Genome Sequence and Analysis of *Thermobifida fusca YX*

A. Lykidis[1], N. Ivanova[1], I. Anderson1, K. Mavromatis[1], P. Richardson[1], S. Lucas[1], A. Lapidus[1], D. Wilson[2], N. Kyrpides[1]

[1]DOE-Joint Genome Institute, Walnut Creek, CA and [2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

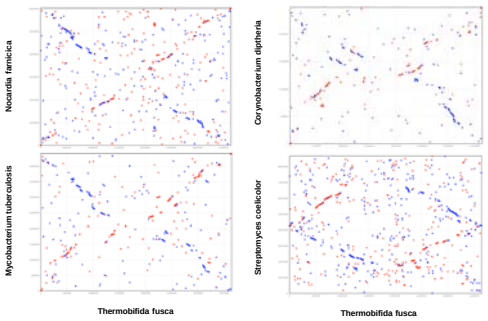
1. Abstract

Thermobifida fusca is a moderate thermophilic soil bacterium that belongs to actinobacteria. It is a major degrader of plant cell walls. Complete genome sequence showed that the *T. fusca* genome consists of a single circular chromosome of 3642249 bp with 85% of the genome sequence predicted to encode 3117 proteins and 65 RNA species. Genome analysis revealed the existence of 29 putative glycosyl hydrolases in addition to the 6 already identified cellulases. The glucosyl hydrolases include enzymes predicted to exhibit mainly dextran/starch and xylan degrading functions. At least two of the secreted cellulases have sequence signatures that indicate their secretion is mediated by the twin-arginine translocation system. Beside the twin-arginine translocation sysem, *T. fusca* possesses components of the *sec* system that allows proteins to cross the inner membrane and appears to lack other major secretion systems. *T. fusca* has the core phosphotransferase system components as other Actinobacteria but lacks PTS transporters. Genome comparisons reveal that 146 genes were laterally transferred to *T. fusca*. *T. fusca* has a small number (39) of unique genes.

2. Characteristics of the *Thermobifida fusca* genome

Characteristic	Value	
Genomic features		
Number of chromosomes	1	
Total no. of sequence bases	3642249	
DNA coding sequence	3122487 (85.73%)	
%GC content	2458622 (67.50%)	
Plasmids	None in this strain	
General statistics		
Total no. of genes	3181	(100.00%)
Protein coding genes	3117	(97.99%)
RNA genes	64	(2.01%)
rRNA genes	12	(0.38%)
5S rRNA	4	(0.13%)
16S rRNA	4	(0.13%)
23S rRNA	4	(0.13%)
tRNA genes	52	(1.63%)
Other RNA genes	0	
Genes with function prediction	2181	(68.54%)
Genes without function prediction	936	(29.41%)
with similarity	898	(28.2%)
without similarity	38	(1.19%)
Pseudogenes	7	(0.22%)
Genes assigned to enzymes	336	(10.56%)
Genes in COGs	2378	(74.73%)
Genes in Pfam	2132	(67.00%)
Genes in ortholog clusters	2824	(88.75%)
Genes in paralog clusters	339	(10.65%)
Genes connected to KEGG pathways	265	(8.33%)

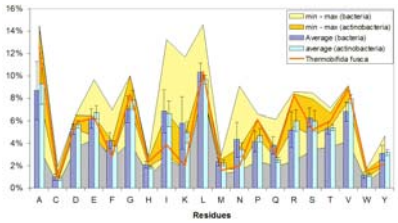
3. Comparison of chromosomal structure of T. fusca versus Nocardia farnicica, Mycobacterium tuberculosis, Corynobacterium glutamicum and Streptomyces coelicolor



4. Comparison of gene sets between selected Actinobacteria genomes

5. Amino acid utilization in thermophilic bacteria

Thermophilic proteomes exhibit an increased content of charged (Asp, Glu, Lys, Arg) versus polar (Asn, Gln, Ser, Thr) residues. This trend is a common feature for organisms with optimal growth temperature higher than 55 °C. *T. fusca* exhibits the same trend with the striking exception of Lys residues, which appear to be close to the minimum for bacteria in IMG. The reverse trend is observed for Ala which is elevated in *T.fusca* while thermophiles have fewer Ala residues

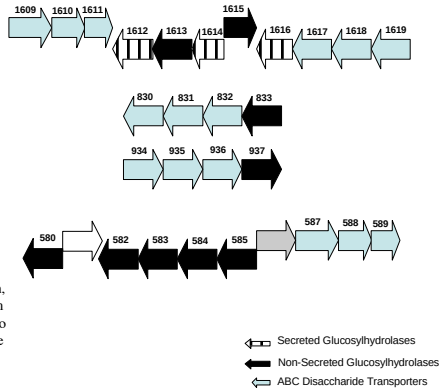


6. T. Fusca utilizes the twin-arginine translocation (TAT) system to secrete cellulases and glucosylhydrolases

			Secreted	SP	TAT
Tfu2923, Xylanase	GH_10	CBM2	Yes	+	+
Tfu2923, Xylanase	GH_10		Yes	+	-
Tfu1629, beta-glucosidase	GH_1		Yes	+	-
Tfu0937, beta-glucosidase	GH_1		No	-	
Tfu2712, Cellulase	Cellulase	CBM3	Yes	+	+
Tfu1959, beta 1,4 exocellulase	CBM2	GH_48	Yes	+	+
Tfu0900, beta-mannanase	Cellulase	CBM2	Yes	+	+
Tfu0901, endocellulase	CBM2	Cellulase	Yes	+	+
Tfu1213, xylanase	GH_11	CBM2	Yes	+	+
Tfu1074, beta 1,4 endoglucanase	GH_6	CBM2	Yes	+	-
Tfu0620, beta 1,4 exocellulase	CBM2	GH_6	Yes	+	+

7. Chromosomal clustering of ABC disaccharide transporters with secreted and non-secreted glucosylhydrolases.

Each arrow represents one gene and locus numbers are indicated above the arrows



This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.