

Abstract

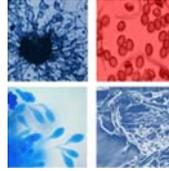
Rapid, efficient, and robust enzymatic degradation of biomass-derived polymers is currently a major obstacle in biofuel production. A key missing component in that process is the availability of enzymes that hydrolyze cellulose, hemicellulose, and other polysaccharides into biofuel substrates at temperatures and chemical conditions suitable for industrial use. Thermophilic fungi are known to excrete enzymes that rapidly degrade polymers including cellulose and hemicellulose at high temperatures, however the genome sequence and full gene complement of thermophilic fungi have not been previously reported. Here we describe the initial sequencing, gene identification, and comparative analysis of two thermophilic fungi, *Thielavia terrestris* and *Sporotrichum thermophile*, with comparison to *Chaetomium globosum*, a closely related non-thermophilic fungal species.



Sporotrichum thermophile
photo Courtesy of Tricia John



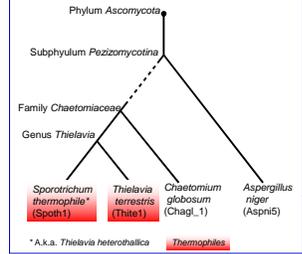
Thielavia terrestris
photo courtesy of Randy Berka



Chaetomium globosum

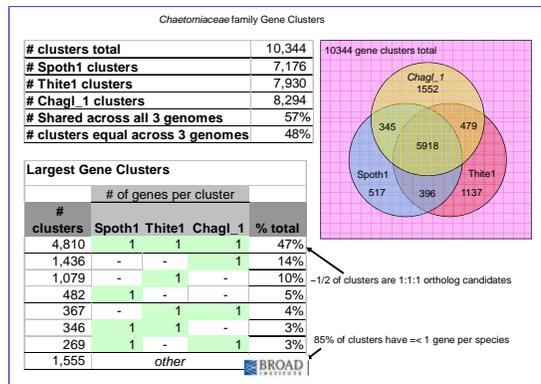


Organisms' Phylogenetic Relationships

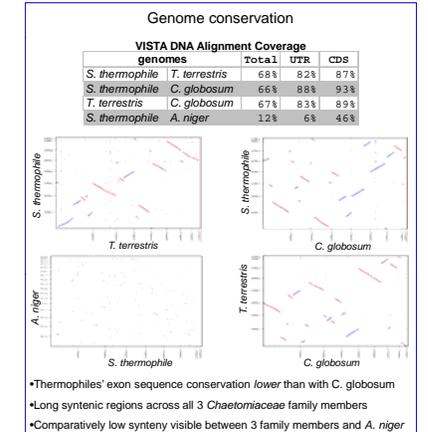
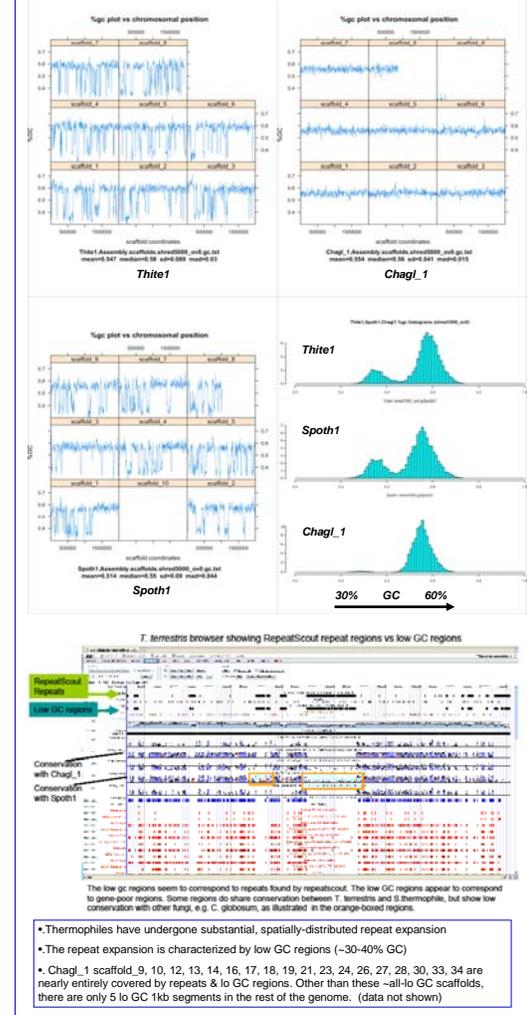


Assembly Comparison	<i>Sporotrichum thermophile</i> Spoth1	<i>Thielavia terrestris</i> Thite1	<i>Chaetomium globosum</i> Chagl_1
Genome size(Mbp)	38.7	36.9	34.9
Sequencing read coverage depth	7.66x	10.15x	7x
Reported # of contigs	293	231	2453
# of scaffolds	10	8	37
# of scaffolds >2 Kbp	10	8	32
Scaffold N50	3/5.4 Mbp	3/4.6 MB	4/4.7 Mbp
Three largest Scaffolds (Mbp)	7.0, 7.0, 5.4	9.5, 6.0, 4.6	6.6, 5.0, 4.7
Sequencing Method	Sanger WGS	Sanger WGS	Sanger WGS
% genome in gaps	0.34%	0.41%	1.58%
% genome in repeats	31%	22%	9%
# genes	8,806	9,815	11,124
Gene density (genes/Mbp)	228	266	319

Note: Organelle DNA was excluded from Spoth1 and Thite1 analysis
Note: Chagl_1 includes 695 gene models that would have been eliminated by JGI length & repeat match filter. Note: Chagl_1 includes 695 gene models that would have been eliminated by JGI length & repeat match filtering.



Thermophile-specific GC & Repeat Distribution



Thermophile-Specific Functional Domains (PFAM)

Thermophile-only domains:

# domains total	Spoth1	Thite1	Chagl_1	Aspn15	descriptions
3	1	2	0	0	Cupin superfamily (DUF985)
3	1	2	0	0	O-Glycosyl hydrolase family 30
2	1	1	0	0	Alanine racemase, N-terminal domain
2	1	1	0	0	Chromo shadow domain
2	1	1	0	0	Conserved hypothetical protein CHP02453
2	1	1	0	0	DASH complex subunit Dac4
2	1	1	0	0	Double-stranded DNA-binding domain
2	1	1	0	0	Dpy-30 motif
2	1	1	0	0	Eukaryotic protein of unknown function (DUF866)
2	1	1	0	0	Glycosyl hydrolases family 25
2	1	1	0	0	GRIM-19 protein
2	1	1	0	0	HILUase/Transferrin family
2	1	1	0	0	Molybdenum dinucleotide binding domain
2	1	1	0	0	NHL repeat
***	(80 total)				

