Department of Energy Final Technical Report

DOE Award number DE-FG02-01ER63133

Title: Gene Transfer and Hybridization Studies in Hyperthermophilic Species

Principal Investigator: Karen E. Nelson, Ph.D.

Institution: The Institute for Genomic Research

In order to investigate the extent of gene transfer across the Thermotogales, we used comparative genomic hybridization (CGH) strains of *Thermotoga* isolated from different locations throughout the world. The Thermotoga genome array was constructed in year one of this award. The Thermotoga strains that were used in the study, along with their place of isolation are below in Table 1.

Strain	Habitat	Temp (C)
MSB8	Geothermal heated seafloor, Vulcano Island, Italy	55-90
LA4	Shore of Lac Abbe, Djibouti	82
LA10	Shore of Lac Abbe, Djibouti	87
RQ2	Geothermal heated seafloor, Ribeira Quente, the Azores	76-82
RQ7	Geothermal heated seafloor, Ribeira Quente, the Azores	76-82
NE2x_L8B	Naples, Italy	-
NE7/L9B	Naples, Italy	-
S1/L12B	Naples, Italy	-
PB1platt	Oil field at the Prudhoe Bay, Alaska, USA	-
VMA1/L2B	Vulcano Island, Italy	-

Table 1. Strains of *Thermotoga* used in the CGH study.

Figure 1. Phylogenetic Representation of the strains that were used in the comparative genome hybridization study.

The outer unknown is strand. The inner to MSB8 displayed 1 seventh

1a MSB8

Figure 3. Region 1 of the *Thermotoga* strains that were verified by amplification and sequencing. Key by Lanes:1, lambda ladder; 2, 1KB ladder;3, SB8-R1; 4, RQ2-R1;5, PB1-R1;6, S1-R1; 7, NE2X-R1;8, NE7-R1; 9, 1 KB ladder; 10, lambda ladder.

Figure 4. The 10 regions that were amplified and sequenced from Strain RQ2. Flanks are identical, regions have been inserted or deleted. Key by lanes:1, lambda ladder;2, 1KB ladder;3, MSB8-R2;4, RQ2-R2;5, MSB8-R5;6, RQ2-R5; 7, MSB8-R7; 8, RQ2-R7;9, MSB8-R8;10, RQ2-R8; 11, MSB8-R9;12, RQ2-R9;13, MSB8-R10;14, RQ2-R10;15, MSB8-R11;16, RQ2-R11;MSB8-R12;RQ2-R12;MSB8-R15;NE7-R15;1 KB ladder;lambda ladder.

Three papers resulted from the above study (Haft, Selengut et al. 2005; Mongodin, Hance et al. 2005; Deboy, Mongodin et al. 2006). In the first entitled **Gene Transfer and Genome Plasticity in *Thermotoga maritima*, a Model Hyperthermophilic Species (Mongodin, Hance et al. 2005)**, the results highlighted the dynamic nature of the genome of members of this genus and support the idea that there has been extensive gene transfer in the *Thermotoga* lineage. The variation in genome composition was found to be independent of the closeness of strains to each other based on 16S rRNA phylogenetic analysis. The mechanisms for gene transfer in this lineage remain to be elucidated. The second manuscript (Haft, Selengut et al. 2005) focused on the repetitive nature of the CRISPR elements, their locations on the chromosome, and their variability among different strains. The CRISPRs were also compared to those found in the genomes of other microbial species. In the third and final manuscript (Deboy, Mongodin et al. 2006) the chromosomes of two members of the Thermotogales were compared. A whole-genome alignment of *T. maritima* MSB8 and *T. neapolitana* NS-E revealed large-scale DNA rearrangements, most of which are associated with CRISPR DNA repeats and/or tRNA genes. The results from this component of the study suggest that the observed

chromosomal rearrangements in the *Thermotogales* likely occurred by successive inversions after their divergence from a common ancestor and before strain diversification.

The results from this award highlight lateral gene transfer as a powerful and major evolutionary tool that has resulted in a significant role in species evolution (Mongodin, Hance et al. 2005). Also, based on the *Thermotoga* analysis, it is evident that substrate utilization and availability in the immediate environment may be one of the main reasons for loss and gain of genetic material (Mongodin, Hance et al. 2005).

Pyrococcus Comparative Genome Hybridization and Suppressive Subtractive Hybridization Studies.

In the second year of the award, a *Pyrococcus* array was created. Using custom designed primers we amplified the 2065 open reading frames (ORFs). The length of the *Pyrococcus furiosus* ORFs varied between a minimum of 65 base pairs to a maximum of 5223 base pairs with an average length of 906 base pairs. There were 1422 ORFs smaller than 1000 base pairs, 574 ORFs between 1000 base pairs and 2000 base pairs, 54 ORFs between 2000 base pairs and 3000 base pairs, and 15 ORFs larger than 3000 base pairs. The PCR products representing these ORFs varied in size between a minimum of 50 base pairs and a maximum of 1374 base pairs. In order to better represent on the array the 69 ORFs larger than 2000 base pairs, we designed additional primer pairs further apart for these. Co-PI DiRuggiero identified 173 ORFs of interest from the *Pyrococcus abyssi* and *Pyrococcus horikoshii* genomes (Lecompte et al, 2001), and we created PCR products from these as well. These ORFs are also included on the array.

The total of 2307 PCR products were checked by agarose gel electrophoresis before and after purification. The purification was done with Millipore MultiScreen Vacuum Manifold for all the products, except for the ones smaller in size than 100 base pairs, which were purified by ethanol precipitation. The PCR products were rearranged from 96 well plates into 384 well plates, mixed with DMSO, and printed on Corning UltraGAPS Coated Slides. The array was printed in duplicate mode. Quality control testing of the array involved staining with Syto 61 stain (Molecular Probes, Inc.) as well as labeling

individual PCR products with fluorescent dyes, hybridizing them to the array and confirming their correct location.

Co-PI DiRuggiero provided us with a 22 new isolates of hyperthermophilic Archaea that are supposed to be related to the different *Pyrococcus* species that have been sequenced. Some of the strains were isolated from different areas of the East and West Pacific, while 7 of them have been isolated from Vulcano, Italy by Nelson and DiRuggiero during the sampling expedition undertaken in 2002 (Table 3). We conducted approximately 25 preliminary hybridizations using controls and 5 of the East Pacific strains, and we are in the process of completing their analysis. We proceeded in our hybridizations with the same stringency used during our previous *Thermotoga* strains/species experiment. A first impression is that the Archeal strains are not as closely related to *P. furiosus* as we previously expected. As a result we are currently experimenting with hybridizing the *P. horikoshii* and *P. abyssi* genomic DNA to the *P. furiosus* array. The *P. horikoshii* and *P. abyssi* genomes have been previously sequenced (Kawarabayasi et al, 1998; and <http://www.genoscope.cns.fr/Pab/>) and we used BLAST to compare their full genomes to the genome of *P. furiosus*. We intend to use this data to support our efforts to lower the stringency of the hybridizations to a level appropriate to the degree of divergence present in the strains we are working with. Examples of parameters we wish to adjust in order to lower the stringency are the amount of formamide present in the hybridization buffer and the composition of the post-hybridization washing buffers (Bowtell and Sambrook, 2003).

Some of the strains were also subjected to Supressive Subtractive Hybridization (SSH) experiments. **Two subtractive libraries were sequenced.** From a total of 10 SSH 96-well plates (960 clones), we generated a total of 2112 sequences (1390 good). For the comparison with strain 12/1, 1152 seqs (726 good) were generated. For the comparison with JT6, we generated 960 seqs (664 good). The SSH libraries were created at the University of Maryland, and sent to TIGR in the form of frozen stocks. The templates were created and sequenced at TIGR and the data was sent back to Co-PI DiRuggiero for analysis.

Strain list - 6/17/03

Origins	Strains	16SrRNA	Microarray List
Vulcano	Pfu	Pyrococcus	Yes
Fidji	Pab	Pyrococcus	Yes
Mariana	Pho	Pyrococcus	Yes
13° Pacific	12/1**	+	Yes
	21/4*	+	Yes
	30/2		
	30/3	+	yes
	30/4	+	
	31/2	+	yes
	32/1*	+	
	32/2*		
	32/3*	+	
	32/4*	gb - Pyrococcus	yes
Pacific	JT1*	gb - Pyrococcus	yes
	JT3	+	yes
	JT6		yes
	JT10**	+	yes
Vulcano	VB2 (t.b.p)		
	VB3 (t.b.p)		
	VB8-I**	DNA	yes
	VB8-II		
	VB8-III		
	VB8-V		
	VB8-VI		
	VB9-I	DNA	yes
	VB9-III		
	VB11-I	DNA	yes
	VB11-II		
	VB11-III		
	VB11-V	DNA	yes
	VB11-VI		

Pacific: Juan de Fuca Ridge, East Pacific (North of 13°)

13° Pacific: East Pacific Rise - 13°N-104°W

Vulcano: Mediterranean Sea

Fidji: West Pacific, South of Mariana Trench

Mariana Trench: West Pacific, North of Fidji

* extrachromosomal elements

*** SSH

tbp: to be plated

gb: genbank

Publications and Presentations Resulting from this Study.

Nelson, K.E. Insights into microbial diversity from comparative genome analysis. Metagenomics **2003**. Darmstadt, Germany. June, 2003.

Mongodin E. Extremophiles Genomics. Workshop on Extremophile Research –Theory and Techniques, July 21, Baltimore MD.

Nelson KE, Mongodin E, Hance I, Hubert R, Stetter K, Gill SR. Comparative Genome Hybridization reveals extensive genomic diversity across the Thermotogales. 103rd General Meeting of the American Society for Microbiology, May 18-22 2003, Washington DC.

Nelson KE, Mongodin E, Hance I, Gill SR. Identification of regions of lateral gene transfer across the Thermotogales. 2003 Genomes to Life Program (GTL) workshop – Department of Energy, February 10-12 2003, Arlington VA.

Mongodin E, Hance I, Huber R, Stetter K, Gill SR, Nelson KE. Comparative Genome Hybridization reveals extensive genomic diversity across the Thermotogales. 3rd ASM and TIGR Conference on Microbial Genomes, January 29-February 1 2003, New Orleans LA

Deboy, R. T., E. F. Mongodin, et al. (2006). "Chromosome Evolution in the Thermotogales: Large-Scale Inversions and Strain Diversification of CRISPR Sequences." *J Bacteriol* **188**(7): 2364-74.

Haft, D. H., J. Selengut, et al. (2005). "A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes." *PLoS Comput Biol* **1**(6): e60.

Mongodin, E. F., I. R. Hance, et al. (2005). "Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species." *J. Bacteriol.* **187**((14)).

Karen E. Nelson, Ph.D.
Associate Investigator
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850