Sulfur/polysulfide reductase (molybdoenzyme)	-	Tpen_1121-1123	Hbut_0371-0373
Sulphydrogenase	-	-	-
Sulfide dehydrogenase	-	-	-
NADPH:sulfur oxidoreductase	-	Tpen_0143	Hbut_0802
mbh/mbx-related	Smar_0018-0030, Smar_0645-0657, Smar_1057-1071	-	-

Same phenotype, different genotype

## Acknowledgments

### Genome Biology Program

Nikos Kyrpides  
Natalia Ivanova  
Thanos Lykidis  
Kostas Mavrommatis  
Sean Hooper  
Edwin Kim  
Natalia Mikhailova

### Virginia Bioinformatics Institute

Biswarup Mukhopadhyay  
Jason Rodriguez  
Dwi Susanti  
Lakshmi Dharmarajan

### University of Georgia

Barry Whitman  
Iris Porat

### University of Illinois

Gary Olsen  
Claudia Reich

### JGI

Alla Lapidus  
Hui Sun  
Susan Lucas  
Alex Copeland  
Matt Nolan

### LANL

Chris Detter  
Linda Thompson

### Oak Ridge

Miriam Land  
Igor Zhulin  
Luke Ulrich  
Jim Elkins

### University of Regensburg

Harald Huber