Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*

Randy M Berka[1,15], Igor V Grigoriev[2,15], Robert Otillar[2], Asaf Salamov[2], Jane Grimwood[3], Ian Reid[4], Nadeeza Ishmael[4], Tricia John[4], Corinne Darmond[4], Marie-Claude Moisan[4], Bernard Henrissat[5], Pedro M Coutinho[5], Vincent Lombard[5], Donald O Natvig[6], Erika Lindquist[2], Jeremy Schmutz[3], Susan Lucas[2], Paul Harris[1], Justin Powlowski[4], Annie Bellemare[4], David Taylor[4], Gregory Butler[4], Ronald P de Vries[7,8], Iris E Allijn[7], Joost van den Brink[7], Sophia Ushinsky[4], Reginald Storms[4], Amy J Powell[9], Ian T Paulsen[10], Liam D H Elbourne[10], Scott E Baker[11], Jon Magnuson[11], Sylvie LaBoissiere[12], A John Clutterbuck[13], Diego Martinez[6, 14], Mark Wogulis[1], Alfredo Lopez de Leon[1], Michael W Rey[1] & Adrian Tsang[4,15]

[1]Novozymes, Inc., Davis, California, USA. [2]US Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [3]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. [4]Centre for Structural and Functional Genomics, Concordia University, Montreal, Quebec, Canada. [5]Architecture et Fonction des Macromolécules Biologiques, CNRS/Universités de Provence/Université de la Mediterranée, Marseille, France. [6]Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA. [7]CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands. [8]Microbiology and Kluyver Centre for Genomics of Industrial Fermentation, Utrecht University, Utrecht, The Netherlands. [9]Sandia National Laboratory, Albuquerque, New Mexico, USA. [10]Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia. [11]Fungal Biotechnology Team, Pacific Northwest National Laboratory, Richland, Washington, USA. [12]McGill University and Génome Québec Innovation Centre, Montreal, Canada. [13]University of Glasgow, Glasgow, UK. [14]Present address: Broad Institute of MIT & Harvard, Cambridge, Massachusetts USA. [15]These authors contributed equally to this work.

October 2011

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*

Randy M Berka[1,15], Igor V Grigoriev[2,15], Robert Otillar[2], Asaf Salamov[2], Jane Grimwood[3], Ian Reid[4], Nadeeza Ishmael[4], Tricia John[4], Corinne Darmond[4], Marie-Claude Moisan[4], Bernard Henrissat[5], Pedro M Coutinho[5], Vincent Lombard[5], Donald O Natvig[6], Erika Lindquist[2], Jeremy Schmutz[3], Susan Lucas[2], Paul Harris[1], Justin Powlowski[4], Annie Bellemare[4], David Taylor[4], Gregory Butler[4], Ronald P de Vries[7,8], Iris E Allijn[7], Joost van den Brink[7], Sophia Ushinsky[4], Reginald Storms[4], Amy J Powell[9], Ian T Paulsen[10], Liam D H Elbourne[10], Scott E Baker[11], Jon Magnuson[11], Sylvie LaBoissiere[12], A John Clutterbuck[13], Diego Martinez[6, 14], Mark Wogulis[1], Alfredo Lopez de Leon[1], Michael W Rey[1] & Adrian Tsang[4,15]

[1]Novozymes, Inc., Davis, California, USA. [2]US Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [3]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. [4]Centre for Structural and Functional Genomics, Concordia University, Montreal, Quebec, Canada. [5]Architecture et Fonction des Macromolécules Biologiques, CNRS/Universités de Provence/Université de la Mediterranée, Marseille, France. [6]Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA. [7]CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands. [8]Microbiology and Kluyver Centre for Genomics of Industrial Fermentation, Utrecht University, Utrecht, The Netherlands. [9]Sandia National Laboratory, Albuquerque, New Mexico, USA. [10]Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia. [11]Fungal Biotechnology Team, Pacific Northwest National Laboratory, Richland, Washington, USA. [12]McGill University and Génome Québec Innovation Centre, Montreal, Canada. [13]University of Glasgow, Glasgow, UK. [14]Present address: Broad Institute of MIT & Harvard, Cambridge, Massachusetts USA. [15]These authors contributed equally to this work. Correspondence should be addressed to A.T. (tsang@gene.concordia.ca).

Received 16 May; accepted 18 August; published online 02 October 2011; doi:10.1038/nbt1976

Thermostable enzymes and thermophilic cell factories may afford economic advantages in the production of many chemicals and biomass-based. Here we describe and compare the genomes of two thermophilic fungi, *Myceliophthora thermophila* and *Thielavia terrestris*. To

our knowledge, these genomes are the first described for thermophilic eukaryotes and the first complete telomere-to-telomere genomes for filamentous fungi. Genome analyses and experimental data suggest that both thermophiles are capable of hydrolyzing all major polysaccharides found in biomass. Examination of transcriptome data and secreted proteins suggests that the two fungi use shared approaches in the hydrolysis of cellulose and xylan but distinct mechanisms in pectin degradation s. Characterization of the biomass-hydrolyzing activity of recombinant enzymes suggests that these organisms are highly efficient in biomass decomposition at both moderate and high temperatures. Furthermore, we present evidence suggesting that aside from representing a potential reservoir of thermostable enzymes, thermophilic fungi are amenable to manipulation using classical and molecular genetics.

Rapid, efficient and robust enzymatic degradation of biomass-derived polysaccharides is currently a major challenge for biofuel production. A prerequisite is the availability of enzymes that hydrolyze cellulose, hemicellulose and other polysaccharides into fermentable sugars at conditions suitable for industrial use. The best studied and most widely used cellulases and hemicellulases are produced by *Trichoderma*, *Aspergillus* and *Penicillium* species, and they are most effective over a temperature range from 40 °C to ~50 °C. At these temperatures, complete saccharification of biomass polysaccharides (>90% conversion to fermentable sugars) requires long reaction times, during which hydrolysis reactors are susceptible to contamination. One way to overcome these obstacles is to raise the reaction temperature, thereby increasing hydrolytic rates and reducing contamination risks. However, implementing higher reaction temperatures requires the deployment of enzymes that are more thermostable than the available preparations from mesophilic fungi. Additional advantages of elevated hydrolysis temperatures include enhanced mass transfer, reduced substrate viscosity, and the potential for enzyme recycling[1].

Thermophilic fungi represent a potential reservoir of thermostable enzymes for industrial applications. They can also potentially be developed into cell factories to support production of chemicals and materials at elevated temperatures. Enzymes from thermophilic fungi often tolerate higher temperatures than enzymes from mesophilic species, and some show stability at 70–80 °C[1,2]. Notably, it has been reported the cellulolytic activity of some thermophilic species was several times higher than that of the most active cellulolytic mesophiles[3]. Furthermore, biomass-degrading enzymes from thermophilic fungi consistently demonstrate higher hydrolytic

capacity[4] despite the fact that extracellular enzyme titers (in grams per liter) are typically lower than those from more conventionally used species such as *Trichoderma* or *Aspergillus*. We describe comparative genomic analyses of two thermophilic ascomycete species, *Thielavia terrestris* and *Myceliophthora thermophila*, to our knowledge the first filamentous fungi with finished genomes. These sequences open the way for new industrial applications of the enzymes from these organisms and potential development of thermophilic fungal production hosts.

RESULTS

Genomes summary

Among thermophilic fungi, *M. thermophila* and *T. terrestris* are two of the best characterized in terms of thermostable enzymes and cellulolytic activity[1–4]. The fermentation characteristics of these two organisms have been examined and found to be suitable for large-scale production[5,6]. The finished 38,744,216-bp genome of *M. thermophila* and 36,912,256 bp genome of *T. terrestris* contain, respectively, seven and six complete telomere-to-telomere chromosomes (**Fig. 1** and **Table 1**). Their telomeres comprise TTAGGG repeats commonly found in telomeres of filamentous fungi. The two genomes are similar in organization. The major difference occurs in chromosome (Ch)1 of *T. terrestris*, which harbors most of the genes located on Ch2 and Ch4 of *M. thermophila*. In addition, extensive translocation is observed between Ch1/Ch6 of *M. thermophila* and Ch2/Ch5 of *T. terrestris*. The protein coding fractions of the genomes include 9,110 genes in *M. thermophila* and 9,813 genes in *T. terrestris* (**Table 1**); both are smaller than average proteomes of other fungi in the class Sordariomycetes, and substantially smaller than the closely related mesophile *Chaetomium globosum*[7], which has 11,124 predicted genes in a 34.9-Mbp genome. These three species within the family Chaetomiaceae share 6,279 three-way orthologs and extensive synteny with >6,000 genes in syntenic blocks between each pair, including four blocks of >400 genes between *M. thermophila* and *T. terrestris* (**Supplementary Fig. 1** and **Supplementary Table 1**). The breakpoints of the synteny blocks often coincide with AT-rich repetitive regions (**Fig. 1**). The largest gene families in the genomes of *M. thermophila* and *T. terrestris* include transporters (e.g., MFS, ABC, AAA and sugar transporters) and proteins involved in signaling (e.g., protein kinases and WD40) as shown in **Supplementary Table 2**, often with more genes of each type in *T. terrestris*. Several Pfam domains appear to be expanded

in the Chaetomiaceae, including glycoside hydrolase families GH61 and GH11, and hypothetical proteins with a DUF1996 domain of unknown function (**Supplementary Table 3**).

Enzymes for biomass degradation

Proteins encoded in the genomes of *T. terrestris* and *M. thermophila* were compared to eight other fungi for genes encoding carbohydrate-active proteins[8] (CAZymes): glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases, glycosyl transferases (GT) and carbohydrate-binding modules (**Supplementary Table 4**). Like the other fungi examined, the two thermophiles harbor large numbers (>210) of glycoside hydrolases and polysaccharide lyases covering most of the recognized families, albeit with important differences (**Supplementary Figs. 2**, **3** and **Supplementary Tables 5**, **6**). For instance *T. terrestris* is poor in pectin and pectate lyases (no PL1, PL3, PL9 and PL11) and relatively rich in polygalacturonases (seven GH28). In contrast, the reverse is true for *M. thermophila* (five PL1, one PL3 and two GH28). Pectin lyases are most active at neutral to alkaline pH whereas GH28 pectin hydrolases are most active in acidic pH. Consistent with their repertoires of pectinolytic enzymes, *M. thermophila* grows best on pectin at neutral to alkaline pH whereas the growth of *T. terrestris* on pectin is best at acidic pH (**Supplementary Fig. 4**). The two thermophiles can be considered all-purpose decomposers with respect to their CAZymes and their ability to degrade plant polysaccharides (**Supplementary Fig. 5**).

Compared to the paradigmatic cellulase producer, *Trichoderma reesei*, the two thermophiles have similar complements of GH proteins. A major difference is the clear expansion of the GH61 family, and to a lesser extent the GH10 and GH11 xylanases, in members of the family Chaetomiaceae examined in this study (at least 18 GH61 proteins for the Chaetomiaceae and three for *T. reesei*) (**Supplementary Tables 5** and **6**). The GH61 family was originally classified on the basis of very weak endo-1,4-β-D-glucanase activity found in one family member[9]. Recently, it was reported that certain GH61 proteins lack measurable hydrolytic activity by themselves, but in the presence of various divalent metal ions, they can substantially enhance lignocellulosic biomass hydrolysis by cellulases and reduce the amount of cellulase required for hydrolysis of biomass polysaccharides[10]. The expansion of GH61 genes in this group of fungi may have evolved as a modified strategy for deconstruction of biomass polysaccharides compared to that of other species such as *T. reesei*. is suggested by the

evolutionary relationship of GH61 proteins among selected species in the order Sordariales, whose members can be divided into 25 orthologous clades (designated A–Y) (**Supplementary Fig. 6**). The fact that these organisms have maintained a diverse array of GH61 genes throughout their evolution intimates their importance for degradation of plant cell wall polysaccharides, with differing GH61 types possibly acting on assorted substrates and/or possessing varied biochemical properties. Differential expression of discrete GH61 subtypes (noted below) supports this view.

Transcript profiles on biomass substrates

To examine the strategy used by these thermophiles for decomposition of plant cell wall polysaccharides, we used RNA-Seq to compare transcript profiles during growth on barley straw or alfalfa straw to growth on glucose. Alfalfa was chosen to represent dicotyledonous plants, whereas barley was used to represent monocotyledon plants. The major difference between these materials is that the carbohydrates from barley cell wall are mainly cellulose and hemicellulose with a negligible amount of pectin[11], whereas alfalfa cell wall contains pectin and xylan in roughly similar proportions, each consisting of 15–20% of total carbohydrates [12,].

We observed notable differences between the transcriptional profiles of genes encoding different classes of carbohydrate-active enzymes (**Fig. 2a** and **Supplementary Tables 5-7**). As expected, the genes encoding enzymes used for modification of fungal cell walls (e.g., GH16, GH17 and GH72 proteins) are expressed at similar levels during growth on glucose and plant straws for the two organisms. In contrast, transcripts encoding enzymes that deconstruct plant cell walls are upregulated only during growth on barley or alfalfa straws. For growth on barley straw, the induced transcripts correspond closely with the substrate composition; genes for cellulolytic and xylanolytic enzymes are highly upregulated, followed by genes for arabinanases, mannanases and to a lesser extent pectinolytic enzymes. However, such simple matching of gene activity and substrate constituents does not extend to growth on alfalfa straw, especially for *T. terrestris*. Though upregulated when compared to growth on glucose, transcripts encoding xylanolytic enzymes during growth on alfalfa straw are kept at a relatively low level in both organisms. Transcripts encoding pectin lyases are highly upregulated for *M. thermophila* but not for *T. terrestris*, accentuating alternative strategies used by these two organisms for degradation of pectin. *T. terrestris* degrades pectin using primarily hydrolase activity (GH28), whereas *M.*

thermophila employs predominantly pectin lyases. When grown on complex substrates, several genes encoding GH61 proteins are highly upregulated in the thermophiles, particularly on barley straw for *M. thermophila*.

Fungi that decompose plant biomass typically possess multiple genes for degradation of a given polysaccharide polymer, and the thermophiles are no exception. Orthologs in these two genomes display similar patterns of expression. Only a subset of genes encoding a given enzyme activity is upregulated. Moreover, the same subset of genes is upregulated in different growth conditions. For example, the orthologs in Clades A, B, E, G and P of GH61 are upregulated under growth in complex substrates for both thermophiles (**Fig. 2b**). An even more striking correlation between transcript levels and orthologs is evident for the GH6 and GH7 cellulases (**Supplementary Fig. 7**) where the transcript profiles for the orthologs of the two organisms are essentially identical. With the exception of the pectinolytic enzymes, the correlation between expression profiles and orthologs extends to many of the lignocellulolytic proteins (**Supplementary Table 7**).

Secretomes and exo-proteomes

In addition to extracellular CAZymes involved in digestion of polysaccharide nutrients, the genomes of *M. thermophila* and *T. terrestris* encode an assortment of hydrolytic and oxidative enzymes that may enhance their ability to forage noncarbohydrate substrates. Collectively, secreted proteins (the secretome) can also provide important information regarding cellular physiology and metabolism in both natural and industrial bioprocessing conditions. The secretomes of *M. thermophila* and *T. terrestris* are predicted to comprise 683 and 789 proteins, respectively (**Supplementary Tables 8** and **9**), of which 569 are homologs. The predicted extracellular proteins (the secretome) include about 180 CAZymes, 40 peptidases, >65 oxidoreductases and >230 proteins of unknown function in each species. Bioinformatic prediction tends to overestimate the number of secreted proteins because the features of some intracellular proteins, in particular those residing in the endoplasmic reticulum, are indistinguishable from secreted proteins. We therefore used mass spectrometry to identify secreted proteins involved in biomass degradation. In general, the genes encoding the identified proteins are expressed at levels higher than those of their paralogs that are not detected extracellularly, especially during growth on agricultural straws (**Supplementary Tables 8** and

9). Based on transcriptome analysis, the few peptidases detected in the exo-proteomes do not display differential regulation, implying that peptidases are not critical components in biomass degradation. Notably, the genes encoding some hypothetical proteins and oxidoreductases that are detected in the exo-proteomes, and which possess predicted signal peptides, are upregulated when these fungi are cultured on agricultural straws as compared to glucose; for example, Mycth_59005, Mycth_2298860, Mycth_2303335 and Thite_2106069. The role of these secreted proteins in lignocellulose degradation is currently being investigated.

Hydrolysis of polysaccharides in alfalfa straw

Thermophilic fungi are major components of the microflora in self-heating composts. They break down cellulose at a faster rate than prodigious, mesophilic cellulase producers such as *T. reesei* at 40–50 °C[3]. Plant biomass–degrading enzymes characterized from thermophilic fungi have temperature optima between 55 °C and 70 °C[13–15]. Consequently, enzymes from thermophiles are expected to break down plant biomass at a faster rate than enzymes from mesophiles at elevated temperatures. We examined the temperature effects on the hydrolysis of alfalfa straw using enzymes from *M. thermophila*, *T. terrestris*, *C. globosum* and *T. reesei* (**Fig. 3**). The optimum temperature of hydrolysis for enzymes from *T. reesei* occurs at 50 °C. The enzyme mixture from *C. globosum* displays a broad temperature optimum from 30–60 °C. Enzymes from the thermophiles release appreciably higher amounts of reducing sugars than do the enzymes from the mesophiles, with peaks at 40 °C and 60 °C. The biphasic temperature profile of cell wall decomposition suggests that the proteins secreted by the thermophiles contain multiple biomass enzymes, some of which have temperature optima at 60 °C, whereas others have optima around 40 °C.

To determine whether biomass-degrading enzymes from these thermophiles possess distinct temperature optima, we successfully cloned and expressed in *Aspergillus niger* the genes encoding seven xylanases from the thermophiles and tested their biochemical properties. The temperature optima for these xylanases range from 45 °C to 70 °C (**Table 2**). These results suggest that *M. thermophila* and *T. terrestris* not only possess diverse enzyme activities for degradation of plant cell wall polysaccharides, they have also evolved diverse properties for these enzymes that enable efficient hydrolysis over a range of temperatures. The diversity of

enzyme activities and properties may help to explain the ubiquity of these organisms in decomposing biomass.

Potential utility of sexual cycle for strain development [AU: OK as edited?]

The ability to do sexual crossing is rare among fungal cells used for bioproduction. Crossing can facilitate strain improvement stemming from recombinant DNA methods, classical mutagenesis, and genome shuffling or natural variation. In this context, we evaluated genes involved in mating and the potential for outcrossing, particularly in *M. thermophila*.

Nearly all thermophilic Chaetomiaceae are either homothallic (self-fertile), as is the case for *C. globosum* and *T. terrestris*, or have been observed only in the asexual state, as is the case for *M. thermophila*. We examined these genomes for homologs of the mating-type genes of *Neurospora crassa*. Each genome possesses a homolog of *N. crassa matA-1* (CHGT_06585, Mycth_2298236 and Thite_2111503), the primary gene responsible for mating-type determination in *mat A* strains. As in *N. crassa*, *matA-2* homologs in the Chaetomiaceae species (CHGT_06585, Mycth_2107654 and Thite_2127265) are immediately adjacent to *matA-1* genes, and presumptive *matA-3* homologs are adjacent to *matA-2* genes (CHGT_06584, 1 Mycth_2115740 and Thite_2111506). We could not identify homologs for *mat a* genes.

The *M. thermophila* genome allowed us to confirm the close relationship between this species and *Myceliophthora heterothallica* (*Thielavia heterothallica*) [16], reported to be heterothallic[17] and for which *mat A* and *mat a* strains have been identified at the molecular level Using molecular markers based on *M. thermophila* sequences, we confirmed heterothallism and marker segregation in isolates of *M. heterothallica* from diverse locations (to date: New Mexico, Indiana and Germany). Attempts to cross the sequenced strain of *M. thermophila* were not successful. The ability to cross *M. thermophila* would permit optimization of gene combinations with natural and engineered gene variants, as well as development of a complete model organism from this industrially important group. Evidence of repeat induced in the sequenced species (**Supplementary Notes**) could represent an obstacle to be overcome to exploit crosses between strains with multicopy transgenes.

Distinguishing thermophilic from mesophilic fungi

Compared to the closely related mesophile *C. globosum*, the genomes of *M. thermophila* and *T. terrestris* contain larger fractions of repetitive sequences that have low GC content, introducing

significant GC variation (**Supplementary Fig. 8**). When comparing the GC content of *M. thermophila* and *T. terrestris* with *C. globusum* and other species within the class Sordariomycetes, it appears that although the genomes of the thermophilic species have a slightly lower genome-level GC content than *C. globosum*, they have a higher GC content in coding regions, which is reflected in the third position of codons (**Supplementary Table 10**). Since G: C pairs are more thermally stable; this may suggest the potential adaptability of protein-coding genes to high temperatures. Approximately 75% of *M. thermophila* codons have a higher GC content at the third nucleotide position (GC3) compared to the corresponding *C. globosum* orthologs. The percentage is even higher when comparing *T. terrestris* with *C. globosum*; 92% of *T. terrestris* codons have a higher GC3. This is in contrast to thermophilic prokaryotes, where analysis of large numbers of sequenced thermophiles and hyperthermophiles did not reveal a correlation of higher GC (and GC3) with thermal adaptability[18]. It remains to be seen whether high GC3 content will hold true for other thermophilic eukaryotes.

Analysis of prokaryotic thermophiles also included multiple attempts to define amino acid 'signatures' of thermophilic adaptations. A seven amino-acid motif IVYWREL has been reported that positively correlates with elevated growth temperatures in 204 complete proteomes of archaea and bacteria[19]. Two groups observed that thermophilic proteins are enriched in glutamine, arginine and lysine amino acid residues and contain lesser amounts of alanine, aspartic acid, asparagine, glutamine, threonine and serine[20,21], and another group found three notable substitutions (lysine to arginine, serine to alanine and serine to threonine) by comparing thermophilic *Corynebacterium efficiens* and mesophilic *C. glutamicum*, and suggested the differences are probably important for thermostability of *C. efficiens* proteins[22]. We searched for these amino acid signatures in filamentous fungi and found that they do not distinguish fungal thermophiles from their mesophilic relatives (**Supplementary Tables 11-14**). On the basis of our comparative analyses of the genomes from two thermophilic fungi, we conclude that their nucleotide and protein features are different from those observed in thermophilic prokaryotes.

We also investigated the possibility that thermophilic fungi possess major differences in processes mediating thermophily including heat shock, oxidative stress, membrane biosynthesis, chromatin structure and modification, and fungal cell wall metabolism. We compared the proteins predicted to be involved in these processes in *C. globosum*, *M. thermophila* and *T. terrestris*, but were unable to find differences that can convincingly be interpreted as the

molecular bases that underpin fungal thermophily (**Supplementary Notes** and **Supplementary Tables 15-22** and **23–25**).

Phylogeny and the origins of thermophily among Ascomycota

Thermophilic fungi, defined as fungi that grow better above 45 °C than at 25 °C, have evolved independently in at least two lineages within the phylum Ascomycota, once each within the orders Sordariales and Eurotiales (**Supplementary Fig. 9**). Within the Sordiariales, thermophily is restricted to subgroups of the family Chaetomiaceae. Among fungi more broadly, thermophily also exists in the Zygomycota, but it appears to be rare or absent in the phyla Basidiomycota and Chytridiomycota. The evolutionary trajectory of thermophily is obscured by chaotic taxonomy[23]. Biosystematic efforts have lagged behind research on thermophiles in industry, resulting in a body of literature for these organisms that lacks accurate taxonomic treatments. Aside from the potential for creating disputes and confusion in the commercial realm, the state of taxonomic study hinders efforts to determine the number of sublineages in which thermophily has been gained and lost in the groups where it is common. The fungi that are the subject of this paper provide excellent examples. *M. thermophila* has been placed alternately in the anamorphic genera *Sporotrichum* and *Chrysosporium*, both of which are inappropriate for fungi in the family Chaetomiaceae. Recognition of this organism as a member of the Chaetomiaceae therefore removes the need to assume a separate origin for thermophily in a third order of filamentous ascomycetes. True thermophiles have been assigned to the genera *Chaetomium* and *Thielavia*, along with mesophilic and thermotolerant species. If current taxonomic conclusions are correct, this would imply multiple independent gains or losses of thermophily within the family Chaetomiaceae. Alternatively, it is possible that current classification does not reflect phylogenetic relationships.

DISCUSSION

Thermophilic fungi are ubiquitous organisms commonly found in decomposing organic matter. The biotechnological utility of these fungi has been recognized for many years. The finished genomes for the two thermophiles may serve not only as reference genomes for studies on genome evolution and structure, but might also support targeting and mapping of changes that could facilitate strain construction and improvement. Selection from natural population variability or from mutagenized strains remains a useful tool for strain development. With a

finished genome as a scaffold and modern sequencing technologies, resequencing these strains and identifying mutations becomes relatively simple but very helpful for identifying beneficial and deleterious genetic changes.

Thermostability alone does not explain why the enzymes from the thermophilic fungi display higher hydrolytic power than enzymes produced by mesophiles. Crude extracellular enzymes from the thermophiles exhibit higher hydrolytic capacity than their counterparts from mesophiles at temperatures ranging from 30 °C to 60 °C (**Fig. 3**). One explanation is that the enzymes from the thermophiles possess higher specific activity toward lignocellulosic biomass. It is also possible that the thermophiles secrete a broader spectrum of accessory proteins that accelerate biomass degradation. In addition to the GH61 proteins[10], the poorly characterized extracellular proteins that are upregulated when the fungi are cultured on straws (**Supplementary Tables 8** and **9**) may represent new lignocellulose-active proteins.

When expressed in a mesophilic heterologous fungal host, the enzymes from thermophilic fungi usually retain their thermostable character[24, 25]. This provides an excellent opportunity to replace specific enzyme components in mesophilic production organisms with better-performing counterparts from thermophiles[10]. Additionally, prospects for further enhancements in the industrial performance of these enzymes through protein engineering appear likely[26].

An intriguing alternative to replacement of individual enzyme components in mesophiles with thermostable orthologs may involve development of thermophilic fungal production hosts. In this regard, such hosts that provide better hyphal morphology in tank fermentations have been shown to result in reduced viscosity and improved productivity, and they can be engineered by using protoplast transformation to introduce recombinant DNA constructs[5, 6]. It is also noteworthy that the well-developed industrial production organism previously known as *Chrysosporium lucknowense* C1 was recently reclassified as an isolate of *M. thermophila*[27]. When combined with the prospect of a functional sexual cycle and a finished genome, both of which enable rational design of improved strains, these observations may provide an adequate toolbox for development of efficient thermophilic fungal host strains.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology

Accession numbers. Assembly and annotation data for *M. thermophila* and *T. terrestris* are available through JGI MycoCosm Genome Portal at http://jgi.doe.gov/fungi and at DDBJ/EMBL/GenBank under chromosome accessions CP003002-CP003008 and CP003009-CP003014, respectively. The transcriptome data are available under GEO accession number GSE27323.

Note: Supplementary information is available on the Nature Biotechnology *website.*

ACKNOWLEDGMENTS

The genome sequencing and analysis were conducted by the US Department of Energy Joint Genome Institute and supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231. The work on transcriptomes, enzyme characterization and the *Myceliophthora* exo-proteome was supported by the Cellulosic Biofuel Network of the Agriculture Bioproducts Innovation Program of Agriculture and Agri-Food Canada, Genome Canada and Genome Québec.

AUTHOR CONTRIBUTIONS

The final text of the manuscript was written by R.M.B. and A.T., and reviewed by I.V.G.; who together also coordinated the overall analysis. I.V.G. coordinated both genome projects at the Joint Genome Institute. R.M.B. prepared the genomic DNA of *T. terrestris* and T.J. the DNA of *M. thermophila*. A.T. coordinated the transcriptome and exo-proteome work, and analyzed the transcriptomes. S.L. and E.L. led genome and cDNA sequencing. J.G. and J.S. finished and assembled both genomes. R.O. and A.S. annotated and analyzed the genomes, synteny, and GC content. I.R. processed the RNA-Seq data and analyzed the cell wall proteins. N.I. coordinated the sample preparation for transcriptome analysis and analyzed the lignocellulolytic proteins. B.H., P.M.C. and V.L. performed the comparative analysis of the carbohydrate-active proteins. C.D. conducted the enzymatic hydrolysis of straws and M-C. M. prepared the samples for transcriptome and exo-proteome analysis. D.O.N. analyzed the mating types and phylogeny of thermophilic fungi. E.L. coordinated the cDNA synthesis and EST analysis. A.B. coordinated the cloning and expression of xylanase genes. D.T. characterized the biochemical properties of the xylanases. R.P. de V., I.E.A, and J. van den B. examined the growth on different substrates. P.H. analyzed the GH61 proteins and J.P. membrane biogenesis. G.B. analyzed the secretomes. S.U.

and R.S. analyzed the chromatin structure and dynamics. A.J.P. examined melanin pigment biogenesis. I.T.P and L.D.H.E. analyzed transporters. S.E.B analyzed secondary metabolism. J.M. examined oxidative stress. M.W. reviewed proteases and peptidases. S.L. examined the exo-proteomes. A.J.C. looked for repeat-induced polymorphisms. D.M. contributed computational tools for viewing *T. terrestris* transcriptome data. A.L.de L. and M.W.R. examined oxidoreductases and chitinases, respectively.

COMPETING FINANCIAL

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at. http://www.nature.com/naturebiotechnology

Published online at http://www.nature.com/nbt/index.html.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

This paper is distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike license, and is freely available to all readers at http://www.nature.com/nbt/index.html.

1. Margaritis, A. & Merchant, R.F.J. Thermostable cellulases from thermophilic microorganisms. *Crit. Rev. Biotechnol.* **4**, 327–367 (1986).

2. Margaritis, A. & Merchant, R. Production and thermal stability characteristics of cellulase and xylanase enzymes from *Thielavia terrestris*. *Biotechnol. Bioeng. Symp.* **13**, 426–428 (1983).

3. Tansey, M.R. Agar-diffusion assay of cellulolytic ability of thermophilic fungi. *Arch. Mikrobiol.* **77**, 1–11 (1971).

4. Wojtczak, G., Breuil, C., Yamada, J. & Saddler, J.N. A comparison of the thermostability of cellulases from various thermophilic fungi. *Appl. Microbiol. Biotechnol.* **17**, 82–87 (1987).

5. Jensen, E.B. & Boominathan, K.C. Thermophilic fungal expression system. US Patent 5,695,985 (1997).

6. Jensen, E.B. & Karuppan, C.B. Thermophilic fungal expression system. US Patent 5,602,004 (1997).

7. Broad_Institute 2005Chaetomium globosum genome database, http://www.broadinstitute.org/annotation/genome/chaetomium_globosum

8. Henrissat, B. & Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**, 637–644 (1997).

9. Karlsson, J. *et al.* Homologous expression and characterization of Cel61A (EG IV) of Trichoderma reesei. *Eur. J. Biochem.* **268**, 6498–6507 (2001).

10. Harris, P.V. *et al.* Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* **49**, 3305–3316 (2010).

11. Pahkala, K. *et al.* Production of bioethanol from barley straw and reed canary grass: a raw material study. *15th European Biomass Conference and Exhibition.* Berlin, Germany (2007).

12. Dien, B.S. *et al.* Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. *Biomass Bioenergy* **30**, 880–891 (2006).

13. Kaur, G., Kumar, S. & Satyanarayana, T. Production, characterization and application of a thermostable polygalacturonase of a thermophilic mould Sporotrichum thermophile Apinis. *Bioresour. Technol.* **94**, 239–243 (2004).

14. Vafiadi, C., Topakas, E., Biely, P. & Christakopoulos, P. Purification, characterization and mass spectrometric sequencing of a thermophilic glucuronoyl esterase from Sporotrichum thermophile. *FEMS Microbiol. Lett.* **296**, 178–184 (2009).

15. Roy, S.K., Dey, S.K., Raha, S.K. & Chakrabarty, S.L. Purification and properties of an extracellular endoglucanase from Myceliophthora thermophila D-14 (ATCC 48104). *J. Gen. Microbiol.* **136**, 1967–1971 (1990).

16. van den Brink, J., Samson, R.A., Hagen, F., Boekhout, T. & de Vries, R.P. Phylogeny of the industrial relevant, thermophilic genera *Myceliophthora* and *Corynascus*. *Fungal Divers.* published online, doi:10.1007/s13225–13011–10107-z (28 May 2011

17. von Klopotek, A. Thielavia heterothallica spec. nov., die perfekte Form von Chrysosporium thermophilum. *Arch. Microbiol.* **107**, 223–224 (1976).

18. Galtier, N. & Lobry, J.R. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J. Mol. Evol.* **44**, 632–636 (1997).

19. Zeldovich, K.B., Berezovsky, I.N. & Shakhnovich, E.I. Protein and DNA sequence determinants of thermophilic adaptation. *PLOS Comput. Biol.* **X**, e5 (2005).

20. Glyakina, A.V., Garbuzynskiy, S.O., Lobanov, M.Y. & Galzitskaya, O.V. Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms. *Bioinformatics* **23**, 2231–2238 (2007).

21. Wang, G.-Z. & Lercher, M.J. Amino acid composition in endothermic vertebrates is biased in the same direction as in thermophilic prokaryotes. *BMC Evol. Biol.* **10**, 263 (2010).

22. Nishio, Y. *et al.* Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of Corynebacterium efficiens. *Genome Res.* **13**, 1572–1579 (2003).

23. Mouchacca, J. Heat-tolerant fungi and applied research work: a synopsis of name changes and synomomies. *World J. Microbiol. Biotechnol.* **16**, 881–888 (2000).

24. Berka, R.M., Rey, M.W., Brown, K.M., Byun, T. & Klotz, A.V. Molecular characterization and expression of a phytase gene from the thermophilic fungus *Thermomyces lanuginosus*. *Appl. Environ. Microbiol.* **64**, 4423–4427 (1998).

25. Murray, P. *et al.* Expression in and characterisation of a thermostable family 3 β-glucosidase from the moderately thermophilic fungus *Talaromyces emersonii*. *Protein Expr. Purif.* **38**, 248–257 (2004).

26. Voutilainen, S.P., Murray, P.G., Tuohy, M.G. & Koivula, A. Expression of *Talaromyces emersonii* cellobiohydrolase Cel7A in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability and activity. *Protein Eng. Des. Sel.* **23**, 69–79 (2010).

27. Visser, H. *et al.* Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense* C1. *Ind. Biotechnol.* **7**, 214–223 (2011).

Figure 1 Genome organization of *M. thermophila* and *T. terrestris*. The six chromosomes of *T. terrestris* are mapped to genomic regions from *M. thermophila* (shown as colored blocks in far-right lane). Only major synteny blocks are represented. For each *T. terrestris* chromosome, left-most lane shows G+C content, second lane from left shows repetitive elements and third lane from left shows regions with high gene density. **Figure 2** Analysis of transcription profiles (**a**) Expression of CAZymes genes of the thermophiles cultured on glucose, alfalfa straw and barley straw. Gene activity is presented as percentage of total CAZymes gene activity of the three culture conditions of each organism. (**b**) Transcript profiles of GH61 orthologs in *M. thermophila* and *T. terrestris*. The homologs of GH61 of *M. thermophila* and *T. terrestris* are organized in clades as shown in **Supplementary Figure 6**. Gene activity is presented as percentage of total CAZymes gene activity of the three culture conditions of each organism. Genes of several clades (e.g., K, O, R, S and T) are not upregulated in any of the growth conditions. None of the genes encoding GH61 proteins are upregulated during growth on glucose (**Supplementary Tables 5-7**).

Figure 3 Release of reducing sugars from alfalfa straw by crude extracellular enzymes from thermophilic and nonthermophilic fungi. The crude extracellular enzymes from the fungi cultured on alfalfa straw were used for hydrolysis. The hydrolysis reactions were performed at the temperatures indicated. The final protein concentrations in the reaction mixtures were: *Chaetomium globosum*, 439 µg/ml; *Myceliophthora thermophila*, 422 µg/ml; *Thielavia terrestris*, 362 µg/ml; and *Trichoderma reesei*, 524 µg/ml.

Table 1 Assembly and gene model statistics

	T. terrestris	***M. thermophila***
G+C content,%	54.7	51.4
Genome size	36.9 Mb	38.7 Mb
No. of chromosomes	6	7
No. of genes	9,813	9,110
Gene length (nt)[a]	1,649	1,733
Exons per gene[a]	2	2

[a]Median values over all genes in an organism.

Table 2 Temperature optima for *M. thermophila* (Mycth) and *T. terrestris* (Thite) xylanases

Gene ID	Family	Optima (°C)
Mycth_100068	GH11	50
Mycth_2121801	GH11	60
Mycth_112050	GH10	60
Mycth_116553	GH10	70
Thite_2117649	GH10	60
Thite_2042100	GH11	50
Thite_2107799	GH11	45

ONLINE METHODS

Additional details on the methods are provided in the **Supplementary Methods**.

Genome sequencing and assembly. Genomic DNA extracted from *Myceliophthora thermophila* (*Sporotrichum thermophile*) strain ATCC 42464 and *Thielavia terrestris* strain NRRL 8126 were used for whole genome shotgun sequencing. All sequencing reads were collected with standard Sanger sequencing protocols on ABI 3730XL capillary sequencing machines and assembled with ARACHNE[28]. Low-quality regions and gaps were computationally selected and sequenced with custom primers. After the completion of the automated rounds, we selected further reactions manually to finish the genomes. We resolved smaller repeats by transposon hopping with 8 kb plasmid clones. We shotgun sequenced and finished fosmid clones to fill large gaps, resolve larger repeats or to resolve chromosome duplications and extend into telomere regions. The finished genomes of *M. thermophila* and *T. terrestris* consist of seven and six chromosomes, respectively, comprising 38,744,216 bp and 36,912,256 bp of finished sequence with an estimated error rate of less than one error in 100,000 base pairs (**Table 1**).

Construction of cDNA libraries and analysis of expressed sequence tags (ESTs). Poly A^+ RNA was isolated from total RNA (pooled RNA from cells grown in rich medium for *T. terrestris* and 1% cellulose and 1% pectin pooled culture from *M. thermophila*) using the Absolutely mRNA Purification Kit and manufacturer's instructions (Stratagene). Synthesis and cloning of cDNA was a modified procedure based on the SuperScript plasmid system with Gateway technology for cDNA synthesis and cloning (Invitrogen). Plasmid DNA for sequencing was produced by rolling circle amplification[29] (TempliPhi, GE Healthcare). Subcloned inserts were sequenced from both ends. A total of 33,559 *T. terrestris* ESTs including 5,548 from external sources and 44,939 *M. thermophila* ESTs including 11,392 from external sources were processed through the JGI EST pipeline separately for each genome.

Genome annotation. Chromosome sequences were masked using REPEATMASKER[30] and the REPBASE library of 234 fungal transposable elements[31]. Gene modeling on the masked assembly was performed *ab initio* FGENESH[32] and GENEMARK-ES[33]; homology-based FGENESH+[32] and GENEWISE[34] seeded by BLASTx alignments of NCBI's nr (nonredundant) protein database against the assembly; cDNA-based EST_map (http://www.softberry.com/) seeded by EST contigs. EST BLAT alignments[35] were used to add or extend exons for gene models. Because multiple gene models were generated for each locus, a single representative model was algorithmically chosen based on model quality. Genes for tRNAs were predicted using tRNAscan-s.e.m[36]. The term gene model refers to protein-coding genes unless otherwise noted. The *C. globosum* genome assembly and gene models, used for comparison to the two thermophiles, were downloaded from the Broad Institute *Chaetomium globosum* Database at http://www.broadinstitute.org/annotation/genome/chaetomium_globosum. All predicted gene models were functionally annotated using SIGNALP[37], TMHMM[38], INTERPROSCAN[39], BLASTp[40] against the nr database, and hardware-accelerated double-affine Smith-Waterman alignments (deCypherSW; http://www.timelogic.com/decypher_sw.html) against SWISSPROT, the Kyoto Encyclopedia of Genes and Genomes (KEGG)[41], and the eukaryotic orthologous groups of proteins database (KOG)[42]. The Enzyme Commission (EC) numbers (http://www.expasy.org/enzyme/) of KEGG hits were assigned to gene models and mapped to corresponding KEGG pathways. INTERPRO and SWISSPROT hits were used to assign Genome Ontology (GO) terms[43] to gene models. Multigene families were predicted with the Markov clustering algorithm (MCL)[44], using BLASTp alignment scores between proteins as a similarity metric. Manual curation of automated annotations was performed using the JGI MycoCosm Genome Portal (http://jgi.doe.gov/fungi).

Transcriptome analysis. The thermophiles were cultured at 45 °C with shaking at 150 r.p.m. in 10× TDM[45] containing 2% glucose or 2% agricultural straws (alfalfa or barley straws ground to 0.5 mm lengths). Mycelia were harvested at early growth phase; 21 h for *M. thermophila* and 24–28 h for *T. terrestris*. Total RNA samples were isolated from mycelia as described[46]. Sequencing was performed using the RNA-Seq method of Illumina's Solexa IG at either the DNA Core Facility of the University of Missouri or at the McGill University-Génome Québec Innovation Centre. The RNA-Seq reads, 38–42 nt in length, from each mRNA sample were mapped against a combination of the genomic sequence and the spliced sequences with Bowtie[47]

using the 'best' strata option. The depth of mapped read coverage at each genome position was calculated using the WIGGLES program bundled with TopHat[48]. To analyze differential expression, we took the transcript and exon definitions from the filtered models of version 2.0 of genome annotation. The number of reads mapping to each transcript was estimated by integrating the coverage depth over the annotated exons of the transcript, dividing by the read length, and rounding to an integer. The Bayesian posterior probability of differential expression was estimated from the read counts for each transcript using the R package baySeq v 1.4 (ref. 49). To aid interpretation, we also calculated FPKM (fragments per kilobase of transcript per million mapped reads) values from the counts using the transcript lengths and the total number of mapped reads from each sample.

Secretome and exo-proteome analysis. Sequence-based prediction of extracellular proteins (secretome) was processed as described[50]. For exo-proteome analysis, proteins secreted by *M. thermophila* after 30 h of growth in barley and alfalfa straws were resolved on one-dimensional SDS-PAGE, fractionated into 12 bands, and in-gel digested with trypsin. Proteins secreted by *T. terrestris* after 96 h of growth in cellulose and xylose were resolved by two-dimensional gel electrophoresis. After staining with Coomassie Blue, over 100 spots from each gel were excised and in-gel digested with trypsin. Peptides eluted from the gel fragments were analyzed by tandem mass spectrometry.

Hydrolysis of polysaccharide biomass. To obtain extracellular proteins, we grew the fungi in 10× TDM containing 2% alfalfa straw (0.5 mm length), at 45 °C for *M. thermophila* and *T. terrestris*, 34 °C for *C. globosum* and 24 °C for *T. reesei*. The cultures were harvested when the peak cellulase activity was reached: 30 h for *M. thermophila*, 40 h for *T. terrestris* and *C. globosum*, and 70 h for *T. reesei*. The culture filtrates containing extracellular proteins were obtained by removing the mycelia and residual substrates by filtering through Miracloth (Calbiochem), and followed by centrifugation at 10,000 *g* for 30 min. Cleared supernatant fluids were concentrated using the Vivaflow200 (Sartorius Stedim) with a polyethersulfone membrane and 10 kDa cutoff. One (1) ml of supernatant was aliquoted into individual tubes containing 2% alfalfa (wt/vol). Reaction mixtures were incubated for 4 h at 40 °C, 50 °C, 60 °C and 70 °C. Reducing sugar was estimated by the 2,2'-bicinchoninate (BCA) method. Protein concentration

was determined by the Bradford method. The activity was calculated as mM of reducing sugar released per mg of protein.

Cloning and expression of xylanase genes. Genes of *M. thermophila* and *T. terrestris* predicted to encode endo-1,4-β-xylanase were cloned and expressed in *Aspergillus niger*. Briefly, oligonucleotides were designed to be complementary to the start and stop regions of the target genes, then used to PCR amplify cDNA prepared from the two thermophiles. The amplified cDNAs were cloned into the *A. niger* expression vector using the Gateway recombination method (Invitrogen) and used to transform *A. niger*.

Enzyme activity assays. Cellulase and xylanase activities were determined by measuring the reducing sugar released from the substrates using the BCA reagent in 96-well microplate format. Carboxymethyl cellulose (CMC-4M) and birchwood xylan, both obtained from Sigma-Aldrich, were used to determine the activity of cellulase and xylanase, respectively. The assays were performed in 50 mM Britton-Robinson buffer (50 mM boric acid, 50 mM acetic acid and 50 mM phosphoric acid) at the indicated temperatures for 30 min. Following incubation, 10 μl of the reaction mixture was added to 190 μl of BCA reagent, incubated at 80 °C for 40 min for color development, and the absorbance of the resultant mixtures was read at 562 nm. Glucose and xylose were used to prepare standard curves.

28. Jaffe, D.B. Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13**, 91–96 (2003).

29. Detter, J.C. *et al.* Isothermal strand-displacement amplification applications for high-throughput genomics. *Genomics* **80**, 691–698 (2002).

30. Smit, A.F.A., Hubley, R. & Green, P. RepeatMasker Open - 3.0. 1996–2010. <http://www.repeatmasker.org> (2010).

31. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).

32. Salamov, A.A. Ab initio gene finding in *Drosophila* Genomic DNA. *Genome Res.* **10**, 516–522 (2000).

33. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O. & Borodovsky, M. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res.* **18**, 1979–1990 (2008).

34. Birney, E. Using GeneWise in the *Drosophila* annotation experiment. *Genome Res.* **10**, 547–548 (2000).

35. Kent, W.J. BLAT—The BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).

36. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).

37. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int. J. Neural Syst.* **8**, 581–599 (1997).

38. Melén, K., Krogh, A. & von Heijne, G. Reliability measures for membrane protein topology prediction algorithms. *J. Mol. Biol.* **327**, 735–744 (2003).

39. Zdobnov, E.M. & Apweiler, R. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).

40. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

41. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004).

42. Koonin, E.V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).

43. Gene Onology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).

44. Enright, A.J., Van Dongen, S. & Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).

45. Roy, B.P. & Archibald, F. Effects of kraft pulp and lignin on Trametes versicolor carbon metabolism. *Appl. Environ. Microbiol.* **59**, 1855–1863 (1993).

46. Semova, N. *et al.* Generation, annotation, and analysis of an extensive *Aspergillus niger* EST collection. *BMC Microbiol.* **6**, 7 (2006).

47. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

48. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).

49. Hardcastle, T.J. & Kelly, K.A. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* **11**, 422 (2010).

50. Tsang, A., Butler, G., Powlowski, J., Panisko, E. & Baker, S. Analytical and computational approaches to define the *Aspergillus niger* secretome. *Fungal Genet. Biol.* **46**, S153–S160 (2009).

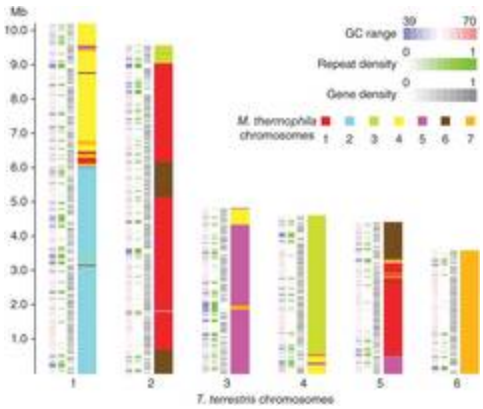


Figure 1

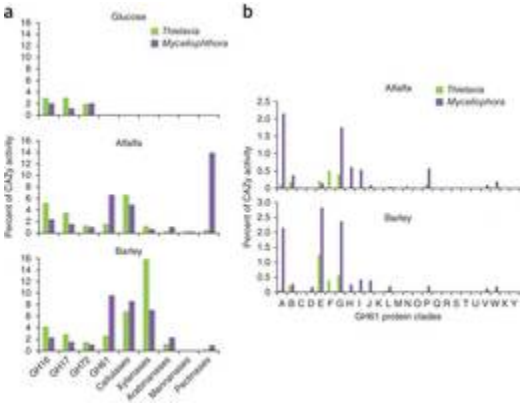


Figure 2

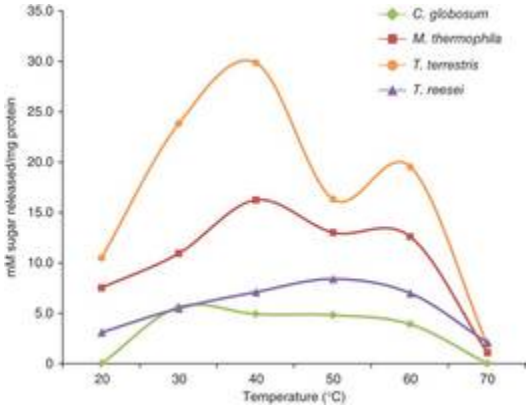


Figure 3

Supplement to: Comparative Genomic Analysis of the Thermophilic Biomass-Degrading Fungi *Myceliophthora thermophila* and *Thielavia terrestris*

Table of Contents

Supplementary Notes **2**
Analysis of GC content and amino acid biases 2
Evidence of RIP in thermophilic fungi 2
Chromatin structure and dynamics 3
Hydrolytic and oxidative enzymes in the exo-proteome 3
Growth and utilization of homogeneous carbohydrates 3
Chitinases 4
Proteases and peptidases 4
Oxidoreductases 5
Oxidative stress proteins 5
Transporters 6
Membrane responses to temperature 6
Fungal cell wall proteins 7
Secondary metabolism 7
Melanin pigment genes 8

Supplementary Methods **9**
Genome sequencing 9
Genome assembly 9
Construction and analysis of ESTs 9
Genome annotation 11
Transcriptome analysis 11
Analysis of *T. terrestris* extracellular proteins by mass spectrometry 12
Analysis of *M. thermophila* extracellular proteins by mass spectrometry 13
Prediction of transport proteins 14
Prediction of fungal cell wall proteins 14
Prediction of Proteins Involved in Chromatin Structure and Dynamics 14

References **15**

Supplementary Tables 1-4 **18**

Supplementary Tables 10-22 **23**

Supplementary Figures 1-9 **38**

Supplementary Tables 5-9 and 23-25, see accompanying Excel Tables

Supplementary Information

SUPPLEMENTARY NOTES

Analysis of GC content and amino acid biases

For analysis of GC content at the third position of codons, we aligned protein sequences of 6279 orthogroups (containing mutual best BLAST hits of each other with more than 40% percent identity) of three genomes (*M. thermophila*, *T. terresris,* and *C. globosum*) using MAFFT [1] and counted the total number of G/C and A/T nucleotides at corresponding third positions of codons for each trio of orthologous sequences. We estimated these numbers only on reliable parts of alignments: 1) with no gaps in the region [-3,+3] around the counted positions at the amino acid level, 2) with at least two of three sequences having the same amino acid at a given position; and 3) for codons, the two first positions of which were identical. In total, 75% of *M. thermophila* genes have a higher GC content at the third codon position relative to corresponding orthologs in *C. globosum*. The percentage is even higher for *T. terrestris* where 92% of its genes have a higher GC content at the third codon position than corresponding orthologs in *C. globosum*.

We computed amino acid composition for complete proteomes of 21 Ascomycete species, and for each pair of genomes we compared amino acid frequencies, computed based on orthologous pairs (reliable bidirectional best BLAST hits, with alignment coverage for both proteins > 95%). Recently, Wang and Lercher [2] observed another bias in amino acid composition of thermophilic prokaryotes, while controlling the confounding factors of phylogeny and GC content. Based on the previous work of Glyakina *et al*. [3] they found that thermophilic proteins are enriched in E, R and K residues and depleted in A, D, N, Q, T, S and H. Interestingly they found the same bias between 'warm-blooded' vertebrates, like mammals and 'cold-blooded' ones, like fish. "ERK - ADNQTSH" content computed for 16 Ascomycetes is presented in **Supplementary Table 14**. In the analysis we included only those proteins which have at least five orthologues in another 20 genomes. Again we did not see any differentiation of thermophilic fungi from others by this feature. Clearly the highest values were observed for *S. cerevisiae* and *S. pombe*, which are not thermophilic.

Another approach to analyze the potential amino acid adaptations in thermophiles is to align closely related thermophilic and mesophilic proteins to detect substitutional asymmetry, *i.e*., when certain aligned amino acids appear to occur substantially more often in either mesophilic or thermophilic proteins [4, 5, 6]. To accomplish this, we aligned 6279 orthogroups (with mutual best BLAST hits having more 40% percent identity) from three genomes and counted all possible 190 substitutions on reliable positions (positions with no gaps in the region [-3, +3] around counted positions) of the alignments. Additionally, we analyzed positions where the two thermophilic species have the same amino acid but differ from the corresponding residue in *C. globosum*. For each thermophilic - mesophilic pair (*M.t - C.g* and *T.t - C.g*) we found correspondingly 29 and 36 pairs of substitutions (out of 190) with significant deviation from an expected 1:1 ratio (Bonferroni-corrected chi-square test $P < 10^{-5}$). However, we also found 20 significantly asymmetric substitutions when comparing thermophilic pairs: *M. thermophila* and *T. terrestris*. As a control, we conducted the same analysis for three related *Trichoderma* genomes with similar GC content: *T. reesei, T. atroviride* and *T. virides*. In these genomes we also found up to 26 significantly asymmetric substitutions. Because none of the *Trichoderma* species are thermophilic and because many asymmetric substitutions coincide in the two analyses, we conclude that most of asymmetric substitutions are probably not related to high temperature adaptability.

Evidence of RIP in thermophilic fungi

An epigenetic mechanism termed RIP (repeat induced point mutation) has been found to operate in a number of Ascomycete fungi in which it is postulated to serve as a mechanism of genome defense against proliferation of selfish DNA elements such as transposons [7]. Both of the thermophilic

species (*M. thermophlia* and *T. terrestris*) examined in this study show evidence of directional mutation that may be attributed to RIP. Sizeable portions of the *T. terrestris* and *M. thermophila* genomes appear to comprise transposable elements degraded by RIP. In stark contrast, the non-thermophilic species *C. globosum* is the first member of the subphylum Pezizomycotina in which no evidence of RIP has been found. Additional analyses are presented in a recent publication by Clutterbuck [7].

Chromatin structure and dynamics

M. thermophila and *T. terrestris* each has at least 178 genes (**Supplementary Table 23**) with clear sequence identity to known yeast genes involved in chromatin structure or dynamics. Approximately 85% of the orthologous pairs (152/178) share 50% or greater identity at the protein level. Similar levels of identity were found in orthologous gene pairs with the closely related mesophile *C. globosum*. Interestingly, there are only two *M. thermophila* and three *T. terrestris* non-orthologous proteins involved in chromatin structure and dynamics. This high level of identity is consistent with the recent divergence of *M. thermophila, T. terrestris* and *C. globosum*. *S. cerevisiae* has 193 genes involved in chromatin structure and dynamics including one histone H1, and two each of histones H2A, H2B, H3 and H4 as noted in the *Saccharomyces* Genome Database [8]. The *M. thermophila* and *T. terrestris* genomes each encode one copy of H2B and H3, and two copies of H1, H2A and H4. The *M. thermophila* and *T. terrestris* genomes also contain H3 and H4 paralogs that differ significantly from the highly conserved sequences typical of H3 and H4. *M. thermophila* and *T. terrestris* have similar numbers of proteins involved in chromatin modification, 19 and 20 methyltransferases, 13 deacetylases and 15 acetyltransferases, respectively. Forty-six chromatin remodeling proteins were identified for both *M. thermophila* and *T. terrestris*, ten members of the SWI/SNF complex, four condensins, ten SAGA complex factors, six INO80 complex factors and one from the FACT complex [9, 10, 11].

Hydrolytic and oxidative enzymes in the exo-proteome

Carbohydrate-active enzymes (CAZymes) is a term that designates the enzymes that deconstruct (glycoside hydrolases, GHs; polysaccharide lyases, PLs; carbohydrate esterases, CEs) and assemble (glycosyltransferases, GTs) complex carbohydrates [12]. These enzymes are characterized by varying modular structures where the catalytic domains are appended to a variable number of other domains that can be catalytic or not. Frequently these ancillary domains display affinity for carbohydrates and they are designated carbohydrate-binding modules (CBMs) [12].

For the two largest enzyme categories, namely GHs and GTs, there appears to be a global increase in the number of enzymes with genome size (**Supplementary Fig. 2**); the dispersion appears larger with the GHs than with the GTs, and this reflects the fact that many GHs are secreted to the external milieu (for digestion of polysaccharides in particular) while the GTs perform more house-keeping functions (energy storage, fungal cell wall synthesis, protein N- and O-glycosylation).

The profiles of enzymes involved in the deconstruction of complex carbohydrates were analyzed by a double clustering procedure to compare the number of enzymes in each CAZyme family among various fungi (**Supplementary Fig. 3**). Despite differences their pectin-degrading enzymes, the two thermophiles clustered together, demonstrating the similarity of their overall CAZyme profiles.

Growth and utilization of homogeneous carbohydrates

The growth profiles of *T. terrestris* and *M. thermophila* were compared to each other and to six other filamentous ascomycetes (**Supplementary Fig. 4, 5**) on various carbon sources. Both *T. terrestris* and *M. thermophila* are true thermophiles, growing much better at 45°C than 34°C, and both species grow on a wide range of carbon sources (**Supplementary Fig. 4**; full profile available at www.fung-growth.org), in particular, sugars that are components of plant polysaccharides. A notable difference between the two species is that *T. terrestris* can grow on galacturonic acid (one of the main components

of pectin) as a sole carbon source, while *M. thermophila* cannot. Both species contain homologs for all the galacturonic acid catabolic pathway genes identified previously [13], suggesting that difference is not caused by the absence of galacturonic acid catabolism in *M. thermophila*. The two species are also similar in their ability to grow on plant polysaccharides with the exception of guar gum (a galactomannan). Growth of *M. thermophila* on this substrate is much better than that of *T. terrestris* which correlates with a higher content of GH26 and GH27 members in the *M. thermophila* genome. The absence of a gene encoding α-glucuronidase (GH67) in *T. terrestris* is surprising as this gene is commonly present in filamentous ascomycetes (www.cazy.org). The growth profile of the two thermophilic species was compared to six other filamentous ascomycetes. Since different fungal species typically exhibit varying growth rates, direct comparison of the species on a specific carbon source cannot be done. However, for all species glucose was the monosaccharide that promoted best growth. Therefore, in our analysis we compared the growth on a specific substrate to the growth on glucose and used the relative difference as a comparison between the species (**Supplementary Fig. 5**). *T. terrestris* and *M. thermophila* can be considered generalist fungi with respect to plant polysaccharide degradation, similar to *A. niger*, *C. globosum* and *H. jecorina*. Although the numbers of genes per CAZyme family differ significantly, all five species appear capable of producing a constellation of enzymes relevant to plant polysaccharide degradation. An exception to this is degradation of inulin, for which both thermophiles and *C. globosum* are poor with respect to both genome content and growth, and galactomannan as noted above. Interestingly, no GH32 candidates were detected for *T. reesei* [14] and *M. thermophila*, which may indicate that the associated activity arises from unknown proteins.

While both thermophiles grow well on pectin, they have developed significantly different approaches to degrading this polysaccharide. *T. terrestris* possesses two pectin lyases and seven GH28 pectin hydrolases, while *M. thermophila* has seven lyases and two GH28 hydrolases. This different strategy is reflected in their growth on pectin. Pectin lyases are most active at neutral to alkaline pH and *M. thermophila* grows best on pectin at alkaline pH (**Supplementary Fig. 4**). Pectin hydrolases are most active at acidic pH, which is reflected in the growth of *T. terrestris* on pectin that is best at acidic pH.

Chitinases

Chitin is the second most abundant polysaccharide in nature where it is a major component of fungal cell walls, the outer skeleton of insects, crabs, shrimp, lobsters and some internal structures in other invertebrates. As a β-1,4-linked polysaccharide composed of *N*-acetyl-glucosamine, it is the main raw material used for production of chitosan, which is employed in a number of applications, such as a flocculating agents, wound healing, sizing and strengthening of paper, and as a delivery vehicle for pharmaceuticals (http://www.euchis.org). In fungi, enzymes which break down chitin (collectively termed chitinases) are believed to have autolytic, nutritional, morphogical and mycoparasitic roles. Fungal chitinases may be comprised of up to five domains including a signal peptide, a catalytic domain, a chitin-binding domain, serine-threonine rich region(s), and carboxy-terminal extension. The *T. terrestris* genome encodes 14 putative chitinases belonging to family GH18, the *M. thermophila* genome harbors 10, and the non-thermophilic species *C. globosum* has 17. None of these three genomes encodes a GH19 protein, but each contains a single gene encoding an enzyme belonging to family GH20.

Proteases and peptidases

A total of approximately 150 peptidase sequences were identified in each genome (143 in *T. terrestris* and 159 in *M. thermophila*). The criterion for identification as peptidases was that they were identified by both InterProScan [15] and the MEROPS batch Blast [16] server (http://merops.sanger.ac.uk) as peptidases. There were approximately 50 additional sequences in each genome identified as peptidases only by InterProScan and not by MEROPS batch Blast. However, these were generally annotated with other activities in addition to peptidase and likely were not peptidases.

Supplementary Information

Of the 150 peptidases, approximately 25-30% had clearly identified signal sequences (46 for *M. thermophila* and 37 for *T. terrestris*). Signal sequences were identified by InterProScan. A protein was considered to have a signal sequence if it and its ten closest homologs had a signal sequence identified by InterProScan. Both organisms had similar sets of secreted peptidases (annotated information shown as part of the secretome analysis in the **Supplementary Tables 8, 9**). Only one of the 40 *M. thermophila* proteins with a signal sequence had no identifiable homolog in *T. terrestris*. That was Mycth_54466, whose nearest homolog (27% identical, 48% similar, 16% gap) in *S. cerevisiae* is YBR286W, a biochemically characterized aminopeptidase.

The group of proteins with signal sequences was enriched for aspartic and glutamic peptidases. Of all glutamic and aspartic peptidases identified, 100% and approximately 75%, respectively, had signal sequences. In contrast, only a quarter of serine peptidases, a fifth of metallo-peptidases and a tenth of cysteine peptidases had signal sequences. No threonine peptidases in either organism had signal sequences.

Oxidoreductases

Basidiomycete fungi produce a variety of extracellular oxidoreductases (e.g., lignin peroxidases and manganese peroxidases) that are believed to play a role in degradation of lignin. Although some Ascomycetes also secrete oxidoreductases, none have been directly implicated in lignin decomposition. We found no obvious homologs of the lignin peroxidases or manganese peroxidases in *M. thermophila* or *T. terrestris*. Nevertheless, we identified seven putative laccase genes in *T. terrestris* and seven in *M. thermophila*. Only three of these appear to be true orthologues of each other, showing 68-77% sequence identity. Each thermophile genome also encodes a likely cellobiose dehydrogenase (EC 1.1.99.18), an extracellular hemoflavoenzyme produced by several wood-degrading fungi [17] and also a likely copper radical oxidase (glyoxal oxidase). **Supplementary Table 15** enumerates the predicted extracellular oxidoreductases encoded in the genomes of five Ascomycete fungi.

Oxidative stress proteins

Oxidative stress is associated with growth in an oxygen rich environment. Reactive oxygen species (ROS) are generated as a by-product of respiratory metabolism and they must be eliminated through the action of a variety of enzymes including superoxide dismutases, peroxidases, catalases and alternative oxidase to name a few. To determine if oxidative stress mechanisms were a point of difference between fungal thermophiles and mesophiles the genes encoding various oxidative stress response proteins were examined.

The genes for all of these enzymes were retrieved from each of the three genomes examined. The results of this analysis are shown in **Supplementary Table 16**. The two main families of superoxide dismutases (SOD, EC 1.15.1.1) are the Cu-Zn SODs and the Fe/Mn SODs where the former contain both metals and the latter contain either Fe or Mn. They catalyze the disproportionation of two molecules of superoxide to molecular oxygen and hydrogen peroxide. There are a total of six SOD genes in both *C. globosum* and *T. terrestris*, whereas, *M. thermophila* has five, missing the Cu-Zn SOD orthologue that is predicted to be secreted.

A large variety of enzymes broadly termed peroxidases (EC 1.11.1._) are capable of reducing peroxides to water, alcohols or oxygen. Catalase and catalase-peroxidase (both EC 1.11.1.6) are the only peroxidases that produce oxygen by the reaction of two molecules of hydrogen peroxide to form molecular oxygen and water. Genes that fit into the broad category of peroxidases were compared across the three genomes, including the catalases, catalase-peroxidases and other peroxidases. The results summarized in **Supplementary Table 17** show that there are some differences in copy number, *C. globosum* has the most, but nothing indicative of a pattern associated with a thermophilic lifestyle. Other peroxidases have one gene copy in each of the three genomes, including glutathione peroxidase

(EC 1.11.1.9), peroxiredoxin (EC 1.11.1.15) and cytochrome-c peroxidase (EC 1.11.1.5). The extra *C. globosum* catalase gene (PID 13820) is unusual in that it most closely aligns with Eurotiomycete sequences, not the Sordariomycetes to which the three fungi examined in this work belong.

Transporters

Bioinformatics analysis of membrane transporters in the genomes of *M. thermophila* and *T. terrestris* identified a total of 201 and 496 predicted cytoplasmic transporters, respectively (**Supplementary Tables 18, 24, 25**) Both genomes encode a broad array of transporters for the uptake of sugars and sugar-phosphates, amino acids, oligopeptides, carboxylates, and nucleosides. Additionally, their genomes encode a large number of major facilitator superfamily (MFS) uptake transporters of unknown specificity, suggestive that they may be capable of uptake of a range of more esoteric carbon sources. Both organisms also include a swathe of multidrug efflux transporters that are presumably involved in secretion of secondary metabolites and protection against exogenous toxic compounds.

The lower number of transporters encoded by *M. thermophila* compared with *T. terrestris* is largely due to decreased numbers of paralogues in large transporter families, for example, 86 *M. thermophila* MFS transporters compared to 221 members in *T. terrestris*. Similarly, there are less than half as many ATP Binding Cassette (ABC) Superfamily transporters encoded by the *M. thermophila* genome (19 versus 44). *T. terrestris* does possess a variety of predicted transport capabilities that are not present in *M. thermophila* including transporters for arsenite (ArsB family), chromate (Chr family), tellurite (TDT family); heavy metals (VIT, NRamp and ILT families); and sodium and calcium ion channels (VIC and Annexin families). *M. thermophila* may be able to transport tricarboxylates through an MTC family transporter which does not have an orthologue in *T. terrestris*.

Membrane responses to temperature

Common strategies for adaptation of the lipid membrane to tolerance of high or low temperatures include changes in sterol content, the ratio of saturated to unsaturated fatty acids, or alterations to fatty acid chain length [18, 19, 20]. Genes encoding enzymes required for ergosterol biosynthesis, fatty acid desaturation, and fatty acid elongation were identified in the *M. thermophila* genome by using BLAST searches with *S. cerevisiae* gene sequences, as well as text searching of the automated annotation. Homologs in *M. thermophila* and *T. terrestris* were identified using the "Cluster" tool on the JGI website.

Each species appears to harbor a complete set of genes for ergosterol biosynthesis (**Supplementary Table 19**). In general, one gene per enzyme activity is present, although there are some exceptions: for example, *M. thermophila* has two genes encoding mevalonate kinase-like proteins, whereas, *T. terrestis* has only one. In temperature-shift transcriptome experiments with either species, none of the transcript levels for the ergosterol pathway genes varied by more than a factor of two in either direction (data not shown). This suggests that changes in ergosterol content of the membrane are not especially significant at the two growth temperatures tested (34°C and 45°C).

Two genes encoding fatty acid elongase (or elongase-like) proteins are present in the genomes of each species (**Supplementary Table 20**), and the transcript levels of these also did not vary appreciably in the temperature shift experiments (data not shown).

Four genes encoding fatty acid desaturases were identified in *M. thermophila,* and three in *T. terrestris*: in addition each genome has one sphingolipid delta-4 desaturase (**Supplementary Table 20**). In *M. thermophila* the transcriptional level of one of these (Mycth_50837) was among the 15 most down-regulated in the switch from low to high temperature (data not shown). A lower content of unsaturated fatty acids would contribute to lower fluidity at higher temperatures, and such observations have been made for a number of thermophilic bacteria and fungi. A similar difference was not observed for any of the fatty acid desaturase genes from *T. terrestris*.

Supplementary Information

Fungal cell wall proteins

Fungal cell walls are mainly polysaccharides, but the synthesis, remodeling and degradation of these polysaccharides are performed by enzymes. Other structural proteins may crosslink the polysaccharides or modify wall properties such as hydrophobicity. Remodeling and degradative enzymes are located in the cell wall outside the plasmalemma, but many of the polysaccharide synthases are embedded in the cell membrane. Certain cell wall proteins are anchored to the outer face of the cell membrane through glycosylphosphatidylinositol (GPI) links; these proteins have both N-terminal signal peptides and C-terminal hydrophobic regions that facilitate linkage to the membrane.

In filamentous ascomycetes the major cell wall polysaccharides include chitin, 1,3-β-glucan, 1,3-β-/1,4-β-glucan, and 1,3-α-glucan. Chitin (poly-(1,4-β-N-acetylglucosamine)) is synthesized by membrane-embedded chitin synthases which extrude the growing chitin chains into the extracellular space. Seven classes of chitin synthase are known in *A. nidulans*; each class has one representative in *M. thermophila* and *T. terrestris*. Chitin is depolymerized by chitinases belong to GH18. *A. nidulans* has 15 annotated chitinases that cluster in 10 groups; *T. terrestris* has 14 similar sequences, and *M. thermophila* has only ten (**Supplementary Table 21**). Despite the similar numbers of chitinases in *A. nidulans* and *T. terrestris*, the correspondence between the two sets of proteins is not one-to-one. For example, there is a cluster of three *Aspergillus* chitinase genes surrounding *chiC*, but only one *T. terrestris* gene model, and no *M. thermophila* model, aligns closely with this group. Conversely, three *T. terrestris* gene models and two *M. thermophila* models cluster tightly with *A. nidulans chiB*.

The enzymes responsible for 1,3-β-glucan synthesis are not well characterized. There is a single putative 1,3-β-glucan synthase in *A. nidulans* (*fksA*) and in *N. crassa* (GLS1) with single homologs in *T. terrestris* (Thite132141) and *M. thermophila* (Mycth140260). For hydrolysis of 1,3-β-glucan, there is one endo-1,3-β-glucanase known in *A. nidulans* (*engA*) and other *Aspergillus* species; *M. thermophila* and *T. terrestris* each have a similar protein. There are also two families of exo-1,3-β-glucanases – *exgA, B, C, D*, and *E* (GH5) and six members of GH55 – known in *A. nidulans*. Two members of the GH5 exo-1,3-β-glucanase family can be found in *M. thermophila* and three in *T. terrestris*. In addition, *M. thermophila* has four members and *T. terrestris* has five members of the GH55 exo-1,3-β-glucanase family. Three families of 1,3-β-transglucosidases (Bgl2, GH17 and Gas, GH72) and 1,3-β-transglycosidases (Crh1, GH16) crosslink the 1,3-β-glucan chains to each other and to chitin. *M. thermophila* and *T. terrestris* both have four orthologous proteins similar to Gas, and four orthologous proteins similar to Crh1. *M. thermophila* and *T. terrestris* each have orthologous proteins similar to EglC, BtgC, and BtgE in the Bgl2 family; *M. thermophila* has two more proteins similar to EglC with no orthologue in *T. terrestris*.

Secondary metabolism

T. terrestris and *M. thermophila* contain less than remarkable numbers of polyketide synthase or non-ribosomal peptide synthetase (NRPS) genes. The genome of *T. terrestris* possesses three NRPS genes and one hybrid NRPS-PKS gene, whereas, *M. thermophila* encodes five NRPSs and three NRPS-PKS hybrids. Eleven PKS encoding genes were found in *T. terrestris* compared to nine in the genome of *M. thermophila.* Interestingly, both genomes encode PKS genes (Thite_35447 and Mycth_101261) that, based on high amino acid identity (~60%), are possibly orthologues of the predicted octaketide producing PKS from *Aspergillus nidulans* (AN0150.2) which is responsible for production of emodin, emodin-derivatives and monodictyphenone [21, 22]. The other PKS genes have putative orthologues in other fungi but were not associated with specific compounds. The genomes of both thermophiles encode putative orthologues of LaeA (Thite_2121390 and Mycth_2294559), the global regulator of secondary metabolism in *Aspergillus* species [23] suggesting a similar mechanism controlling secondary metabolite gene clusters among these diverse ascomycete species.

Supplementary Information

Melanin pigment genes

Melanogenesis is required in development, stress management and pathogenesis in filamentous ascomycetes, and is thus considered a key secondary metabolic pathway. *T. terrestris* possesses the genetic machinery required for melanin production, and certain genes of this pathway show a physical clustering. *T. terrestris* encodes two genes (Thite133621 and Thite153956) that are closely related to melanogenesis master regulatory genes in *Magnaporthe grisea* and *Cocholiobolus heterostrophus* (*pig-1* and *cmr-1*, respectively), organisms in which this biosynthetic pathway has been characterized. In total, 18 predicted genes in *T. terrestris* showed significant similarity to *pig-1* and *cmr-1*, and each of these ORFs contained either *GAL4*-like and/or zinc-finger DNA binding domains. At least 19 BUF1 (*M. grisea*)/*PKS1*8 (*C. heterostrophus*) homologs were also evident in *T. terrestris*; these sequences encode a polyketide synthase I (PKS I), and seven *T. terrestris* predicted genes (Thite 2109607, Thite52153, Thite2106274, Thite13940, Thite44861, Thite35447 and Thite2110824) having appreciable sequence similarity to PKS I genes with confirmed roles in melanin biosynthesis in other organisms. Additionally, the *T. terrestris* genome contains homologs of scytalone dehydratase I (Thite2120303), hydroxy naphthalene reductases (approximately 47 homologs, including two sequences, Thite2145571 and Thite2130018, showing significant phylogenetic affinity to *C. heterostrophus* melanogenesis genes BRN1 and BRN2), and tyrosine/polyphenoloxidases (~10 homologs), where four predicted genes, (Thite2131928, Thite62739, Thite124715 and Thite2118068) were similar to known *M. grisea* tyrosinase sequences). Initial investigations also revealed that *T. terrestris* homologs potentially involved in melanin biosynthesis appear to show at least partial clustering; chromosome 3 contains at least four (of 16 total) potential melanin biosynthesis genes [PKS (Thite52153), tyrosinase (Thite2118068) and two naphthalene reductases (Thite2145571 and Thite2130018)] distributed over approximately 1 MB. Chromosome 1 contains copies of *cmr-1*-related predicted gene products (Thite2106273 and Thite2083982) and PKS homologs (Thite2109607 and Thite2110824). In sum, the presence of *T. terrestris* melanogenesis genes is consistent with phenotypic observations, where species of *Thielavia* produce pigmented perithecia and ascospores, suggesting a role for melanin in development. The extent to which this molecule is involved in general stress management in this, and other members of the Chaetomiaceae, is unclear. Genomic co-localization of some of the melanin biosynthesis genes on Chromsomes 1 and 3 in the *T. terrestris* genome points to the possible evolution and/or maintenance of some degree of clustering in this, and possibly other secondary metabolic pathways, reminiscent of similar patterns in the distantly-related *M. grisea*, *C. heterostrophus* and filamentous ascomycetes, generally.

Supplementary Information

SUPPLEMENTARY METHODS

Genome Sequencing

All sequencing reads for the whole genome shotgun sequencing were collected with standard Sanger sequencing protocols on ABI 3730XL capillary sequencing machines. For both genomes, three different sized libraries were used as templates for the plasmid subclone sequencing process, and both ends were sequenced. *M. thermophila*: 202,837 reads from the 3.3 kb sized library, 211,776 reads from the 6.7 kb sized library, and 54,528 reads from a 36.0 kb fosmid library. *T. terrestris:* 339,194 reads from the 2.5 kb sized library, 271,871 reads from the 8.7 kb sized library, and 88,508 reads from a 33.3 kb fosmid library were sequenced (**Supplementary Table 22**).

In order to improve and finish the genomes of *M. thermophila* and *T. terrestris,* we broke the whole genome shotgun assembly into scaffold-sized pieces, and each piece was reassembled using PHRAP [24]. We then finished these scaffold pieces using our pipeline based on PHRED/PHRAP/CONSED [25]. Initially, we targeted all low quality regions and gaps with computationally selected sequencing reactions completed with 4:1 BigDye terminator: dGTP chemistry (Applied Biosystems). These automated rounds included directed primer walking on plasmid subclones using custom primers.

After the completion of the automated rounds, a trained finisher manually inspected each assembly. We selected further reactions manually to complete the genome. These reactions included additional custom primer walks on plasmid subclones or fosmids. We sequenced these templates using 4:1 BigDye terminator:dGTP chemistry. We resolved smaller repeats by transposon-hopping with 8 kb plasmid clones. We shotgun sequenced and finished fosmid clones to fill large gaps, resolve larger repeats or to resolve chromosome duplications and extend into telomere regions.

We validated each assembly by independent quality assessment that included a visual examination of subclone paired ends and visual inspection of high quality discrepancies and all remaining low quality areas. All available EST resources were also placed on the assembly to ensure completeness. The finished genomes of *M. thermophila* and *T. terrestris* consist of seven and six chromosomes, respectively, comprising 38,744,216 bp and 36,912,256 bp of finished sequence with an estimated error rate of less than one error in 100,000 base pairs (**Table 1**).

Genome assembly

For both genomes**,** the sequencing reads were assembled using a modified version of ARACHNE v.20071016 [26] with parameters maxcliq1=100, correct1_passes=0 and BINGE_AND_PURGE=True. For *M. thermophila*, this produced 11 scaffold sequences with an L50 of 5.4 Mb, eight scaffolds larger than 100 kb, and total scaffold size of 38.8 Mb. For *T. terrestris*, 12 scaffold sequences were produced with an L50 of 4.6 Mb, 8 scaffolds larger than 100 kb, and total scaffold size of 37.0 Mb. Each scaffold was screened against bacterial proteins, organelle sequences and GenBank and removed if found to be a contaminant. Additional scaffolds were removed if the scaffold contained only unanchored rDNA sequences.

Construction and analysis of ESTs

Poly A+ RNA was isolated from total RNA (pooled RNA from cells grown in MY50 (rich medium) for *T. terrestris* and 1% cellulose and 1% pectin pooled culture from *M. thermophila*) using the Absolutely mRNA Purification Kit and manufacturer's instructions (Stratagene, La Jolla, CA). cDNA synthesis and cloning was a modified procedure based on the "SuperScript plasmid system with Gateway technology for cDNA synthesis and cloning" (Invitrogen, Carsbad, CA). Approximately 1-2 µg of poly A+ RNA, reverse transcriptase SuperScript II (Invitrogen) and oligo dT-*Not*I primer (5' GACTAGTTCTAGATCGCGAGCGGCCGCCCT15VN 3') were used to synthesize first strand cDNA. Second

strand synthesis was achieved with *E. coli* DNA ligase, polymerase I, and RNaseH followed by end repair using T4 DNA polymerase. The *Sal*I adaptor (5' TCGACCCACGCGTCCG and 5' CGGACGCGTGGG) was ligated to the cDNA, digested with *Not*I (New England Biolabs, Ipswich, MA), and size-fractionated by gel electrophoresis (1.1% agarose). Size ranges of cDNA were excised from the gel for the RNA sample yielding two cDNA libraries (0.6 – 2 kb and >2 kb). The cDNA inserts were directionally cloned into the *Sal*I and *Not*I digested vector pCMVsport6 (Invitrogen). The ligation reactions were used to transform *E. coli* ElectroMAX T1 DH10B competent cells (Invitrogen).

Library quality was first assessed by randomly selecting 24 clones and PCR amplifying the cDNA inserts with the primers M13-F (5' GTAAAACGACGGCCAGT) and M13-R (5' AGGAAACAGCTATGACCAT) to determine the fraction of insert-less clones. Cells from each library were plated onto agarose plates (254 mm plates from Teknova, Hollister, CA) at a density of approximately 1000 colonies per plate. Plates were grown at 37°C for 18 hours then individual colonies were picked and used to inoculate a well containing LB medium [27] with appropriate antibiotic in a 384 well plate (Nunc, Rochester, NY). Clones in 384 well plates were grown at 37°C for 18 hours. Plasmid DNA for sequencing was produced by rolling circle amplification [28] (TempliPhi, GE Healthcare, Piscataway, NJ). Subcloned inserts were sequenced from both ends using primers complementary to the flanking vector sequence (Fw: 5' ATTTAGGTGACACTATAGAA 3' Rv: 5' TAATACGACTCACTATAGGG 3') and Big Dye terminator chemistry with ABI 3730 instruments (Applied Biosystems, Foster City, CA).

To trim vector and adaptor sequences, common sequence patterns at the ends of ESTs were identified and removed using an internally developed tool. Insert-less clones were identified if either of the following criteria were met: >200 bases of vector sequence at the 5' end or less than 100 bases of non-vector sequence remained. ESTs were trimmed for quality using a sliding window trimmer (window = 11 bases). Once the average quality score in the window was below the threshold (Q15) the EST was split and the longest remaining sequence segment was retained as the trimmed EST. EST sequences with less than 100 high quality bases were removed. ESTs were evaluated for the presence of polyA or polyT tails (which if present were removed) and the EST reevaluated for length, removing ESTs with less than 100 bases remaining. ESTs consisting of more than 50% low complexity sequence were also removed from the final set of "good ESTs." In the case of re-sequencing, the longest high quality EST was retained. Sister ESTs (end pair reads) were categorized as follows: if one EST was insert-less or a contaminant then by default the second sister was categorized as the same. However, each sister EST was treated separately for complexity and quality scores. Finally, EST sequences were compared to the Genbank nucleotide database in order to identify contaminants; non-desirable ESTs such as those matching non-cellular and rRNA sequences were removed.

For clustering, ESTs were evaluated with MALIGN, a kmer based alignment tool (Chapman, unpublished), which clusters ESTs based on sequence overlap (kmer = 16, seed length requirement = 32 alignment ID >= 98%). Clusters of ESTs were further merged based on sister ESTs using double linkage. Double linkage requires that two or more matching sister ESTs exist in both clusters to be merged. EST clusters were then each assembled using CAP3 [29] to form consensus sequences. Clusters may have more than one consensus sequence for various reasons to include; the clone has a long insert, clones are splice variants or consensus sequences are erroneously not assembled. Cluster singlets are clusters of one EST, whereas CAP3 singlets are single ESTs which had joined a cluster but during cluster assembly were isolated into a separate singlet consensus sequence. ESTs from each separate cDNA library were clustered and assembled individually, and subsequently the entire set of ESTs for all cDNA libraries were clustered and assembled together. For cluster consensus sequence annotation, the consensus sequences were compared to Swissprot [30] using BLASTX [31] and the hits were reported. Clustering and assembly of all 33,539 *Thielavia* ESTs resulted in 11,186 consensus sequences and 1,940 singlets. Clustering and assembly of 33,539 *Myceliophthora* ESTs resulted in 10,567 consensus sequences and 2,459 singlets.

Supplementary Information

Genome annotation

Genomic assembly scaffolds were masked using RepeatMasker [32] and the RepBase library of 234 fungal transposable elements [33]. Gene modeling on that repeat-masked assembly was performed *ab initio* Fgenesh [34] and Genemark-ES [35]; homology-based Fgenesh+ [34] and Genewise [36] seeded by BLASTx alignments of NCBI's nr (non-redundant) protein database against the assembly; cDNA-based EST_map (http://www.softberry.com/) seeded by EST contigs. Genewise models were extended where possible using scaffold data to find start and stop codons. EST BLAT alignments [37] were used to add or extend exons for gene models, including the addition of 5' and/or 3' untranslated regions. Since multiple gene models were generated for each locus, a single representative model was algorithmically chosen based on model quality. Measures of model quality for this initial 'GeneCatalog' set included proportions of the models complete with start and stop codons (89-95% of models were observed complete in *M. thermophila* and *T. terrestris*), consistency with ESTs and EST assembly consensus sequences (25-39% of models were covered≥75%), support by protein similarity with NCBI nr database proteins (91 -95% of models), and matches to Pfam domains (48-59% of models) [38]. Using tRNAscan-SE [39] 97 tRNAs were predicted in *T. terrestris*, and 204 in *M. thermophila*. Herein, however, the term gene model refers to protein coding genes unless otherwise noted. Subsequent to gene modeling, RepeatScout [40] was used for additional repeat family-detection and estimation of repeat coverage of the genome; this later-stage repeat detection was not included in the masked assembly sequence. The *C. globosum* genome assembly and gene models, used for comparison to the two thermophiles, were downloaded from the Broad Institute *Chaetomium globosum* Database at http://www.broadinstitute.org/annotation/genome/chaetomium_globosum.

All predicted gene models were functionally annotated using SignalP [41], TMHMM [42], InterProScan[15], BLASTp [31] against the nr database, and hardware-accelerated double-affine Smith-Waterman alignments (deCypherSW; http://www.timelogic.com/decypher_sw.html) against SwissProt (http://www.expasy.org/sprot/), the Kyoto Encyclopedia of Genes and Genomes (KEGG) [43], and the eukaryotic orthologous groups of proteins database (KOG) [44]. The Enzyme Commission (EC) numbers (http://www.expasy.org/enzyme/) of KEGG hits were assigned to gene models and mapped to corresponding KEGG pathways. InterPro and SwissProt hits were used to assign Genome Ontology (GO) terms [45] to gene models. Multigene families were predicted with the Markov clustering algorithm (MCL) [46], using BLASTp alignment scores between proteins as a similarity metric. Manual curation of automated annotations was performed using the JGI Genome Portal's web-based interactive tools to edit predicted gene structures, assign gene functions, and report supporting evidence.

Tandem genes were detected as genomic fragments with two or more adjacent genes homologous to each other (*E*-value threshold 1e-05 and alignment coverage for all proteins > 50%), with maximum allowance of two non-homologous genes in a given tandem gene region. Segmental duplications were selected as duplicated genome fragments with minimum of three genes in each fragment with at least of 50% of genes between fragments being homologs to each other.

Transcriptome analysis

For RNA extraction fungal spores were inoculated to a final concentration of 2.5 x 10^5/ml in 10x TDM [47] containing 2% glucose or 2% agricultural straws (alfalfa or barley straws ground to 0.5 cm lengths), and incubated at either 34°C or 45°C with shaking at 150 rpm. Mycelia were harvested at early growth phase; 21 h for *M. thermophila* at both temperatures, and 24-28 h at 45°C and 56 h at 34°C for *T. terrestris*. Total RNA samples were isolated from mycelia as described [48]. Sequencing was performed using the RNA-Seq method of Illumina's Solexa IG at either the DNA Core Facility of the University of Missouri or at the McGill University-Génome Québec Innovation Centre.

Supplementary Information

In the first step of analyzing the RNA-Seq data, the 38-42 nt reads were mapped to the genomic sequence assemblies of the two thermophiles. To identify splice junctions, all the available reads for each species were aligned to its genome with BLAT, requiring at least 93% sequence identity and intron length less than 2000 nt. The best alignments for each read were selected with pslCDnaFilter [37]; only reads that aligned over 95% or more of their length were kept. After the mapped reads were sorted by genomic position, potential splice junctions were extracted from the reads with gapped alignments, and filtered. To pass the filter, the splice junctions needed to have introns with GT-AG, GC-AG, or AT-AC donor-acceptor pairs, and to score above 0.5 in a discriminant function that combined the donor-acceptor type, the number and distribution of reads spanning the junction, the number of reads mapping inside the intron relative to the number of spanning reads, and the presence of overlapping splice junctions. The coefficients of the discriminant function were estimated by logistic regression on a set of potential junctions that had been manually classified as valid or invalid. A file containing spliced sequences specific to the length of the reads to be mapped was generated by concatenating flanking genomic sequences of length 1 less than the read length from both sides of the splice; if another splice site occurred within this flanking sequence, variants with spliced and unspliced flanking sequences were both generated. The reads from each mRNA sample were then remapped against a combination of the genomic sequence and the spliced sequences with Bowtie [49] using the 'best' strata option. Mappings to the spliced sequences were translated into gapped alignments to the genomic sequence, and the mapped reads were separated into those with a unique best mapping (ca. 90%) and those with more than one equally good mapping. Reads with >25 best mappings were discarded; reads with 2-24 best mappings were randomly assigned to one of their possible mappings with probability proportional to the density of uniquely mapped reads around each position. The depth of mapped read coverage at each genome position was calculated using the WIGGLES program bundled with TopHat [50].

To analyze differential expression, the transcript and exon definitions were taken from the filtered models of version 1 of genome annotation. The number of reads mapping to each transcript was estimated by integrating the coverage depth over the annotated exons of the transcript, dividing by the read length, and rounding to an integer. The Bayesian posterior probability of differential expression was estimated from the read counts for each transcript using the R package baySeq v 1.4 [51]. To aid interpretation, FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values were also calculated from the counts using the transcript lengths and the total number of mapped reads from each sample.

Analysis of *T. terrestris* extracellular proteins by mass spectrometry

An amount of protein equal to 400 µg from the concentrated samples collected after 96 h growth in xylose and cellulose was precipitated using methanol-chloroform and resolved by two-dimensional gel electrophoresis. The first dimension, isoelectric focusing (IEF), was performed by cup-loading the redissolved protein onto ReadyStrip IPG strips pH 4-7, 11 cm (Bio-Rad). All gels were run in duplicate. IEF was then done using an IPGphor system (Pharmacia Biotech). Following this, the second dimension, SDS-PAGE was run in a Criterion Dodeca Cell (Bio-Rad). Gels were stained with Coomassie Blue and scanned using a Perfection V750 Pro scanner (Epson) and images prepared with Adobe Photoshop version 6.0.1 (Adobe Systems Inc.) and Photodraw 2000 version 2.0.0.0915 (Microsoft Corporation). Up to 96 spots were excised from 2D gels using a ProteomeWorks Spot Cutter (Bio-Rad). Proteins were digested with trypsin using a MassPrep Station robotic liquid handling system (Micromass). Proteins in the gel spots were reduced with 50 µl of 10 mM DTT in 100 mM ammonium bicarbonate (AMBIC) for 30 min, and then alkylated with 50 µl of 55 mM iodoacetamide in 100 mM AMBIC for 20 min. The dried gel spots were next re-hydrated in trypsin digestion solution (6 ng/l sequencing grade trypsin in 50 mM AMBIC) for 30 min at ambient temperature, and then the proteins digested for an additional 8 h at 40°C. Acetonitrile (50 µl) was used to de-hydrate gel spots between reactions, and they were air-dried. Once

digested, peptides were extracted twice with 1% formic acid/2% acetonitrile for 30 min. A Synapt MS mass spectrometer (Waters Corporation) with a NanoAcquity UPLC (Waters Corporation) was used for nano-UPLC-MS/MS analysis of digested peptides. The Synapt MS was controlled using MassLynx software version 4.1 (Waters). Samples were loaded onto a Symmetry C18 trapping column (180 m ID X 20 mm, 5 m; Waters), fitted in the injection loop and washed with 0.1% formic acid in water at 15 µl per minute for 1 min. Peptides were separated on a BEH13 C18 nanoflow fused capillary column (100 µm ID x 100 mm, C18, 1.7m, Waters) at a flow rate of 400 nl/minute. A step elution gradient of 1% to 85% acetonitrile in 0.1% formic acid was applied over a 40 minute interval. The column eluent was monitored at 214 nm and introduced into the Synapt MS through an electrospray ion source fitted with nanospray interface. Data were acquired in survey scan mode from a mass range of m/z 400 to 1990 with switching criteria for MS to MS/MS to include an ion intensity of greater than 10.0 counts per second and charge states of +2, +3, and +4. Analysis spectra of up to six co-eluting species with a scan time of 1.9 seconds and inter-scan time of 0.1 seconds could be obtained.

Raw data files were processed into peak lists using ProteinLynx Global Server version 2.4 (Waters). The assembled peak lists where then analyzed using Mascot (Matrix Sciences) and X!Tandem (The GPM, thegpm.org; version 2007.01.01.1) software searching the *T. terrestris* version 1.0 filtered gene model database. The database originally downloaded contained 9,815 gene model sequences, to which were added sequences for trypsin and keratin, for a total of 9,817 sequences. Search parameters included: (1) trypsin digestion with a maximum of one missed cleavage allowed, (2) fixed carbamidomethyl (C) and variable oxidation (M) modifications, (3) peptide tolerance of 0.3 Da, and (4) MS/MS tolerance of 0.3 Da. Search results were then analyzed using SCAFFOLD software (Proteome Software Inc., version SCAFFOLD_3_00_05) to validate protein identifications. Protein identifications were accepted if they had a protein probability≥95% assigned by the Protein Prophet algorithm, a minimum of two unique peptides, and peptides with probabilities≥95% assigned b y the Peptide Prophet algorithm.

Analysis of *M. thermophila* extracellular proteins by mass spectrometry

Proteins in supernatants were acetone precipitated prior to determination of protein contents using the 2D-Quant kit (GE Healthcare Life Sciences). Equal amounts of proteins (45 µg) were fractionated by SDS-PAGE on 2.4 cm long, 7-15% polyacrylamide gels. Proteins were stained with Coomassie Brilliant Blue G (Sigma). Each lane was excised into 12 bands on a robotic workstation (ProXCISION™ Proteomics Gel Cutting Robot, (PerkinElmer Life and Analytical Sciences). Following the transfer of gel pieces to a 96-well tray, proteins were subjected to reduction, cysteine-alkylation and in-gel tryptic digestion by on a MassPrep Workstation (Waters) as described previously [52]. Peptides were analyzed using a Nanopump series 1000 (Agilent Technologies) coupled to a Q-TOF *micro*™ (Waters) equipped with a Nanosource modified with a nanospray adapter (New Objective). A volume of 20 µl of peptide extracts was applied onto a PicoFrit™ column (75 µm id, New Objective), filled with BioBasic® C18 packing (10 cm bed length, 5 µm, 300 Å) as described [53]. Gradient length was set at 60 min. Mass spectrometric data were acquired by Data Directed Analysis on MassLynx (Waters) with a 1,1,4 duty cycle (1 s in MS Survey mode, 1 peptide selected for fragmentation, maximum of 4 s in MS/MS acquisition mode). Doubly and triply charged ions were selected for fragmentation.

Distiller version 2.0.0 (Matrix Science) was used to generate the peak list, with the following detection parameters: correlation threshold 0.4; minimum S/N 5; precursor selection tolerance 3 Da. The concatenated peak list were searched against a copy of *M. thermophila* filtered gene model version 2.0 database (9110 sequences), using MASCOT CLUSTER version 2.1.04 (Matrix Science) with the following parameters: trypsin; one missed cleavage; fixed carbamidomethyl alkylation of Cys; variable oxidation of Met; 0.5 mass unit tolerances on parent and fragment ions; monoisotopic. Peptide search files were loaded onto SCAFFOLD version 3 (Proteome Software), in order to validate identification and use spectral

counting as the method for quantitative proteomics. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm [54]. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped by the algorithm to satisfy the principles of parsimony. Statistics were performed with SCAFFOLD; *p* values on *t*-tests equal to or less than 0.05 were considered significant. SCAFFOLD was also used to calculate fold changes.

Prediction of transport proteins

Complete protein sequence datasets from both genomes were analyzed using the TransAAP pipeline [55] for their predicted complement of membrane transport proteins. This approach combines BLAST searches against a curated membrane transport protein database (Transport DB), as well as HMM searches and COG-based searches against membrane transporter protein families. Membrane transporters were assigned to protein families based on sequence similarities [55].

Prediction of fungal cell wall proteins

In the absence of any experimental study of *M. thermophila* and *T. terrestris* cell walls, identification of probable cell-wall proteins depended on homology to previously characterized proteins in the walls of related fungi. For *M. thermophila* and *T. terrestris*, the closest relative with well-characterized cell wall proteins is *Aspergillus*, especially *A. nidulans* [56]. A proteomic analysis of proteins extracted from cell walls of *N. crassa* [57] provided complementary evidence. Putative cell-wall proteins were found by BLASTP searching of the predicted proteins of *M. thermophila* and *T. terrestris* with the identified cell wall proteins of *A. nidulans* and *N. crassa* as queries. To identify orthologs within protein families, the sequences of the query proteins and the BLASTP hits were aligned with T-COFFEE, and an average distance tree (using BLOSUM62 distances) was calculated from the multiple alignments with JALVIEW. GPI anchor sites were predicted with Big-Pi [58].

Prediction of Proteins Involved in Chromatin Structure and Dynamics

The search was initiated using the KOG database [59]. All gene models in the COG Chromatin Structure and Dynamics section were analyzed using BLASTP [31] and CDD (conserved domain database) [60]. Function was assigned based either on having over 45% identity at the protein level as determined using BLASTP or a conserved domain as identified by CDD or both. A local BLASTP was also preformed with the downloaded genomes of *M. thermophila, T. terrestris* and *C. globosum.*

REFERENCES

1. Katoh, K. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**, 511-518 (2005).
2. Wang, G.-Z. & Lercher, M.J. Amino acid composition in endothermic vertebrates is biased in the same direction as in thermophilic prokaryotes. *BMC Evolutionary Biology* **10**, 263 (2010).
3. Glyakina, A.V., Garbuzynskiy, S.O., Lobanov, M.Y. & Galzitskaya, O.V. Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms. *Bioinformatics* **23**, 2231-2238 (2007).
4. Nishio, Y. et al. Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of Corynebacterium efficiens. *Genome Res* **13**, 1572–1579 (2003).
5. Mizuguchi, K., Sele, M. & Cubellis, M.V. Environment specific substitution tables for thermophilic proteins. *BMC Bioinformatics* **8**, S15 (2007).
6. McDonald, J.H. Temperature Adaptation at Homologous Sites in Proteins from Nine Thermophile-Mesophile Species Pairs. *Genome Biology and Evolution* **2**, 267-276 (2010).
7. Clutterbuck, A.J. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genetics and Biology* **48**, 306-326 (2011).
8. Cherry, J.M. et al. Genetic and physical maps of Saccharomyces cerevisiae. *Nature* **387(6632 Suppl)**, 67-73 (1997).
9. Morrison, A.J. & Shen, X. Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nat Rev Mol Cell Biol* **10**, 373-384 (2009).
10. Baker, S.P. & Grant, P.A. The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. *Oncogene* **26**, 5329-5340 (2007).
11. Biswas, D., Yu, Y., Prall, M., Formosa, T. & Stillman, D.J. The yeast FACT complex has a role in transcriptional initiation. *Mol Cell Biol* **25**, 5812-5822 (2005).
12. Cantarel, B.L. et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37**, D233-D238 (2009).
13. Martensuzunova, E. & Schaap, P. An evolutionary conserved d-galacturonic acid metabolic pathway operates across filamentous fungi capable of pectin degradation. *Fungal Genetics and Biology* **45**, 1449-1457 (2008).
14. Martinez, D. et al. Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). *Nat Biotechnol* **26**, 553-560 (2008).
15. Zdobnov, E.M. & Apweiler, R. InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848 (2001).
16. Rawlings, N.D. & Morton, F.R. The MEROPS batch BLAST: A tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie* **90**, 243-259 (2008).
17. Baminger, U., Subramaniam, S.S., Renganathan, V. & Haltrich, D. Purification and characterization of cellobiose dehydrogenase from the plant pathogen Sclerotium (Athelia) rolfsii. *Appl Environ Microbiol* **67**, 1766-1774 (2001).
18. Crisan, E.V. Current concepts of thermophilism and the thermophilic fungi. *Mycologia* **65**, 1171-1198 (1973).
19. Mumma, R.O., Fergus, C.L. & Sekura, R.D. The lipids of thermophilic fungi: lipid composition comparisons between thermophilic and mesophilic fungi. *Lipids* **5**, 100-103 (1970).
20. Arthur, H. & Watson, K. Thermal adaptation in yeast: growth temperatures, membrane lipid, and cytochrome composition of psychrophilic, mesophilic, and thermophilic yeasts. *J Bacteriol* **128**, 56-68 (1976).

21. Bok, J.W. et al. Chromatin-level regulation of biosynthetic gene clusters. *Nature Chemical Biology* **5**, 462-464 (2009).
22. Chiang, Y.M. et al. Characterization of the Aspergillus nidulans Monodictyphenone Gene Cluster. *Applied and Environmental Microbiology* **76**, 2067-2074 (2010).
23. Bok, J.W. & Keller, N.P. LaeA, a regulator of secondary metabolism in Aspergillus spp. *Eukaryot Cell* **3**, 527-535 (2004).
24. Green, P. Phrap, version 0.990329. http://phrap.org (1999).
25. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res* **8**, 195-202 (1998).
26. Jaffe, D.B. Whole-Genome Sequence Assembly for Mammalian Genomes: Arachne 2. *Genome Research* **13**, 91-96 (2003).
27. Davis, R.W., Botstein, D. & Roth, J.R. Advanced Bacterial Genetics. *Cold Spring Harbor Press, Cold Spring Harbor, NY* (1980).
28. Detter, J.C. et al. Isothermal strand-displacement amplification applications for high-throughput genomics. *Genomics* **80**, 691-698 (2002).
29. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res* **9**, 868-877 (1999).
30. Gasteiger, E., Jung, E. & Bairoch, A. SWISS-PROT: connecting biomolecular knowledge via a protein database. *Curr Issues Mol Biol* **3**, 47-55 (2001).
31. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
32. Smit, A.F.A., Hubley, R. & Green, P. RepeatMasker Open - 3.0. 1996-2010. *URL:* http://www.repeatmasker.org (2010).
33. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**, 462-467 (2005).
34. Salamov, A.A. Ab initio Gene Finding in Drosophila Genomic DNA. *Genome Research* **10**, 516-522 (2000).
35. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O. & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research* **18**, 1979-1990 (2008).
36. Birney, E. Using GeneWise in the Drosophila Annotation Experiment. *Genome Research* **10**, 547-548 (2000).
37. Kent, W.J. BLAT---The BLAST-Like Alignment Tool. *Genome Research* **12**, 656-664 (2002).
38. Finn, R.D. et al. The Pfam protein families database. *Nucleic Acids Research* **38**, D211-D222 (2009).
39. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-964 (1997).
40. Price, A.L. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351-i358 (2005).
41. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* **8**, 581-599 (1997).
42. Melén, K., Krogh, A. & von Heijne, G. Reliability Measures for Membrane Protein Topology Prediction Algorithms. *Journal of Molecular Biology* **327**, 735-744 (2003).
43. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32** D277-D280 (2004).
44. Koonin, E.V. et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol* **5**, R7 (2004).

45. Consortium, G.O. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258-D261 (2004).
46. Enright, A.J., Van Dongen, S. & Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575-1584 (2002).
47. Roy, B.P. & Archibald, F. Effects of Kraft Pulp and Lignin on Trametes versicolor Carbon Metabolism. *Appl Environ Microbiol* **59**, 1855-1863 (1993).
48. Semova, N. et al. Generation, annotation, and analysis of an extensive Aspergillus niger EST collection. *BMC Microbiol* **6**, 7 (2006).
49. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**, 511-515 (2010).
50. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
51. Hardcastle, T.J. & Kelly, K.A. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* **11**, 422 (2010).
52. Wasiak, S. et al. Enthoprotin: a novel clathrin-associated protein identified through subcellular proteomics. *J Cell Biol* **158**, 855-862 (2002).
53. Gilchrist, A. et al. Quantitative proteomics analysis of the secretory pathway. *Cell* **127**, 1265-1281 (2006).
54. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**, 5383-5392 (2002).
55. Ren, Q., Chen, K. & Paulsen, I.T. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Research* **35**, D274-D279 (2007).
56. de Groot, P.W. et al. Comprehensive genomic analysis of cell wall genes in Aspergillus nidulans. *Fungal Genet Biol* **46 Suppl 1**, S72-S81 (2009).
57. Maddi, A., Bowman, S.M. & Free, S.J. Trifluoromethanesulfonic acid-based proteomic analysis of cell wall and secreted proteins of the ascomycetous fungi Neurospora crassa and Candida albicans. *Fungal Genet Biol* **46**, 768-781 (2009).
58. Eisenhaber, F. Prediction of lipid posttranslational modifications and localization signals from protein sequences: Big-Pi , NMT and PTS1. *Nucleic Acids Research* **31**, 3631-3634 (2003).
59. Tatusov, R.L. et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41 (2003).
60. Marchler-Bauer, A. et al. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35**, D237-240 (2007).
61. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066 (2002).
62. Le, S.Q., Lartillot, N. & Gascuel, O. Phylogenetic mixture models for proteins. *Philos Trans R Soc Lond B Biol Sci* **363**, 3965-3976 (2008).
63. Chenna, R. et al. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**, 3497-3500 (2003).
64. Willis, L., G,, Winston, M.L. & Honda, B.M. Phylogenetic relationships in the honeybee (genus Apis) as determined by the sequence of the cytochrome oxidase II region of mitochondrial DNA. *Mol Phylogenet Evol* **1**, 169-178 (1989).

SUPPLEMENTARY TABLES

Organism pair	# of Orthologs (avg % aa identity)	# of Syntenic blocks	# of Genes in syntenic blocks	# of Genes in largest block
Mycth-Thite	7017 (72%)	221	5950	485
Mycth -Chagl	6996 (73%)	284	6270	358
Thite -Chagl	7203 (70%)	313	5721	396

Supplementary Table 1. Orthology and synteny between Chaetomiaceae. Chagl, *Chaetomium globosum*; Mycth, *Myceliophthora thermophila*; Thite, *Thielavia terrestris*.

Supplementary Information

Pfam ID	Mycth	Thite	Chag	Neucr	Necha	Trire	Anig	Pfam Desciription
PF07690	152	202	159	152	610	222	432	MFS_1
PF00172	110	132	106	99	477	192	243	Zn_clus
PF00400	95	97	100	93	107	105	105	WD40
PF00106	70	118	104	72	291	135	189	adh_short
PF00069	81	89	70	86	98	86	90	Pkinase
PF00271	71	73	85	75	87	75	73	Helicase_C
PF04082	57	72	64	61	337	113	197	Fungal_trans
PF00096	49	63	51	53	86	47	54	zf-C2H2
PF00076	55	56	56	64	54	54	59	RRM_1
PF00067	47	56	69	39	142	69	132	p450
PF00083	48	53	53	46	215	69	125	Sugar_tr
PF08659	37	60	52	37	107	64	94	KR
PF08241	43	51	44	51	119	46	60	Methyltransf_11
PF08242	39	52	43	48	127	48	67	Methyltransf_12
PF00004	40	41	40	40	55	44	42	AAA
PF00646	37	41	42	37	99	46	54	F-box
PF00270	38	40	44	43	51	42	40	DEAD
PF00005	37	35	41	35	76	46	64	ABC_tran
PF00097	33	38	29	33	38	32	32	zf-C3HC4
PF02985	35	36	34	39	34	38	44	HEAT
PF00023	30	38	94	49	124	58	72	Ank
PF01266	31	37	33	29	120	30	70	DAO
PF08240	29	38	50	31	141	52	114	ADH_N
PF00153	33	32	35	36	47	35	38	Mito_carr
PF00071	31	32	26	36	37	32	27	Ras
PF00107	28	35	38	28	133	45	98	ADH_zinc_N
PF08477	30	31	23	36	35	31	27	Miro
PF01370	29	32	33	24	101	52	78	Epimerase
PF01494	23	38	32	22	89	30	70	FAD_binding_3
PF00702	29	29	28	29	50	31	39	Hydrolase

PF00501	28	26	39	25	50	37	75	AMP-binding
PF07719	27	26	32	25	26	28	26	TPR_2
PF00515	25	25	27	24	24	25	29	TPR_1
PF07993	24	26	23	13	51	31	55	NAD_binding_4
PF01565	19	30	42	22	71	27	48	FAD_binding_4
PF00226	24	25	25	28	28	26	27	DnaJ
PF00149	24	25	21	26	31	27	28	Metallophos
PF00583	23	26	18	23	63	40	28	Acetyltransf_1
PF00018	24	25	25	24	23	24	25	SH3_1
PF00176	22	25	24	25	33	25	26	SNF2_N
PF00173	21	21	18	20	35	21	29	Cyt-b5
PF00734	17	23	33	21	12	14	8	CBM_1
PF00026	16	23	18	17	18	17	11	Asp
PF00561	19	20	23	19	61	33	36	Abhydrolase_1
PF00664	20	18	19	17	36	22	31	ABC_membrane
PF00179	19	19	17	20	19	20	21	UQ_con
PF01073	17	21	17	16	53	23	41	3Beta_HSD
PF07653	18	20	18	19	17	18	19	SH3_2
PF00324	18	19	17	20	87	33	64	AA_permease
PF07992	17	19	20	12	30	18	24	Pyr_redox_2

Supplementary Table 2. Top 50 PFAM domains in Chaetomiaceae compared to select Ascomycetes. Anig, *Aspergillus niger*; Chagl, *Chaetomium globosum*; Mycth, *Myceliophthora thermophila*; Necha, *Nectria haematococca*; Neucr, *Neurospora crassa*; Thite, *Thielavia terrestris;* Trire, *Trichoderma reesei*.

Supplementary Information

PfamID	Mycth	Thite	Chag	Neucr	Necha	Trire	Anig	Pfam Desciription
PF03443	14	12	17	11	9	2	5	Glyco_hydro_61
PF09362	14	14	20	9	4	3	2	DUF1996
PF00457	7	5	9	2	3	3	3	Glyco_hydro_11
PF00733	3	3	3	2	2	2	2	Asn_synthase
PF00352	2	2	2	1	1	1	1	TBP
PF02896	1	1	1	0	0	0	0	PEP-utilizers_C
PF00112	1	1	1	0	0	0	0	Peptidase_C1
PF08982	1	1	1	0	0	0	0	DUF1857

Supplementary Table 3. PFAM domains expanded in Chaetomiaceae. Anig, *Aspergillus niger*; Chagl, *Chaetomium globosum*; Mycth, *Myceliophthora thermophila*; Necha, *Nectria haematococca*; Neucr, *Neurospora crassa*; Thite, *Thielavia terrestris;* Trire, *Trichoderma reesei*.

Supplementary Information

Organism	Number of genes encoding					
	GH	**GT**	**PL**	**CE**	**CBM**	**EXPN**
Trichoderma reesei QM6a	200	103	6	17	49	4
Nectria haematococca mpVI	339	132	33	44	65	6
Gibberella zeae PH-1	248	110	21	44	67	4
Myceliophthora thermophila	205	83	8	27	52	2
Thielavia terrestris	214	89	3	26	66	2
Chaetomium globosum CBS 148.51	266	96	15	42	85	1
Neurospora crassa OR74A	174	78	4	22	42	1
Neurospora tetrasperma FGSC 2508	180	90	4	24	56	2
Podospora anserina S mat+	230	89	7	41	97	2
Magnaporthe grisea 70-15	232	94	5	47	65	1

Supplementary Table 4. Comparison of the number of predicted CAZymes for *Myceliophthora thermophila* and *Thielavia terrestris* with eight mesophilic, filamentous fungi: GH, glycoside hydrolase; GT, glycosyl transferase; PL, polysaccharide lyase; CE, carbohydrate esterase; CBM, carbohydrate-binding module; and EXPN, expansin.

Supplementary Information

Organism	Genome GC, %	Coding GC, %	3rd position GC, %
Thielavia terrestris	54	63 (+9)	80 (+17)
Myceliophthora thermophila	51	62 (+9)	77 (+15)
Chaetomium globosum	55	60 (+5)	72 (+12)
Neurospora crassa	50	56 (+6)	66 (+10)
Trichoderma reesei	52	59 (+7)	72 (+13)
Aspergilus niger	50	54 (+4)	61 (+7)

Supplementary Table 10. GC content for genomes, coding regions and 3rd positions of codons (differences between the two columns are shown in parenthesis).

Supplementary Information

Organism	**"IVYWREL" motif**
Myceliophthora thermophila	36.3
Thielavia terrestris	36.1
Chaetomium globosum	35.9
Neurospora crassa	35.4
Nectria haematococca	36.8
Trichoderma reesei	36.4
Trichoderma atroviride	36.4
Trichoderma virens	36.7
Aspergillus niger	37.2
Aspergillus fumigatus	37.0
Mycosphaerella graminicola	36.1
Stagonospora nodorum	35.9
Candida albicans	36.4
Pichia stipidis	38.1
Saccharomyces cerevisiae	36.9
Schizosaccharomces pombe	38.0

Supplementary Table 11. Average content of amino acid "IVYWREL" motifs in the proteomes of 16 Ascomycete species.

Supplementary Information

Organism	A	V	L	I	C	P	M	Y	F	H	W	D	N	E	Q	S	T	R	K	G
M. thermophila	9.8	6.2	8.7	4.2	1.1	6.7	2.0	2.6	3.4	2.3	1.4	5.7	3.2	6.3	3.9	7.8	5.7	6.9	4.5	7.5
T. terrestris	10.5	6.3	8.9	4.0	1.2	6.8	2.0	2.5	3.4	2.3	1.4	5.6	3.1	6.1	3.9	7.7	5.6	6.9	4.2	7.5
C. globosum	9.5	6.2	8.7	4.2	1.2	6.7	2.1	2.5	3.4	2.4	1.5	5.6	3.4	6.1	4.0	7.6	6.1	6.8	4.4	7.6
N. crassa	8.7	6.0	8.4	4.4	1.1	6.5	2.2	2.6	3.4	2.4	1.4	5.6	3.7	6.5	4.3	8.3	6.1	6.1	5.1	7.2
N. hematococca	8.5	6.2	9.0	5.0	1.4	5.9	2.2	2.8	3.9	2.4	1.6	5.8	3.6	6.2	3.9	7.8	5.9	6.0	4.8	6.9
T. reesei	9.2	6.1	9.0	4.8	1.2	6.0	2.2	2.7	3.7	2.4	1.5	5.8	3.5	6.2	4.0	8.1	5.7	6.2	4.8	6.9
T. atrovide	8.8	6.0	9.0	5.2	1.3	5.7	2.2	2.8	3.8	2.4	1.5	5.7	3.8	6.1	4.0	8.2	5.7	5.9	4.9	6.8
T. virens	8.7	6.1	9.1	5.3	1.3	5.7	2.2	2.8	3.8	2.4	1.5	5.7	3.8	6.1	4.0	8.1	5.8	5.8	4.9	6.8
A. niger	8.5	6.3	9.2	5.0	1.3	5.9	2.2	3.0	3.8	2.4	1.5	5.6	3.6	6.1	4.0	8.2	6.0	6.1	4.4	6.8
A. fumigatus	8.6	6.2	9.2	4.9	1.3	6.0	2.1	2.8	3.7	2.4	1.5	5.5	3.6	6.1	4.1	8.4	5.9	6.4	4.6	6.7
M. gramenicula	9.1	6.1	8.8	4.7	1.3	5.8	2.2	2.7	3.6	2.4	1.5	5.8	3.6	6.3	4.0	7.9	6.1	6.2	4.9	7.0
S. nodorum	8.9	6.1	8.6	4.8	1.3	5.9	2.3	2.8	3.7	2.5	1.5	5.7	3.7	6.2	4.1	7.9	6.1	6.0	5.1	6.7
C. albicans	5.0	5.4	9.2	7.1	1.1	4.5	1.8	3.5	4.4	2.1	1.0	5.9	6.7	6.4	4.5	9.0	6.1	3.8	7.3	5.0
P. stipidis	5.9	6.1	9.9	6.6	1.0	4.4	1.8	3.6	4.6	2.1	1.0	5.9	5.7	6.6	3.8	8.7	5.7	4.2	6.8	5.2
S. cerevisiae	5.5	5.5	9.5	6.5	1.3	4.4	2.1	3.4	4.4	2.2	1.0	5.8	6.2	6.5	3.9	9.0	5.9	4.4	7.3	
S. pombe	6.2	6.0	9.9	6.2	1.5	4.7	2.1	3.4	4.6	2.3	1.1	5.3	5.2	6.5	3.8	9.4	5.5	4.9	6.4	4.9

Supplementary Table 12. Amino acid composition of 16 Ascomycete proteomes. See **Supplementary Table 6** for genus names.

Amino acid Substitution	Organism pair					
	Mycth-Chagl	**Mycth-Thite**	**Thite - Chagl**	**Trire - Triat**	**Trire - Trivi**	**Triat – Trivi**
A -> S	4580-3901*	5169-3519*	3654-4940*	6139-4402*	6295-4164*	5483-5558
R -> K	4619-3165*	4755-2674*	3770-4644*	4865-3348*	5076-2940*	4148-4118

Supplementary Table 13. Amino acid substitutional asymmetry in pairs of orthologues between the two thermophilic fungi and eight mesophilic fungi: *M. thermophila* (Mycth), *T. terrestris* (Thite), *C. globosum* (Chagl), *T. reseei* (Trire), *T. atroviride* (Triat), *T. virens* (Trivi). For each pair of fungi, the number of substitutions is shown in reverse and direct order. Statistically significant deviation from 1:1 ratio, $p <$ 1e-05 (Bonferroni-corrected chi-square) is indicated by an asterisk (*).

Supplementary Information

Organism	**“ERK” - “DNQTSHA” content**
Myceliophthora thermophila	20.1 ± 7.1
Thielavia terrestris	20.8 ± 7.2
Chaetomium globosum	20.7 ± 6.5
Neurospora crassa	20.8 ± 7.3
Nectria haematococca	20.5+/-5.7
Trichoderma reesei	20.8 ± 7.1
Trichoderma atroviride	21.0 ± 6.8
Trichoderma virens	20.8 ± 6.8
Aspergillus niger	20.8 ± 6.8
Aspergillus fumigatus	20.6 ± 6.4
Mycosphaerella graminicola	20.6 ± 6.7
Stagonospora nodorum	20.8 ± 6.4
Candida albicans	20.5 ± 7.3
Pichia stipidis	19.1 ± 6.8
Saccharomyces cerevisiae	18.7 ± 7.0
Schizosaccharomces pombe	18.7 ± 6.9

Supplementary Table 14. Average "ERK" - "DNQTSHA" content in the proteomes of 16 Ascomycetes.

Supplementary Information

Organism	Lignin peroxidase	Manganese Peroxidase	Laccase	Copper Radical Oxidase	Aryl-alcohol Oxidase	Cellobiose Dehydrogenase
T. terrestris	0	0	7	1	0	1
M. thermophila	0	0	7	1		1
N. crassa	0	0	7	1	0	0
A. niger	0	0	9	0	0	0
T. reesei	0	0	3	0	0	0

Supplementary Table 15. A comparison of predicted extracellular oxidoreductases encoded in the genomes of five filamentous Ascomycete fungi.

Supplementary Information

C. globosum	***M. thermophila***	***T. terrestris***	**Annotation**
Chagl_10760	Mycth_2296038	2113716	Cu-Zn, secreted
Chagl_16224	Mycth_2297816	Thite_2142959	Cu-Zn
Chagl_11148	***none***	Thite_2122167	Cu-Zn, Secreted
Chagl_17245	Mycth_2301830	Thite_2108025	Fe/Mn
Chagl_19922	Mycth_2308056	Thite_2117870	Fe/Mn, Mitochondrial
Chagl_10814	Mycth_2295915	Thite_2143700	Fe/Mn, Secreted

Supplementary Table 16. Superoxide dismutases encoded in the genomes of three Chaetomiaceae. The closest orthologues, as determined by ClustalW alignments are located on the same row of the table.

C. globosum	***M. thermophila***	***T. terrestris***	**Annotation**
Chagl_13388	Mycth_2299739	Thite_42931	Catalase
Chagl_12800	Mycth_80916	***none***	Catalase
Chagl_13820	***none***	***none***	Catalase
Chagl_15098	Mycth_2303088	Thite_2120501	Catalase-Peroxidase
none	***none***	Thite_2088638	Catalase-Peroxidase

Supplementary Table 17. Catalases and catalase-peroxidases. The closest orthologues, as determined by ClustalW alignments are located on the same row of the table.

Supplementary Information

Transporter family	Substrate	*M. thermophila*	*T. terrestris*
AAAP	amino acid	4	6
ABC	multidrug	19	44
ACR3	arsenite	0	1
AE	chloride/bicarbonate anion exchange	0	1
AEC		0	2
Annexin		0	1
APC	amino acid	5	18
ArsAB	arsenite (ArsA)	0	1
ArsB	arsenite (ArsB)	1	0
ATP-E	ATP release	0	1
CaCA	proton:calcium ion antiporter	3	8
CCC	potassium ion:chloride ion symporter	0	1
CDF	cation efflux	3	7
CHR	chromate ion	0	1
ClC	chloride ion channel	2	2
CNT	sodium ion:nucleoside symporter	0	1
CPA1	sodium ion:proton antiporter	2	1
CPA2	potassium/sodium ion:proton antiporter	2	1
CTL	sodium ion:choline symporter?	0	1
Ctr	copper ion uptake	1	1
DASS	sodium ion:dicarboxylate/sulfate symporter	0	1
DMT	UDP-galactose:UMP antiporter	4	13
ENT	nucleoside	1	0
F-ATPase	protons	6	24
FeT		0	2
FNT	formate/nitrite	0	1
GPH	sucrose	1	2
HCC	hemolysin C (HlyC) homolog	0	1
IISP	signal recognition particle receptor beta subunit (SR-beta)	2	12
ILT	iron ion	0	1
KUP	potassium ion uptake	0	1
LCT		0	2
LTE	lipid export	2	6
MagT1	magnesium ion	1	0
MC		10	22
MFS	multidrug efflux	86	221
Mid1		0	1
MIP	glycerol uptake	1	1
MIT	magnesium/cobalt ion	3	3

MOP	nuclear division RFT1 homolog (OLF subfamily)	2	4
MPT	inner membrane translocase (import) Tim10	8	11
MscL	large-conductance mechanosensitive ion channel	1	0
MscS	small-conductance mechanosensitive ion channel	0	1
MTC	tricarboxylates	1	0
NCS1	cytosine/purines/uracil/thiamine/allantoin	1	2
NCS2	xanthine/uracil	1	3
NiCoT	nickel ion	1	1
NIPA	magnesium ion uptake	1	3
NSCC2	preprotein translocase Sec62	0	1
OPT	oligopeptide	3	9
Oxa1	60 KD inner membrane protein OxaA homolog	0	1
P-ATPase	potassium/sodium ion	7	17
Pho1	phosphate	0	1
PiT	phosphate	1	1
POT	proton:dipeptide/tripeptide symporter	0	1
PPI	peroxisomal protein import (Pex14)	0	3
PPI2	peroxisomal protein import (Pex3)	1	3
RND	Niemann-Pick C disease protein homolog	1	3
SSS	sodium ion:proline symporter	2	3
SulP	sulfate	2	2
TDT	tellurite	0	1
Trk	potassium ion uptake	1	1
TRP-CC	transient receptor potential calcium ion channel	1	1
VIC	potassium ion channel	0	2
VIT	vacuolar iron uptake transporter homolog	0	3
YaaH	YaaH family acetate transporter?	0	2
ZIP	zinc ion	7	3

Supplementary Table 18. Number of membrane transporter proteins per organism by transporter family and substrate.

SGD Name	EC Number	Enzyme Name	*M. thermophila* Protein ID	*T. terrestris* Protein ID
Erg10	2.3.1.9	acetyl-CoA C-acetyltransferase	105847	2156279
			2298614	2115011
Erg13	2.3.3.10	HMG-CoA synthase	111934	127590
	1.1.1.34	3-Hydroxy-3-methylglutaryl-CoA reductase	114008	135005
Erg12	2.7.1.36	mevalonate kinase like protein	66022	57831
Erg12	2.7.1.36	mevalonate kinase like protein	71219	-
Erg8	2.7.4.2	phosphomevalonate kinase-like	38472	35750
	4.1.1.33	diphosphomevalonate decarboxylase	83048	136252
Erg20	2.5.1.10	farnesyl pyrophosphate synthetase	115096	133921
Erg9	2.5.1.21	farnesyl-diphosphate farnesyl transferase	112159	76823
Erg1	1.14.99.7	squalene epoxidase	2299088	2115532
				165536
Erg7	5.4.99.7	oxidosqualene:lanosterol cyclase	2301508	129749
Erg11	1.14.13.70	eburicol 14a-demethylase	110187	135170
Erg24	1.3.1.70	delta(14)-sterol reductase like protein	50083	155704
Erg25	1.14.13.72	methylsterol monooxygenase	109638	125790
Erg26	1.1.1.170	sterol-4-alpha-carboxylate 3-dehydrogenase (decarboxylating) like protein	122939	129164
Erg27	1.1.1.270	3-keto steroid reductase-like protein	69410	164350
Erg6	2.1.1.41	Sterol 24-C-methyltransferase	110670	68084
Erg2	5.-.-.-	C-8 sterol isomerase	64893	70538
Erg3	1.3.3.-	C-5 sterol desaturase-like protein	112329	64954
Erg5	1.14.14.-	C-22 sterol desaturase	73679	128412
Erg4/ Erg24	1.3.1.7	delta(24(24(1)))-sterol reductase like protein	111214	112329

Supplementary Table 19. Enzymes of the ergosterol biosynthesis pathway. The order in this table is the same order as in the biosynthetic pathway. Three additional acetyl-CoA C-acetyltransferases were present in the automated annotation of each genome, but the two shown are the most similar to the *S. cerevisiae* Erg10p.

Supplementary Information

EC number	Enzyme Name	*M. thermophila* Protein ID	*T. terrestris* Protein ID
2.3.1.16	fatty acid elongase	79157	69856
2.3.1.16	fatty acid elongase-like protein	67390	59436
1.14.19.6	delta-12 acyl-CoA desaturase	88639	76756
	bifunctional D12/D15 fatty acid desaturase	50837	-
1.14.19.1	delta-9 acyl-CoA desaturase	109620	125777
1.14.19.3	delta-6 acyl-CoA desaturase	2309213	2116695
1.14.19.-	sphingolipid delta-4 desaturase like protein	2314931	2109119

Supplementary Table 20. Fatty acid elongases and desaturases.

Supplementary Information

	A. nidulans[1]	*M. thermophila* Protein ID	*T. terrestris* Protein ID	Notes
Chitin synthase				
	7032 (*chsA*)	2309764	2147675	
	2523 (*chsB*), 4367	79885	2106408	
	799, 4566 (*chsC*)	2302583	2108832	
	1555 (*chsD/E*)	2296388	2113371	
	6318 (*csmA*)	2299402	2115877	
	6317 (*csmB*)	112916	2115879	
	1046	2299267	66300	
Chitinase	8241 (*chiA*), 11059	113450	2111816, 35493	GPI anchor?
	11063	2067358	52693	
		2300996	2041478	
	221	54618	2116882	
	299	98773	2121296	
	517, 541	--	--	
	549, 9390 (*chiC*), 8481	--	2054382	
	4871 (*chiB*), 5454	50608	35183, 2121983	
		2308241	2117724	
	509	94536	46100, 154425	
	7613	--	2057765	
		95226	2128646	
	7886	2042719	--	
1,3-β-glucan synthase	3729 (*fksA*)	140260	132141	
endo-1,3-β-glucanase	472 (*engA*)	2306588	2109946	
exo-1,3-β-glucanase	1332 (*exgA*), 3777 (*exgB*), 4052 (*exgC*)	2131308	--	
	7533 (*exgD*)	82558	2109869	
	8947 (*exgE*)	--	2122186, 2090256	
	550, 8480, 4825	102522	2040561	
		2299780	2109436	
		98521	120170	
		--	2122792	
	779	2305407	2118359	
	7869	--	--	
1,3-β-transglucosidase	7950 (*eglC*)	2315007	2171105	GPI anchor
		2124399	--	
		108391	--	
	1551 (*btgE*)	2296382	2113377	
	10150 (*btgA*), 3727 (*btgD*)	--	--	
	4700 (*btgC*)	2297176	68116	
	7657 (*gelA*)	60294	2114708	

		66349	2073214	
	558 (*gelB*)	71748	2124545	
	3730 (*gelC*), 11152 (*gelD*), 7511 (*gelE*)	106819	2169375	
1,3-β-transglycosidase	3914 (*crhA*)	--	--	
	4515 (*crhB*)	2079097	2069256	
	933 (*crhC*)	61486	2111204	
	3053 (*crhD*), 6948 (*crhE*)	62175	2121558	
		2315557	2169619	

Supplementary Table 21. Cell wall proteins documented in *A. nidulans*, and their best matches in *M. thermophila* and *T. terrestris.*

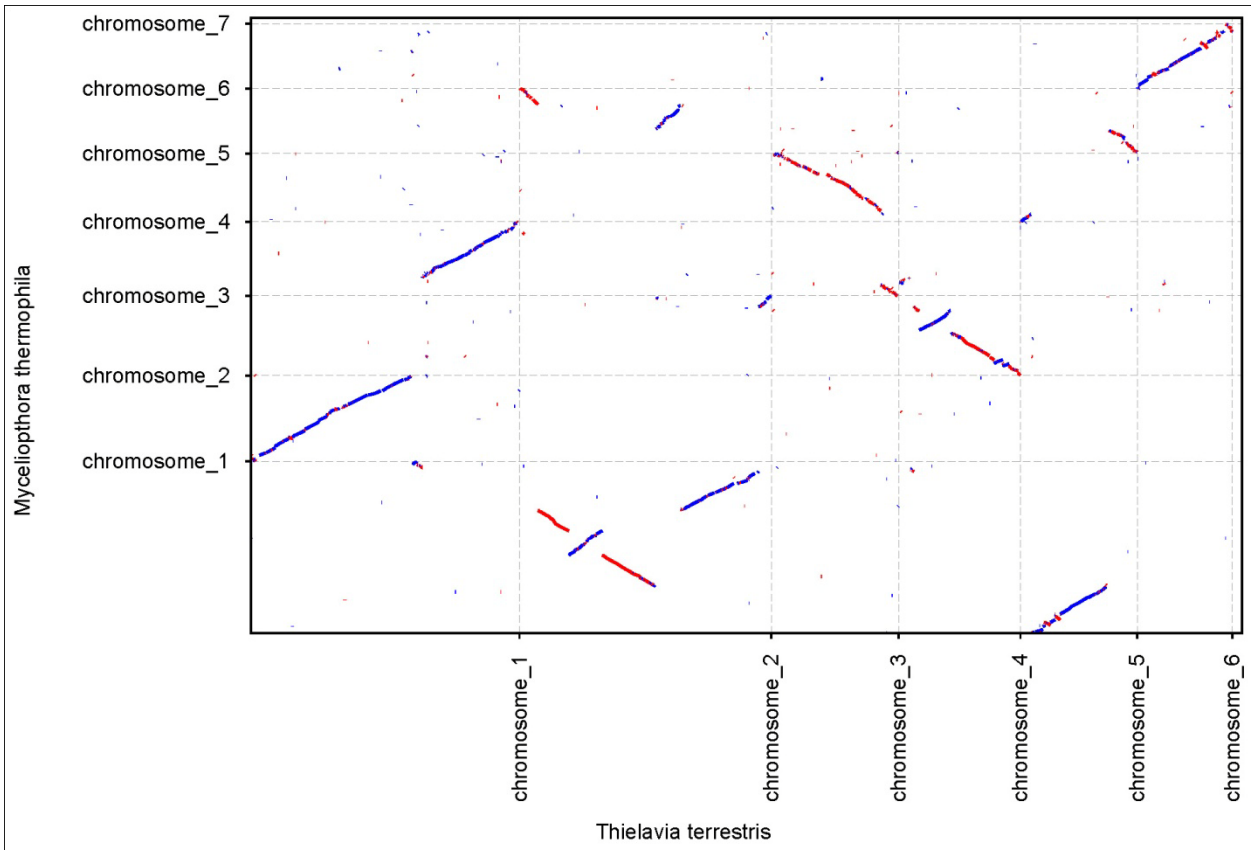
Supplementary Information

Library Type	*M. thermophila*			*T. terrestris*		
	Average Insert Size, bp	**Read Number**	**Assembled Sequence Coverage (X)**	**Average Insert Size, bp**	**Read Number**	**Assembled Sequence Coverage (X)**
3kb	3,265	202,837	3.37	2,530	339,194	5.57
8kb (1)	6,692	211,776	3.39	8,692	271,871	3.90
Fosmid	36,072	54,528	0.90	33,260	88,508	0.67
Total		469,141	7.66		470,079	10.15

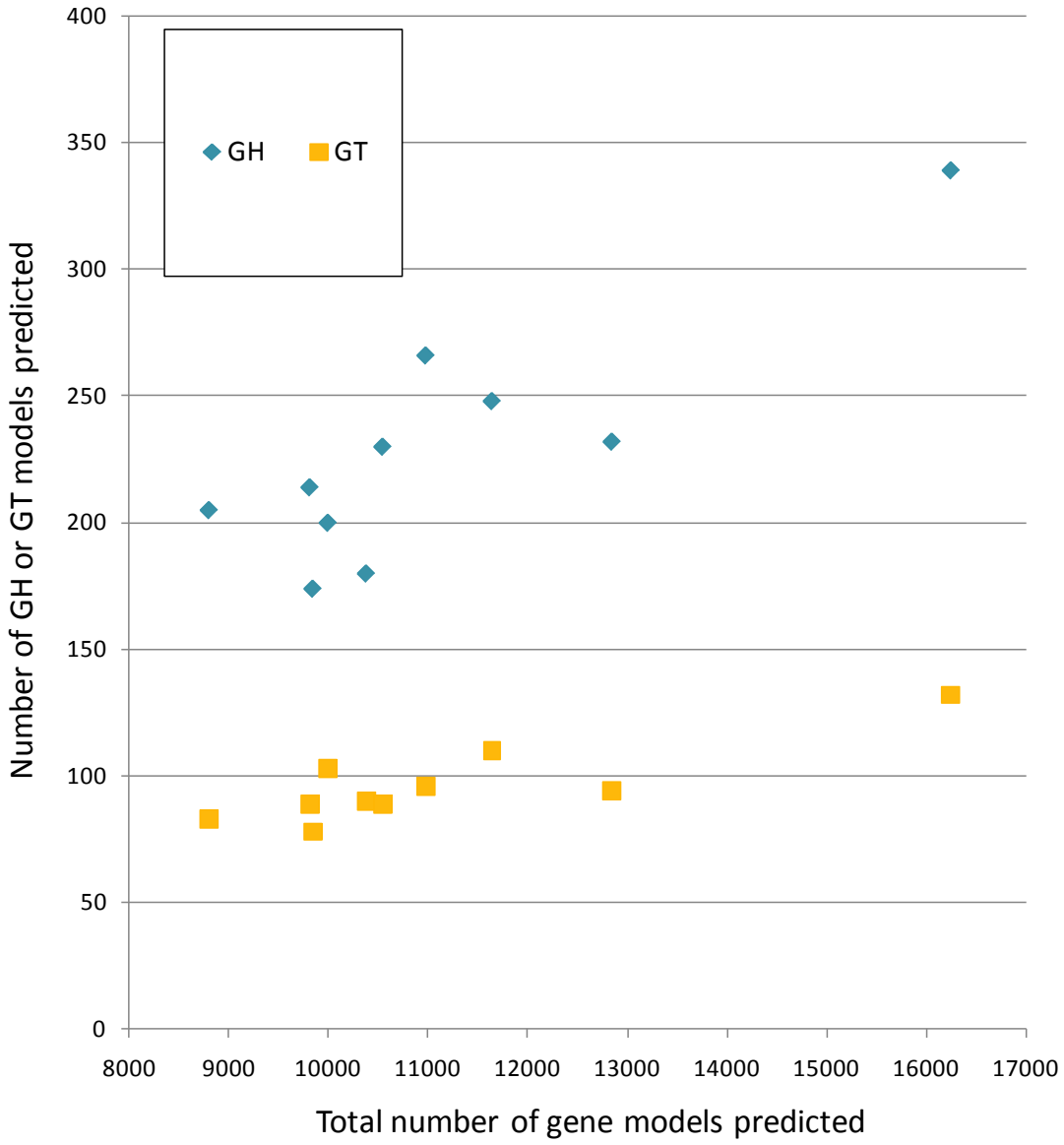
Supplementary Table 22. Genomic libraries included in the genome assemblies of *M. thermophila* and *T. terrestris*. Also shown are their respective assembled sequence coverage levels in the whole genome shotgun assembly.

Supplementary Tables 5-9 and 23-25 can be found in the accompanying Excel file.

SUPPLEMENTARY FIGURES



Supplementary Figure 1. Dotplots indicating extended regions of nucleotide-level synteny between *M. thermophila* and *T. terrestris* according to the VISTA DNA alignments. Same-strand alignments are shown in blue, opposite-strand reads are in red.

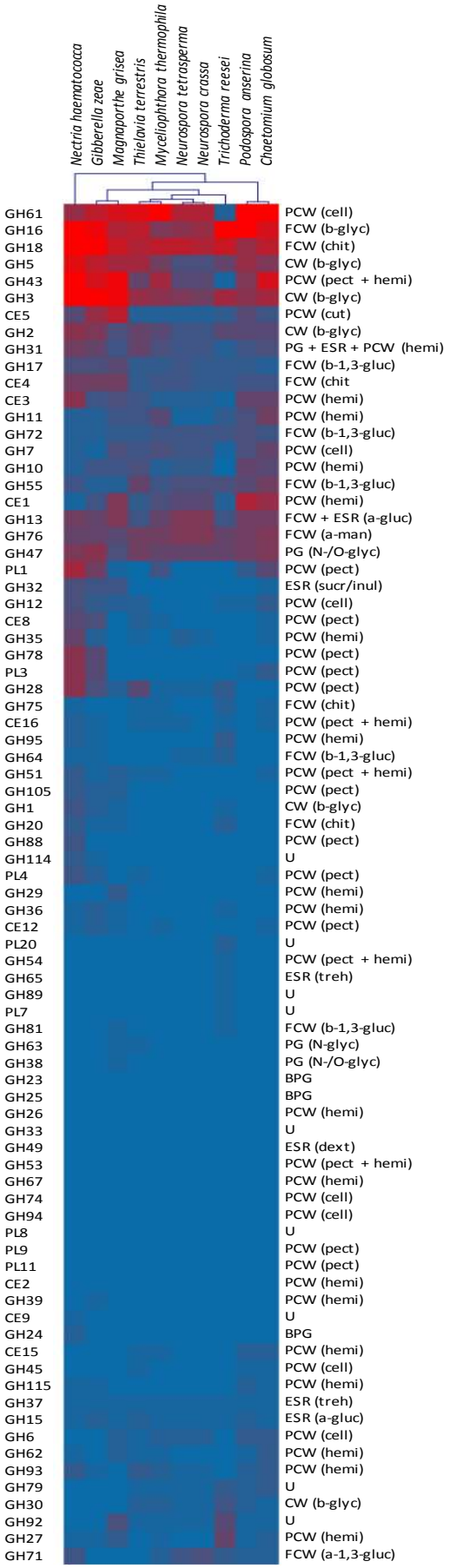


Supplementary Figure 2. Number of glycoside hydrolases (GH) and glycosyltransferases (GT) genes in ten selected fungi [*Trichoderma reesei (Hypocrea jecorina)* QM6a; *Nectria haematococca* mpVI; *Gibberella zeae* PH-1; *Myceliophthora thermophila*; *Thielavia terrestris*; *Chaetomium globosum* CBS 148.51; *Neurospora crassa* OR74A; *Neurospora tetrasperma* FGSC 2508; *Podospora anserina* S mat+; *Magnaporthe grisea* 70-15].

Supplementary Information

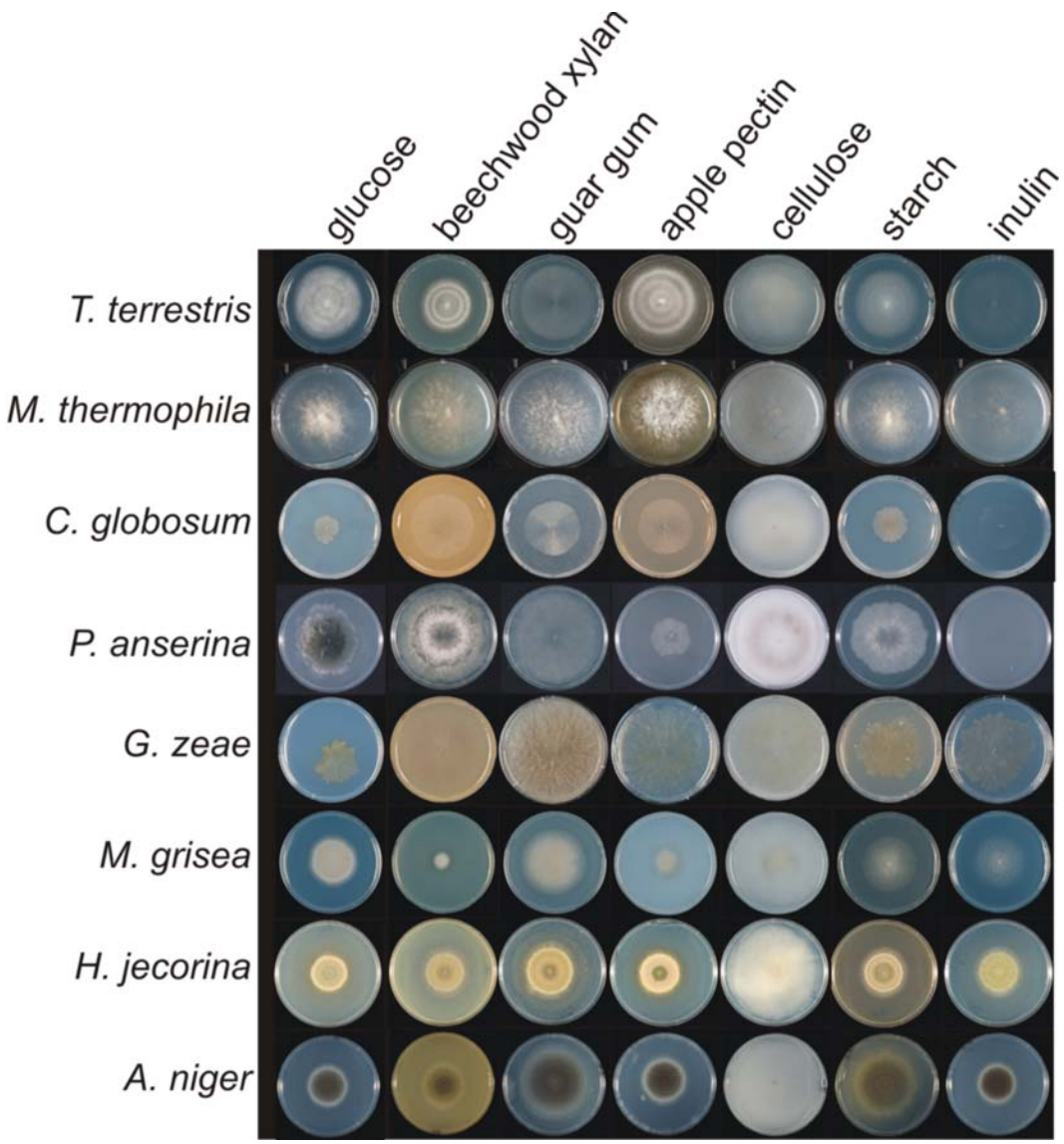
Supplementary Figure 3. Double clustering of the carbohydrate-cleaving families from representative fungal genomes to illustrate abundance of CAZymes in each family. Abundance of the different enzymes within a family is represented by a colour scale from 0 (blue) to > 20 occurrences (red) per species.

Top tree: the fungi named are *Trichoderma reesei (Hypocrea jecorina)* QM6a; *Nectria haematococca* mpVI; *Gibberella zeae* PH-1; *Myceliophthora thermophila*; *Thielavia terrestris*; *Chaetomium globosum* CBS 148.51; *Neurospora crassa* OR74A; *Neurospora tetrasperma* FGSC 2508; *Podospora anserina* S mat+; *Magnaporthe grisea* 70-15. Left: The enzyme families are represented by their class (GH, glycoside hydrolase; PL, polysaccharide lyase; CE, carbohydrate esterase) and family number according to the carbohydrate-active enzyme database [12]. Right side: known substrate of CAZy families (most common forms in brackets): BPG, bacterial peptidoglycan; BEPS, bacterial exopolysaccharides; CW, cell wall; ESR, energy storage and recovery; FCW, fungal cell wall; PCW, plant cell wall; PG, protein glycosylation; U, undetermined; α-gluc, α-glucans (including starch/glycogen); β-glyc, β-glycans;b-1,3-gluc, β-1,3-glucan; cell, cellulose; chit, chitin/chitosan; dext, dextran; hemi, hemicelluloses; inul, inulin; N-glyc, N-glycans; N-/O-glyc, N- / O-glycans; pect, pectin; sucr, sucrose; and tre, trehalose.



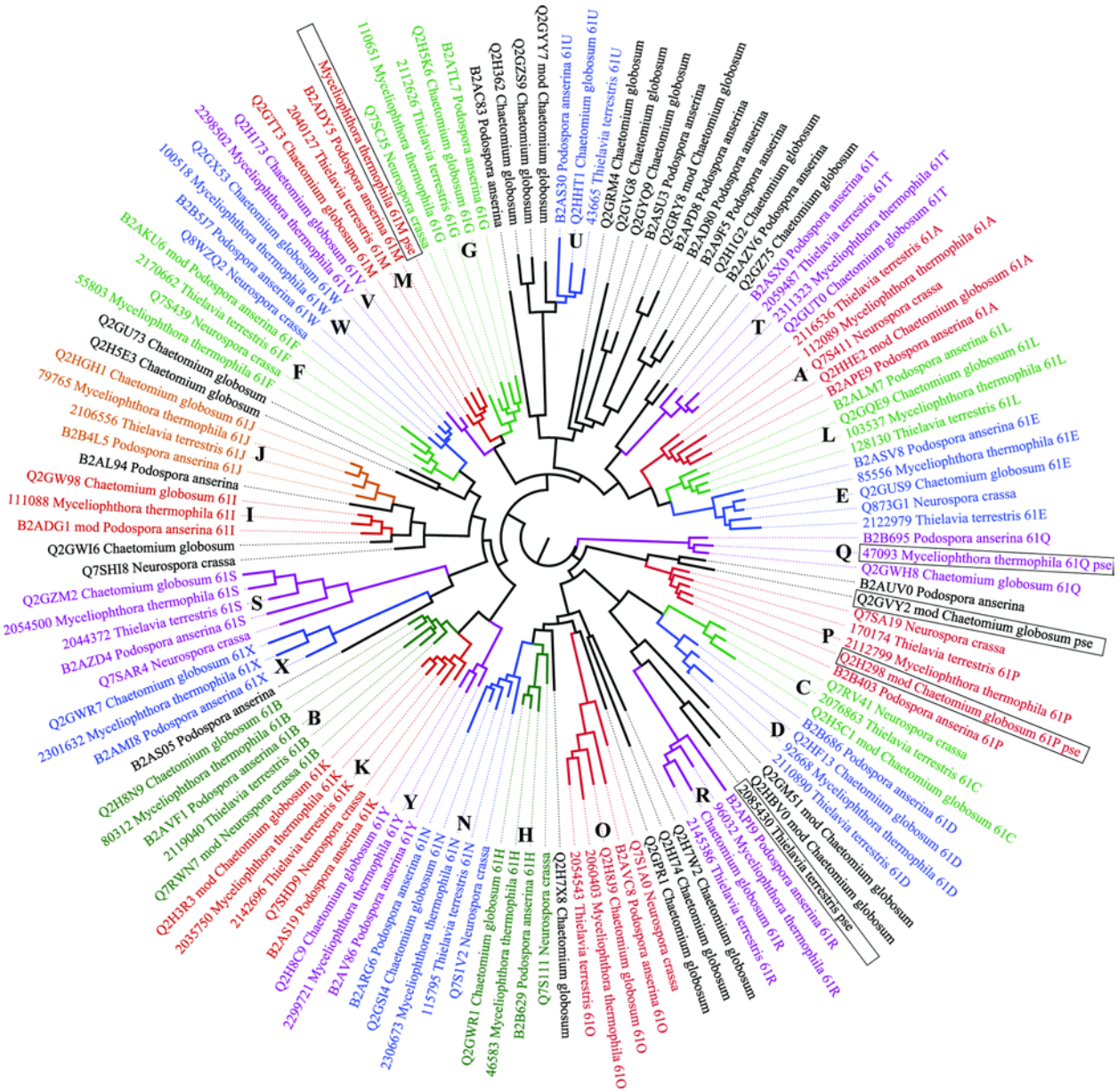
Supplementary Information

Supplementary Figure 4. Growth of *T. terrestris* and *M. thermophila* on a number of different carbon sources at 34°C and 45°C. The figure shows that growth of both species is much better at 45°C than at 34°C, demonstrating that they are true thermophiles. The *T. terrestris* genome contains mainly pectin hydrolases (most active at acidic pH), while the *M. thermphila* genome contains mainly pectin lyases (mainly active at alkaline pH). This is reflected in their growth as *T. terrestris* grows best on pectin at low pH and *M. thermophila* at high pH**.**

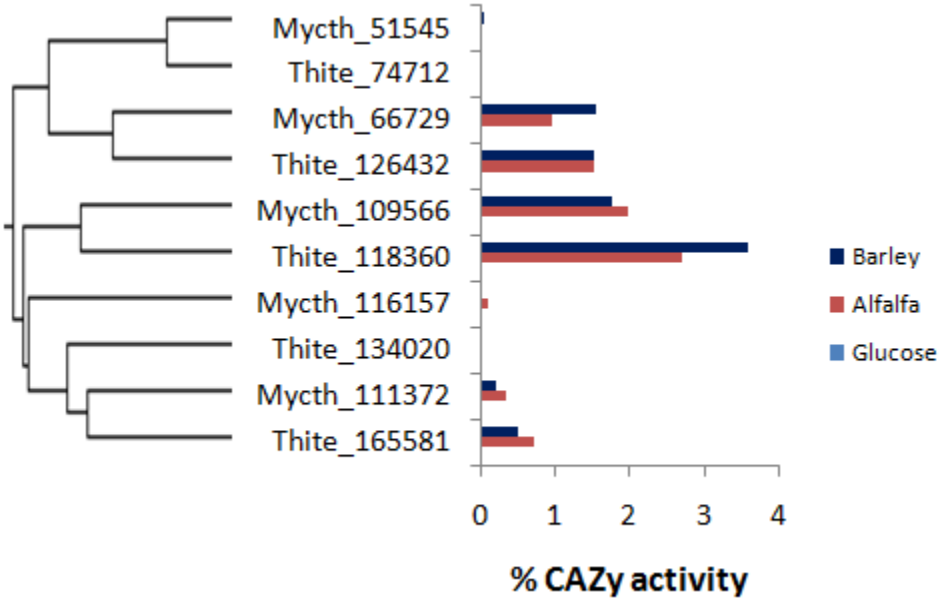


Supplementary Figure 5. Comparison of the growth of filamentous fungi on different substrates. The growth of *T. terrestris* and *M. thermophila* on polymeric carbon sources were compared to that of *Chaetomium globosum*, *Podospora anserine*, *Gibberella zeae*, *Magnaporthe grisea*, *Hypocrea jecorina* and *Aspergillus niger*. As different fungi have different growth rates, growth on glucose was used as an internal standard to enable the comparison of relative growth between the fungi. Guar gum is a galactomannan, similar in structure to softwood cell wall galactomannans.

Supplementary Information

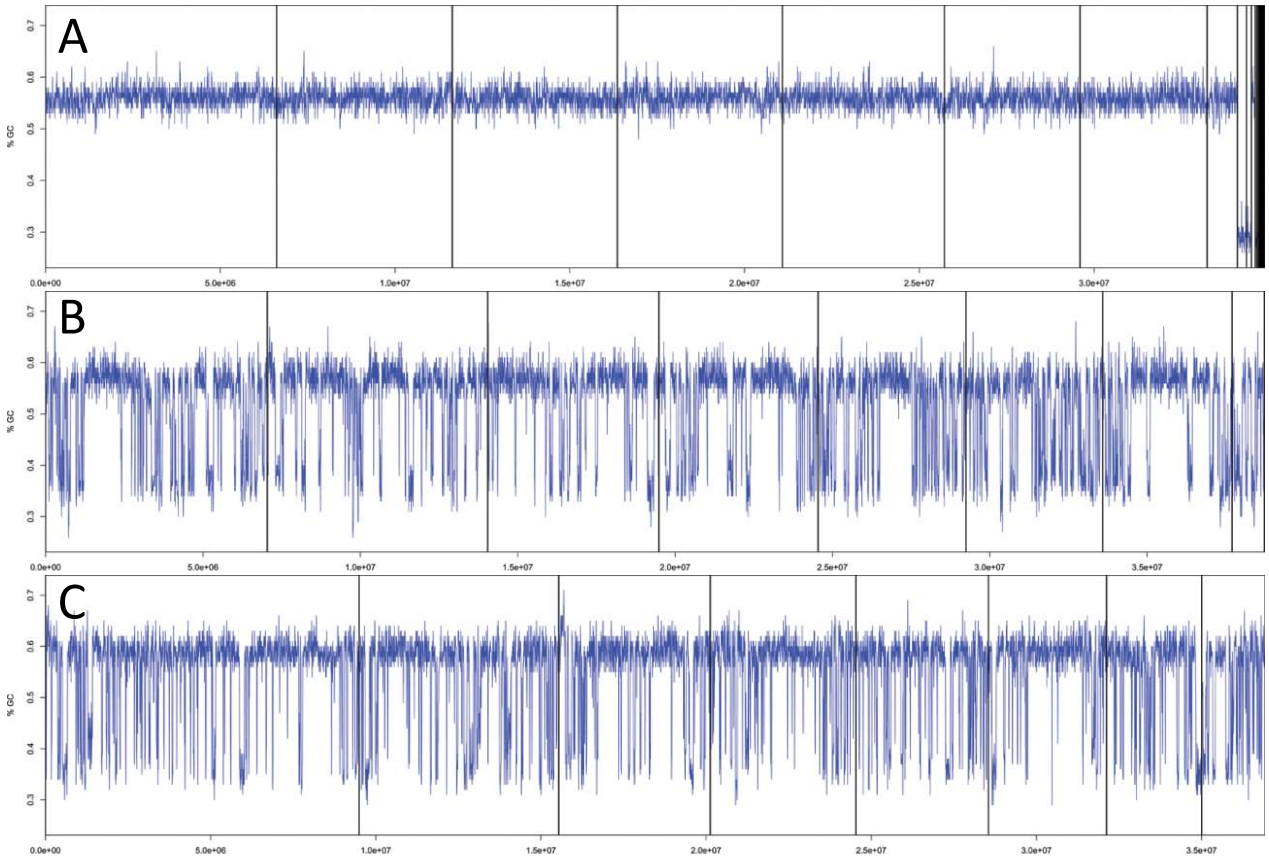


Supplementary Figure 6. Maximum likelihood phylogram of GH61 proteins in selected Sordariales. Orthologous clades present in *T. terrestris* and/or *M. thermophila* are given a letter designation and colored, and remaining clades are left black. For *T. terrestris* and *M. thermophila* the proteins are identifed by their protein ID number in version 2 of the genome release, and for other species identified by their Uniprot accession number. In some cases the gene models were substantially modified from the database release, and this is indicated by "mod" after the accession number. Several of the gene models come from likely pseudogenes based on the presence of numerous frameshifts required to encode a full-length GH61 protein. These are boxed and appended with "pse". Protein alignments were created with the linsi algorithm of MAFFT version 6.850 [61], and the tree was computed with PhyML version 3.0 using the UL3 mixture model and subtree pruning and regrafting [62]. Rendered with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).



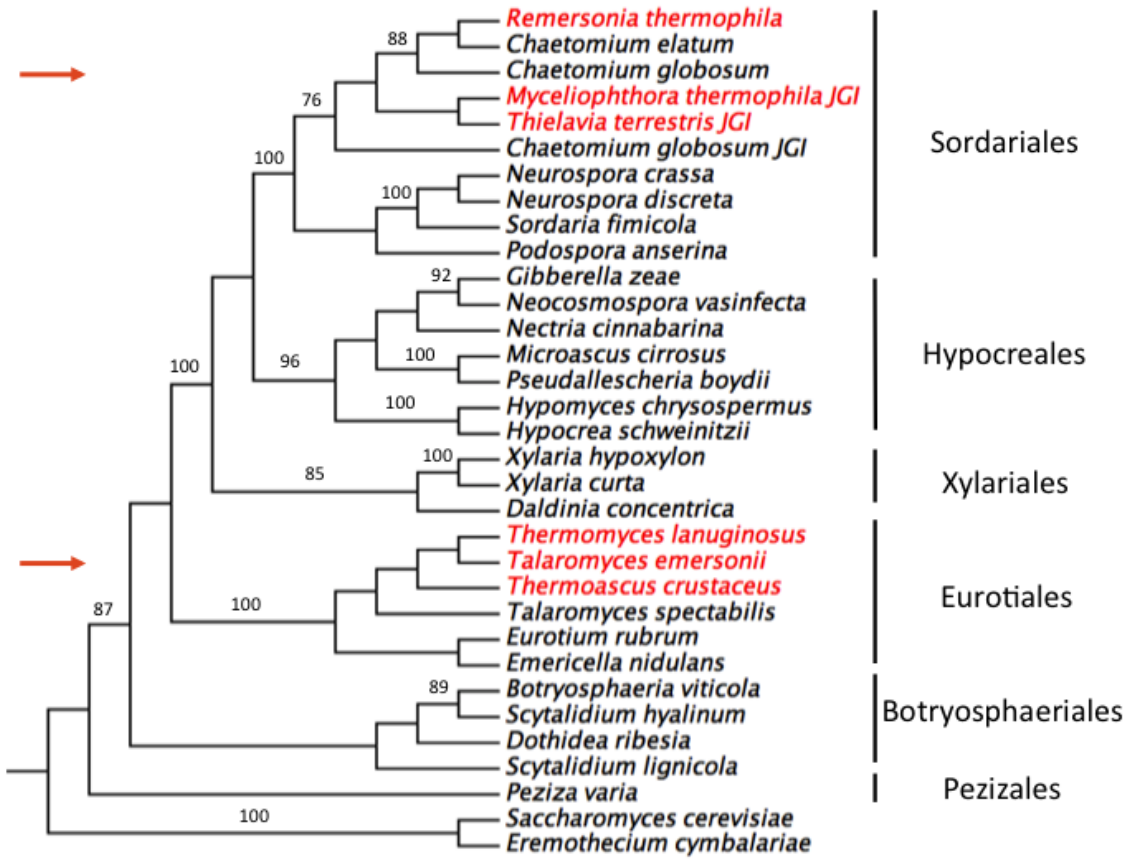
Supplementary Figure 7. Differential regulation of GH6 and GH7 orthologues. The orthologues from GH6 and GH7 endoglucanases (egl) and cellobiohydrolases (cbh) of the two thermophiles were aligned using MUSCLE. Shown here is evolutionarily relatedness of these genes and their transcript levels presented as percent of total CAZy activity. Glucose, blue; read, alfalfa straw; and green, barley straw.

Supplementary Information



Supplementary Figure 8. Percent G+C distribution along genome coordinates. All graphs share the same Y-axis range (%GC). The X-axis range (scaffold position, cumulative across all scaffolds) equals the range of the data, i.e. 0 to 35-39Mb. The individual data points are the average scaffold %GC, with scaffolds partitioned into 5kb non-overlapping segments. The black vertical bars indicate boundaries between scaffolds. A = *C. globosum*; B = *M. thermophila*; C = *T. terrestris*.

Supplementary Information



Supplementary Figure 9. Thermophily has arisen at least twice among the Ascomycota (known thermophiles shown in red type). This tree is based on bootstrap parsimony analysis of small-subunit (18S) rRNA genes. The alignment was created in ClustalW [63] and subjected to parsimony analysis using the PHYLIP (Phylogeny Inference Package) program *DNAPARS* [64]. Shown is a majority-rule consensus tree derived from 1000 bootstrap replicates. Numbers indicate nodes with bootstrap support greater than 70%. The tree could be interpreted to suggest that either thermophily has been lost in some Chaetomiaceae lineages or it has arisen more than once among species in this group. We caution against this conclusion, given that the tree is based on a single gene and we lack temperature-growth information for key presumptive mesophilic strains. We also note that the tree reveals several apparent taxonomic problems among thermophiles in both the Sordariales and Eurotiales. The following sequences were employed (parentheses indicate GenBank accession numbers or sequences obtained from the Joint Genome Institute, JGI): *Botryosphaeria viticola* (EF204510), *Chaetomium elatum* (M83257), *C. globosum* (U20379), *C. globosum* (JGI v1.0), *Daldinia concentrica* (U32402), *Dothidea ribesia* (AY016343), *Emericella nidulans* (AB008403), *Eremothecium cymbalariae* (AY046268), *Eurotium rubrum* (U00970), *Gibberella zeae* (AB250414), *Hypocrea schweinitzii* (L36986), *Hypomyces chrysospermus* (M89993), *Microascus cirrosus* (M89994), *Myceliophthora thermophila* (Sporotrichum thermophile, JGI v1.0), *Nectria cinnabarina* (AB237663), *Neocosmospora vasinfecta* (U32414), *Neurospora crassa* (X04971), *N. discreta* (JGI FGSC 8579 mat A, v1.0), *Peziza varia* (AY789390), *Podospora anserina* (X54864), *Pseudallescheria boydii* (U43915), *Remersonia thermophile* (AF044319), *Saccharomyces cerevisiae* (GQ458028), *Scytalidium hyalinum* (AF258607), *S. lignicola* (AY762623), *Sordaria fimicola* (X69851), *Talaromyces emersonii* (D88321), *Talaromyces spectabilis* (AY526473), *Thermoascus crustaceus* (AY526486), *Thermomyces lanuginosus* (EF468714), *Thielavia terrestris* (JGI v2.0), *Xylaria curta* (U32417), *X. hypoxylon* (U2037

Supplementary Table 22. Secretome and exo-proteome of *Thielavia terrestris* . List of predicted e
Functional annotation, both electronic and manual, where available is provided. The transcript ex
and barley straw were normalized and presented as RPKM. Extracellular proteins detected by ma
"identified" in the "Exo-proteome" column.

Protein_ID	Description	InterPro_ID	InterPro_description
Thite_116253	Hypothetical protein		
Thite_118360	cellobiohydrolase	IPR000254	Cellulose-binding region, fungal
Thite_124330	Hypothetical protein	IPR002818	ThiJ/PfpI
Thite_127477	trehalase	IPR001661	Glycoside hydrolase, family 37
Thite_135113	Glyoxal oxidase-like	IPR002889	Carbohydrate-binding WSC
Thite_165581	endoglucanase	IPR000254	Cellulose-binding region, fungal
Thite_2049447	Hypothetical protein		
Thite_2051196	acetylxylan esterase	IPR000254	Cellulose-binding region, fungal
Thite_2076863	Glycoside Hydrolase Family 61C protein	IPR005103	Glycoside hydrolase, family 61
Thite_2088638	Catalase/peroxidase	IPR002016	Haem peroxidase, plant/fungal/bacterial
Thite_2096194	Glycoside Hydrolase Family 114 protein		
Thite_2106069	FAD-linked oxygenase	IPR006094	FAD linked oxidase, N-terminal
Thite_2107221	Peptidase G1, eqolisin	IPR000250	Peptidase G1, eqolisin
Thite_2108049	FAD-dependent oxygenase	IPR006094	FAD linked oxidase, N-terminal
Thite_2109539	alcohol oxidase	IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_2109740	Hsp70/GRP78/bipA-homolog	IPR013126	Heat shock protein 70
Thite_2109930	alpha-mannosidase	IPR001382	Glycoside hydrolase, family 47
Thite_2110055	oligoxyloglucanase	IPR000254	Cellulose-binding region, fungal
Thite_2110162	Ceramidase	IPR006823	Neutral/alkaline nonlysosomal ceramidase
Thite_2110742	Glutaminase	IPR014870	Region of unknown function DUF1793
Thite_2110890	Glycoside Hydrolase Family 61D protein	IPR005103	Glycoside hydrolase, family 61
Thite_2110968	Hypothetical protein		

Thite_2111816	chitinase	IPR001223	Glycoside hydrolase, family 18, catalytic domain
Thite_2112302	endo-1,4-beta-galactanase	IPR011683	Glycosyl hydrolase 53
Thite_2112614	beta-galactosidase/glucoronidase	IPR006102	Glycoside hydrolase family 2, immunoglobulin-like beta-sandwich
Thite_2112626	Glycoside Hydrolase Family 61G protein	IPR000254	Cellulose-binding region, fungal
Thite_2113741	Protein disulfide isomerase	IPR006662	Thioredoxin-related
Thite_2116333	xylanase	IPR000254	Cellulose-binding region, fungal
Thite_2117106	Carboxylesterase	IPR002018	Carboxylesterase, type B
Thite_2117115	Carbohydrate-Binding Module Family 1 protein	IPR000254	Cellulose-binding region, fungal
Thite_2117251	Peptidase A1; putative vacuolar proteinase A	IPR001461	Peptidase A1
Thite_2117593	4-O-methyl-glucuronoyl methylesterase		
Thite_2117613	Glycoside Hydrolase Family 115 protein		
Thite_2117885	Carbohydrate Esterase Family 16 protein	IPR000254	Cellulose-binding region, fungal
Thite_2118148	xylanase	IPR001000	Glycoside hydrolase, family 10
Thite_2118189	oxidoreductase	IPR003953	Fumarate reductase/succinate dehydrogenase flavoprotein, N-terminal
Thite_2119040	Glycoside Hydrolase Family 61B protein	IPR000254	Cellulose-binding region, fungal
Thite_2119055	beta-glucosidase	IPR001764	Glycoside hydrolase, family 3, N-terminal
Thite_2121574	N-acetyl-glucosaminidase	IPR001540	Glycoside hydrolase, family 20
Thite_2122015	Hypothetical protein		
Thite_2122346	Hypothetical protein		
Thite_2122621	Glycoside Hydrolase Family 43 protein	IPR006710	Glycoside hydrolase, family 43
Thite_2122979	Glycoside Hydrolase Family 61E protein	IPR005103	Glycoside hydrolase, family 61
Thite_2123321	acid phosphatase	IPR007312	Phosphoesterase
Thite_2124663	Carbohydrate-Binding Module Family 1 protein	IPR000254	Cellulose-binding region, fungal
Thite_2126203	Hypothetical protein		

Thite_2131267	Hypothetical protein		
Thite_2169619	crosslinking transglycosidase	IPR000757	Glycoside hydrolase, family 16
Thite_2170113	Hypothetical protein		
Thite_2170472	atidylinositol transfer protein	IPR003172	recognition
Thite_2170662	Glycoside Hydrolase Family 61F protein	IPR000254	Cellulose-binding region, fungal
Thite_35183	chitinase	IPR015821	Glycoside hydrolase, family 18, N-terminal
Thite_52603	alpha-glucosidase	IPR000322	Glycoside hydrolase, family 31
Thite_54138	endoglucanase	IPR001722	Glycoside hydrolase, family 7
Thite_59724	Cellobiose dehydrogenase	IPR000254	Cellulose-binding region,
Thite_72629	endoglucanase	IPR000254	Cellulose-binding region, fungal
Thite_110056	Hypothetical protein		
Thite_110207	Hypothetical protein		
Thite_110267	Hypothetical protein		
Thite_111021	Hypothetical protein		
Thite_111374	Hypothetical protein		
Thite_112118	Hypothetical protein		
Thite_112906	Hypothetical protein		
Thite_113168	Hypothetical protein		
Thite_113414		IPR001214	SET
Thite_113949	Hypothetical protein		
Thite_114193	Hypothetical protein		
Thite_114312		IPR002110	Ankyrin
Thite_114680	Hypothetical protein		
Thite_115269	Hypothetical protein		
Thite_115524	Hypothetical protein		
Thite_115795	Glycoside Hydrolase Family 61N protein	IPR005103	Glycoside hydrolase, family 61
Thite_115847		IPR001969	Peptidase aspartic, active site
Thite_116047		IPR002345	Lipocalin
Thite_116456	Hypothetical protein		
Thite_117079	Hypothetical protein		
Thite_117183	Hypothetical protein		
Thite_117234		IPR001461	Peptidase A1
Thite_117394		IPR000209	Peptidase S8 and S53, subtilisin, kexin, sedolisin
Thite_118257		IPR001128	Cytochrome P450
Thite_118353	endoglucanase	IPR001547	Glycoside hydrolase, family 5
Thite_118376		IPR007197	Radical SAM

Thite_118864		IPR000772	Ricin B lectin
Thite_120170	exo-beta-1,3-glucanase	IPR011050	Pectin lyase fold/virulence factor
Thite_120212		IPR006045	Cupin 1
Thite_120596		IPR002921	Lipase, class 3
Thite_120654	Hypothetical protein		
Thite_120962	xylanase	IPR001000	Glycoside hydrolase, family 10
Thite_121141	Peptidase S10, serine carboxypeptidase	IPR001563	Peptidase S10, serine carboxypeptidase
Thite_121467	Carbohydrate-Binding Module Family 50 protein	IPR002482	Peptidoglycan-binding LysM
Thite_122288		IPR006094	FAD linked oxidase, N-terminal
Thite_122346	Hypothetical protein		
Thite_122351	exo-polygalacturonase	IPR000743	Glycoside hydrolase, family 28
Thite_122381		IPR006094	FAD linked oxidase, N-terminal
Thite_122458		IPR013027	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
Thite_122459	Hypothetical protein		
Thite_122540		IPR002372	Pyrrolo-quinoline quinone
Thite_122544	pectin methyl esterase	IPR000070	Pectinesterase, catalytic
Thite_122696		IPR003822	Paired amphipathic helix
Thite_123205	endoglucanase	IPR000254	Cellulose-binding region, fungal
Thite_123323	Peptidase A1; aspartic protease	IPR001461	Peptidase A1
Thite_123328		IPR006094	FAD linked oxidase, N-terminal
Thite_123413	Hypothetical protein		
Thite_123924		IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_124055		IPR001117	Multicopper oxidase, type 1
Thite_124715		IPR002227	Tyrosinase
Thite_124733	beta-galactosidase	IPR006101	Glycoside hydrolase, family 2
Thite_124883	endoglucanase	IPR001547	Glycoside hydrolase, family 5
Thite_124911		IPR001117	Multicopper oxidase, type 1

Thite_124995		IPR000627	Intradiol ring-cleavage dioxygenase, C-terminal
Thite_125001		IPR006680	Amidohydrolase 1
Thite_125074		IPR000028	Chloroperoxidase
Thite_125239	beta-glucosidase	IPR001764	Glycoside hydrolase, family 3, N-terminal
Thite_125400		IPR001199	Cytochrome b5
Thite_125550	Hypothetical protein		
Thite_125571	exo-beta-1,3-glucanase	IPR011050	Pectin lyase fold/virulence factor
Thite_125575		IPR002227	Tyrosinase
Thite_125640		IPR003042	Aromatic-ring hydroxylase
Thite_125668		IPR002018	Carboxylesterase, type B
Thite_126254		IPR005556	SUN
Thite_126432	cellobiohydrolase	IPR000254	Cellulose-binding region, fungal
Thite_127241	Hypothetical protein		
Thite_128130	Glycoside Hydrolase Family 61L protein	IPR005103	Glycoside hydrolase, family 61
Thite_131920	Hypothetical protein		
Thite_132424	Hypothetical protein		
Thite_133795	Hypothetical protein		
Thite_134020	endoglucanase	IPR001722	Glycoside hydrolase, family 7
Thite_152002	Hypothetical protein		
Thite_152112		IPR001461	Peptidase A1
Thite_152269		IPR000408	Regulator of chromosome condensation, RCC1
Thite_152679		IPR011058	Cyanovirin-N
Thite_152693	alpha-amylase	IPR002044	Glycoside hydrolase, carbohydrate-binding
Thite_153552		IPR008960	Carbohydrate-binding family 9/cellobiose dehydrogenase, cytochrome
Thite_154280		IPR008914	Phosphatidylethanolamine-binding protein PEBP
Thite_154425	chitinase	IPR001002	Chitin-binding, type 1
Thite_155164		IPR000627	Intradiol ring-cleavage dioxygenase, C-terminal
Thite_156041	Hypothetical protein		
Thite_156464	Hypothetical protein		
Thite_156478	Hypothetical protein		
Thite_156520	Hypothetical protein		
Thite_156575	rhamnogalacturonase	IPR000743	Glycoside hydrolase, family 28

Thite_157561	alpha-glucosidase	IPR004888	Glycoside hydrolase, family 63
Thite_157646	Hypothetical protein		
Thite_157813		IPR000675	Cutinase
Thite_157851	Carbohydrate-Binding Module Family 18 protein	IPR001002	Chitin-binding, type 1
Thite_157867	Hypothetical protein		
Thite_158355	Hypothetical protein		
Thite_158792		IPR007867	Glucose-methanol-choline oxidoreductase, C-terminal
Thite_159055		IPR001128	Cytochrome P450
Thite_159718		IPR000209	Peptidase S8 and S53, subtilisin, kexin, sedolisin
Thite_159943	Hypothetical protein		
Thite_159967		IPR001338	Fungal hydrophobin
Thite_160100	Hypothetical protein		
Thite_160282	Hypothetical protein		
Thite_160560		IPR012307	Xylose isomerase-type TIM barrel
Thite_160890	Hypothetical protein		
Thite_160953		IPR013027	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
Thite_160963		IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_162651	chitinase	IPR001223	Glycoside hydrolase, family 18, catalytic domain
Thite_164568	glucoamylase	IPR002044	Glycoside hydrolase, carbohydrate-binding
Thite_164628	Hypothetical protein		
Thite_165277	endopolygalacturonase	IPR000743	Glycoside hydrolase, family 28
Thite_16657	Hypothetical protein		
Thite_170174	Glycoside Hydrolase Family 61P protein	IPR005103	Glycoside hydrolase, family 61
Thite_18434	Hypothetical protein		
Thite_2021175		IPR006162	Phosphopantetheine attachment site
Thite_2022036	Hypothetical protein		
Thite_2022936	Peptidase A1; aspartic protease	IPR001461	Peptidase A1
Thite_2023982	Hypothetical protein		
Thite_2025173	Peptidase S8 and S53, subtilisin, kexin, sedolisin	IPR015366	Peptidase S53, propeptide

Thite_2029826		IPR000782	Beta-Ig-H3/fasciclin
Thite_2037112		IPR001087	Lipase, GDSL
Thite_2037128	Hypothetical protein		
Thite_2037188		IPR002018	Carboxylesterase, type B
Thite_2037683		IPR001810	Cyclin-like F-box
Thite_2037763		IPR006094	FAD linked oxidase, N-terminal
Thite_2037962	exo-polygalacturonase	IPR000743	Glycoside hydrolase, family 28
Thite_2038308		IPR001128	Cytochrome P450
Thite_2038446		IPR001117	Multicopper oxidase, type 1
Thite_2038768	Hypothetical protein		
Thite_2038876		IPR001117	Multicopper oxidase, type 1
Thite_2039222		IPR013027	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
Thite_2039882		IPR006094	FAD linked oxidase, N-terminal
Thite_2040058		IPR002227	Tyrosinase
Thite_2040127	Glycoside Hydrolase Family 61M protein	IPR005103	Glycoside hydrolase, family 61
Thite_2040192		IPR006094	FAD linked oxidase, N-terminal
Thite_2040826	Hypothetical protein		
Thite_2040964		IPR001128	Cytochrome P450
Thite_2040983	Hypothetical protein		
Thite_2040984	Glycoside Hydrolase Family 43 protein	IPR006710	Glycoside hydrolase, family 43
Thite_2041205		IPR002018	Carboxylesterase, type B
Thite_2041678		IPR000028	Chloroperoxidase
Thite_2041884	alpha-L-arabinofuranosidase	IPR007934	Alpha-L-arabinofuranosidase B
Thite_2041986	Hypothetical protein		
Thite_2041988	Hypothetical protein		
Thite_2042100	xylanase	IPR001137	Glycoside hydrolase, family 11
Thite_2042687	Hypothetical protein		
Thite_2042744	Carbohydrate Esterase Family 3 protein	IPR001087	Lipase, GDSL
Thite_2042795	Glycoside Hydrolase Family 16 protein	IPR000254	Cellulose-binding region, fungal
Thite_2043158		IPR011058	Cyanovirin-N
Thite_2044372	Glycoside Hydrolase Family 61S protein	IPR005103	Glycoside hydrolase, family 61
Thite_2044512	Hypothetical protein		

Thite_2044737	Hypothetical protein		
Thite_2044936	beta-glucosidase	IPR001764	Glycoside hydrolase, family 3, N-terminal
Thite_2045005	Beta-galactosidase	IPR001944	Glycoside hydrolase, family 35
Thite_2045382		IPR014778	Myb, DNA-binding
Thite_2045470	rhamnogalacturonan lyase	IPR013784	Carbohydrate-binding-like fold
Thite_2046186		IPR013658	SMP-30/Gluconolaconase/LRE-like region
Thite_2046339	beta-galactosidase	IPR006102	Glycoside hydrolase family 2, immunoglobulin-like beta-sandwich
Thite_2046868		IPR002227	Tyrosinase
Thite_2047269	Hypothetical protein		
Thite_2047488		IPR005152	Secretory lipase
Thite_2047729	Beta-galactosidase	IPR001944	Glycoside hydrolase, family 35
Thite_2047936	Hypothetical protein		
Thite_2048092	Hypothetical protein		
Thite_2048418		IPR009078	Ferritin/ribonucleotide reductase-like
Thite_2048962		IPR016040	NAD(P)-binding
Thite_2049026		IPR002403	Cytochrome P450, E-class, group IV
Thite_2049710		IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_2049779	exo-glucosaminidase	IPR006102	Glycoside hydrolase family 2, immunoglobulin-like beta-sandwich
Thite_2050221		IPR008183	Aldose 1-epimerase
Thite_2050866		IPR001568	Ribonuclease T2
Thite_2050870	xylanase	IPR001137	Glycoside hydrolase, family 11
Thite_2051538	Hypothetical protein		
Thite_2051582	Hypothetical protein		
Thite_2052408	Hypothetical protein		
Thite_2052779	Hypothetical protein		
Thite_2053289		IPR013830	Esterase, SGNH hydrolase-type
Thite_2053343	Hypothetical protein		
Thite_2053585		IPR001578	Peptidase C12, ubiquitin carboxyl-terminal hydrolase 1
Thite_2053822	Hypothetical protein		

Thite_2053903		IPR002018	Carboxylesterase, type B
Thite_2053998		IPR006025	Peptidase M, neutral zinc metallopeptidases, zinc-binding site
Thite_2054124	Polysaccharide Lyase Family 7 protein	IPR014895	Alginate lyase 2
Thite_2054382	chitinase	IPR001002	Chitin-binding, type 1
Thite_2054431	Hypothetical protein		
Thite_2054543	Glycoside Hydrolase Family 61O protein	IPR005103	Glycoside hydrolase, family 61
Thite_2054659	Hypothetical protein		
Thite_2055159	Peptidase C1A, papain-like cysteine protease	IPR000668	Peptidase C1A, papain C-terminal
Thite_2055217	Hypothetical protein		
Thite_2055322		IPR006620	Prolyl 4-hydroxylase, alpha subunit
Thite_2055580	Carbohydrate Esterase Family 15 protein		
Thite_2055725	Hypothetical protein		
Thite_2056623		IPR001466	Beta-lactamase
Thite_2056626	Hypothetical protein		
Thite_2056800		IPR008960	Carbohydrate-binding family 9/cellobiose dehydrogenase, cytochrome
Thite_2057025	Hypothetical protein		
Thite_2057091		IPR001214	SET
Thite_2057343	Carbohydrate Esterase Family 3 protein	IPR001087	Lipase, GDSL
Thite_2057535	Hypothetical protein		
Thite_2057949		IPR013027	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
Thite_2058287	Hypothetical protein		
Thite_2058539	Hypothetical protein		
Thite_2058635	Hypothetical protein		
Thite_2059294		IPR008960	Carbohydrate-binding family 9/cellobiose dehydrogenase, cytochrome
Thite_2059487	Glycoside Hydrolase Family 61T protein	IPR005103	Glycoside hydrolase, family 61
Thite_2061473	mixed-linked glucanase	IPR000757	Glycoside hydrolase, family 16
Thite_2064577		IPR000782	Beta-Ig-H3/fasciclin
Thite_2064646	Hypothetical protein		
Thite_2065711	Hypothetical protein		

Thite_2067702		IPR004843	Metallophosphoesterase
Thite_2071939		IPR004843	Metallophosphoesterase
Thite_2076439	Peptidase S8 and S53, subtilisin, kexin, sedolisin	IPR000209	Peptidase S8 and S53, subtilisin, kexin, sedolisin
Thite_2077609	Hypothetical protein		
Thite_2077826		IPR002472	Palmitoyl protein thioesterase
Thite_2078736		IPR000991	Glutamine amidotransferase class-I, C-terminal
Thite_2083723	Hypothetical protein		
Thite_2084410		IPR009078	Ferritin/ribonucleotide reductase-like
Thite_2084523	Hypothetical protein		
Thite_2084620	Hypothetical protein		
Thite_2084692	Hypothetical protein		
Thite_2085213	Hypothetical protein		
Thite_2085302	Hypothetical protein		
Thite_2085304	Hypothetical protein		
Thite_2085311	Hypothetical protein		
Thite_2085319		IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_2085357	Hypothetical protein		
Thite_2085417	Hypothetical protein		
Thite_2085430	Glycoside Hydrolase Family 61 protein	IPR005103	Glycoside hydrolase, family 61
Thite_2085916	Hypothetical protein		
Thite_2086101	Hypothetical protein		
Thite_2086109	Hypothetical protein		
Thite_2086141	Hypothetical protein		
Thite_2086167	Hypothetical protein		
Thite_2087701	Hypothetical protein		
Thite_2087754	Carbohydrate-Binding Module Family 18 protein	IPR001002	Chitin-binding, type 1
Thite_2087862	Hypothetical protein		
Thite_2088011	Hypothetical protein		
Thite_2088815		IPR002227	Tyrosinase
Thite_2089046	Hypothetical protein		
Thite_2089066	Hypothetical protein		
Thite_2089087	Hypothetical protein		
Thite_2089133	Hypothetical protein		
Thite_2089412	Hypothetical protein		
Thite_2089704		IPR009327	Protein of unknown function DUF985
Thite_2089794	Hypothetical protein		

Thite_2089905	Glycosyltransferase Family 1 protein		
Thite_2090199	Hypothetical protein		
Thite_2090214	Hypothetical protein		
Thite_2091099	Hypothetical protein		
Thite_2091336		IPR008427	Extracellular membrane protein, 8-cysteine region, CFEM
Thite_2091422	Hypothetical protein		
Thite_2092533	Peptidase A1; aspartic protease	IPR001461	Peptidase A1
Thite_2093413	Hypothetical protein		
Thite_2095073	Glycosyltransferase Family 32 protein	IPR007577	Glycosyltransferase sugar-binding region containing DXD motif
Thite_2096446		IPR000719	Protein kinase, core
Thite_2097422	Hypothetical protein		
Thite_2097476		IPR002110	Ankyrin
Thite_2097551		IPR010816	Heterokaryon incompatibility Het-C
Thite_2106082		IPR001128	Cytochrome P450
Thite_2106106	Hypothetical protein		
Thite_2106207	Hypothetical protein		
Thite_2106220	Hypothetical protein		
Thite_2106238		IPR002885	Pentatricopeptide repeat
Thite_2106442	Hypothetical protein		
Thite_2106455	Hypothetical protein		
Thite_2106536	Hypothetical protein		
Thite_2106556	Glycoside Hydrolase Family 61J protein	IPR005103	Glycoside hydrolase, family 61
Thite_2106630		IPR001623	Heat shock protein DnaJ, N-terminal
Thite_2106686		IPR001128	Cytochrome P450
Thite_2106743		IPR001128	Cytochrome P450
Thite_2107066		IPR006662	Thioredoxin-related
Thite_2107105		IPR002016	Haem peroxidase, plant/fungal/bacterial
Thite_2107118	beta-glucosidase	IPR017853	Glycoside hydrolase, catalytic core
Thite_2107144	Hypothetical protein		
Thite_2107170	Hypothetical protein		
Thite_2107217	Glycoside Hydrolase Family 31 protein	IPR000322	Glycoside hydrolase, family 31
Thite_2107235		IPR013244	Secretory pathway Sec39
Thite_2107410	Peptidase M14, carboxypeptidase A	IPR000834	Peptidase M14, carboxypeptidase A
Thite_2107492	Hypothetical protein		

Thite_2107548	Hypothetical protein		
Thite_2107700		IPR002130	Peptidyl-prolyl cis-trans isomerase, cyclophilin-type
Thite_2107714		IPR006094	FAD linked oxidase, N-terminal
Thite_2107757	Hypothetical protein		
Thite_2107758	endoglucanase	IPR001547	Glycoside hydrolase, family 5
Thite_2107766	Peptidase S8 and S53; possible tripeptidyl-peptidase	IPR015366	Peptidase S53, propeptide
Thite_2107785		IPR006771	Blastomyces yeast-phase-specific protein
Thite_2107793		IPR009030	Growth factor, receptor
Thite_2107799	xylanase	IPR001137	Glycoside hydrolase, family 11
Thite_2107804		IPR002143	Ribosomal protein L1
Thite_2107854	Hypothetical protein		
Thite_2107902	Hypothetical protein		
Thite_2107926	exo-α-L-1,5-arabinanase	IPR002860	BNR repeat
Thite_2107935	alpha-mannosidase	IPR001382	Glycoside hydrolase, family 47
Thite_2107939	Hypothetical protein		
Thite_2107967	xylanase	IPR000254	Cellulose-binding region, fungal
Thite_2107973		IPR000026	Guanine-specific ribonuclease N1 and T1
Thite_2108040	Hypothetical protein		
Thite_2108084		IPR008427	Extracellular membrane protein, 8-cysteine region, CFEM
Thite_2108097		IPR010829	Cerato-platanin
Thite_2108246	Hypothetical protein		
Thite_2108322		IPR002302	Leucyl-tRNA synthetase, class Ia, bacterial/mitochondrial
Thite_2108499		IPR010993	Sterile alpha motif homology
Thite_2108507	alpha-glucosidase	IPR004888	Glycoside hydrolase, family 63
Thite_2108701		IPR014870	Region of unknown function DUF1793
Thite_2108762	xylanase	IPR001000	Glycoside hydrolase, family 10
Thite_2108763		IPR000836	Phosphoribosyltransferase

Thite_2108799	Hypothetical protein		
Thite_2108822	Hypothetical protein		
Thite_2109107		IPR003095	Heat shock protein DnaJ
Thite_2109170	Hypothetical protein		
Thite_2109181	Peptidase G1, eqolisin	IPR000250	Peptidase G1, eqolisin
Thite_2109230		IPR000209	Peptidase S8 and S53, subtilisin, kexin, sedolisin
Thite_2109377	beta-galactosidase	IPR006101	Glycoside hydrolase, family 2
Thite_2109390		IPR000683	Oxidoreductase, N-terminal
Thite_2109436	exo-beta-1,3-glucanase	IPR011050	Pectin lyase fold/virulence factor
Thite_2109515	Hypothetical protein		
Thite_2109545	N,O-diacetylmuramidase	IPR002053	Glycoside hydrolase, family 25
Thite_2109753		IPR002579	Methionine sulphoxide reductase B
Thite_2109832		IPR013094	Alpha/beta hydrolase fold-3
Thite_2109849		IPR013094	Alpha/beta hydrolase fold-3
Thite_2109939		IPR002052	N-6 adenine-specific DNA methylase, conserved site
Thite_2109946	endo-1,3-beta-glucanase	IPR005200	Glycoside hydrolase, family 81
Thite_2110040	Hypothetical protein		
Thite_2110076	Hypothetical protein		
Thite_2110097	Hypothetical protein		
Thite_2110172	Hypothetical protein		
Thite_2110450		IPR001128	Cytochrome P450
Thite_2110587	Hypothetical protein		
Thite_2110616		IPR002347	Glucose/ribitol dehydrogenase
Thite_2110654	Hypothetical protein		
Thite_2110681		IPR012913	Glucosidase II beta subunit-like
Thite_2110760	Hypothetical protein		
Thite_2110854		IPR002198	Short-chain dehydrogenase/reductase SDR
Thite_2110888		IPR011118	Tannase and feruloyl esterase
Thite_2110902		IPR013247	SH3, type 3
Thite_2110908	Glycosyltransferase Family 32 protein	IPR007577	Glycosyltransferase sugar-binding region containing DXD motif

Thite_2110911	Hypothetical protein		
Thite_2110918	Glycosyltransferase Family 69 protein		
Thite_2110920	Glycosyltransferase Family 25 protein	IPR002654	Glycosyl transferase, family 25
Thite_2110954	arabinofuranosidase	IPR005193	Glycoside hydrolase, family 62
Thite_2110956		IPR008972	Cupredoxin
Thite_2110957	endoglucanase	IPR000254	Cellulose-binding region, fungal
Thite_2110967	Hypothetical protein		
Thite_2110971	Peptidase S28	IPR008758	Peptidase S28
Thite_2110984	Hypothetical protein		
Thite_2111001	Hypothetical protein		
Thite_2111054	Hypothetical protein		
Thite_2111099	Hypothetical protein		
Thite_2111173	Carbohydrate-Binding Module Family 18 protein	IPR001002	Chitin-binding, type 1
Thite_2111175	Hypothetical protein		
Thite_2111241		IPR001283	Allergen V5/Tpx-1 related
Thite_2111353	Carbohydrate Esterase Family 2 protein	IPR000254	Cellulose-binding region, fungal
Thite_2111513	Hypothetical protein		
Thite_2111662		IPR004183	Extradiol ring-cleavage dioxygenase, class III enzyme, subunit B
Thite_2111669	Hypothetical protein		
Thite_2111687		IPR000120	Amidase signature enzyme
Thite_2111723	Hypothetical protein		
Thite_2111726	alpha-mannosidase	IPR001382	Glycoside hydrolase, family 47
Thite_2111903	Glycosyltransferase Family 69 protein		
Thite_2112001		IPR004843	Metallophosphoesterase
Thite_2112061		IPR003154	S1/P1 nuclease
Thite_2112063	Hypothetical protein		
Thite_2112109	Hypothetical protein		
Thite_2112145		IPR000073	Alpha/beta hydrolase fold-1
Thite_2112176		IPR012420	CBP4
Thite_2112292	Hypothetical protein		
Thite_2112307		IPR014025	Glutaredoxin subgroup
Thite_2112326		IPR002889	Carbohydrate-binding WSC
Thite_2112401		IPR004843	Metallophosphoesterase
Thite_2112556	Hypothetical protein		

Thite_2112596		IPR001509	NAD-dependent epimerase/dehydratase
Thite_2112613	Hypothetical protein		
Thite_2112801	Hypothetical protein		
Thite_2112829	Hypothetical protein		
Thite_2113399	mixed-linked glucanase	IPR008985	Concanavalin A-like lectin/glucanase
Thite_2113460	Hypothetical protein		
Thite_2113568		IPR010816	Heterokaryon incompatibility Het-C
Thite_2113631	Hypothetical protein		
Thite_2113646		IPR001932	Protein phosphatase 2C-related
Thite_2113690	Hypothetical protein		
Thite_2113699		IPR013838	Beta tubulin, autoregulation binding site
Thite_2113709	Hypothetical protein		
Thite_2113716		IPR001424	Superoxide dismutase, copper/zinc binding
Thite_2113751	Hypothetical protein		
Thite_2113854		IPR002889	Carbohydrate-binding WSC
Thite_2113898	Peptidase A1; aspartic protease	IPR001461	Peptidase A1
Thite_2113936	Glycosyltransferase Family 17 protein	IPR006813	Glycosyl transferase, family 17
Thite_2114094	Hypothetical protein		
Thite_2114176		IPR000276	GPCR, rhodopsin-like
Thite_2114229		IPR001461	Peptidase A1
Thite_2114232	mixed-linked glucanase	IPR000757	Glycoside hydrolase, family 16
Thite_2114474	mannanase	IPR001547	Glycoside hydrolase, family 5
Thite_2114488	Glycosyltransferase Family 69 protein		
Thite_2114499	Hypothetical protein		
Thite_2114542	Peptidase M28	IPR007484	Peptidase M28
Thite_2114550		IPR008972	Cupredoxin
Thite_2114605	Carbohydrate Esterase Family 3 protein	IPR001087	Lipase, GDSL
Thite_2114620	rhamnogalacturonase	IPR000743	Glycoside hydrolase, family 28
Thite_2114681		IPR001763	Rhodanese-like
Thite_2114718	Hypothetical protein		

Thite_2114740	Glycosyltransferase Family 32 protein	IPR007577	Glycosyltransferase sugar-binding region containing DXD motif
Thite_2114758	alpha-mannosidase	IPR012939	Glycosyl hydrolase 92
Thite_2114773	Hypothetical protein		
Thite_2114807	Hypothetical protein		
Thite_2114879	Hypothetical protein		
Thite_2114943		IPR001453	Molybdopterin binding
Thite_2115007		IPR005556	SUN
Thite_2115086		IPR008313	Uncharacterised conserved protein UCP028846
Thite_2115245	Hypothetical protein		
Thite_2115447	Hypothetical protein		
Thite_2115456	Hypothetical protein		
Thite_2115533	Hypothetical protein		
Thite_2115598		IPR000215	Protease inhibitor I4, serpin
Thite_2115764	Hypothetical protein		
Thite_2115891	Glycoside Hydrolase Family 39 protein		
Thite_2115931		IPR008427	Extracellular membrane protein, 8-cysteine region, CFEM
Thite_2115960		IPR003042	Aromatic-ring hydroxylase
Thite_2115964		IPR008427	Extracellular membrane protein, 8-cysteine region, CFEM
Thite_2115973	Hypothetical protein		
Thite_2116109		IPR006139	D-isomer specific 2-hydroxyacid dehydrogenase, catalytic region
Thite_2116193	Hypothetical protein		
Thite_2116225	Glycosyltransferase Family 71 protein		
Thite_2116251		IPR006094	FAD linked oxidase, N-terminal
Thite_2116283		IPR002401	Cytochrome P450, E-class, group I
Thite_2116294		IPR017853	Glycoside hydrolase, catalytic core
Thite_2116299		IPR001077	O-methyltransferase, family 2
Thite_2116300		IPR004792	HI0933-like protein

Thite_2116324		IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_2116371	exo-α-L-1,5-arabinanase	IPR002860	BNR repeat
Thite_2116387	Hypothetical protein		
Thite_2116431	pectinesterase/polygalacturonase bifunctional protein	IPR000743	Glycoside hydrolase, family 28
Thite_2116458		IPR003042	Aromatic-ring hydroxylase
Thite_2116487		IPR003042	Aromatic-ring hydroxylase
Thite_2116536	Glycoside Hydrolase Family 61A protein	IPR005103	Glycoside hydrolase, family 61
Thite_2116556		IPR004843	Metallophosphoesterase
Thite_2116565	Glycosyltransferase Family 24 protein	IPR009448	UDP-glucose:Glycoprotein Glucosyltransferase
Thite_2116660	Hypothetical protein		
Thite_2116676	Hypothetical protein		
Thite_2116781		IPR001623	Heat shock protein DnaJ, N-terminal
Thite_2116855		IPR001179	Peptidyl-prolyl cis-trans isomerase, FKBP-type
Thite_2116882	chitinase	IPR001223	Glycoside hydrolase, family 18, catalytic domain
Thite_2116896		IPR003953	Fumarate reductase/succinate dehydrogenase flavoprotein, N-terminal
Thite_2116923	Peptidase S8 and S53; putative vacuolar proteinase B	IPR000209	Peptidase S8 and S53, subtilisin, kexin, sedolisin
Thite_2116952		IPR000873	AMP-dependent synthetase and ligase
Thite_2116964		IPR000917	Sulphatase
Thite_2117004		IPR006863	Erv1/Alr
Thite_2117008	Hypothetical protein		
Thite_2117031	Glycosyltransferase Family 31 protein		
Thite_2117036		IPR011058	Cyanovirin-N
Thite_2117052		IPR007484	Peptidase M28
Thite_2117069	Hypothetical protein		
Thite_2117076	mannanase	IPR000254	Cellulose-binding region, fungal
Thite_2117103	Hypothetical protein		
Thite_2117121	Hypothetical protein		
Thite_2117203	alpha-mannosidase	IPR001382	Glycoside hydrolase, family 47

Thite_2117388	Hypothetical protein		
Thite_2117509		IPR002018	Carboxylesterase, type B
Thite_2117515		IPR000782	Beta-Ig-H3/fasciclin
Thite_2117640	Glycoside Hydrolase Family 30 protein	IPR001139	Glycoside hydrolase, family 30
Thite_2117649	xylanase	IPR000254	Cellulose-binding region, fungal
Thite_2117656	Hypothetical protein		
Thite_2117762	endoglucanase	IPR002594	Glycoside hydrolase, family 12
Thite_2117790	feruloyl esterase		
Thite_2117792		IPR009030	Growth factor, receptor
Thite_2117845	Hypothetical protein		
Thite_2117874		IPR007197	Radical SAM
Thite_2117895	polysaccharide deacetylase	IPR002509	Polysaccharide deacetylase
Thite_2117982		IPR003892	Ubiquitin system component Cue
Thite_2117994		IPR004843	Metallophosphoesterase
Thite_2118041	Glycosyltransferase Family 90 protein	IPR006598	Lipopolysaccharide-modifying protein
Thite_2118057	Glycosyltransferase Family 31 protein		
Thite_2118068		IPR002227	Tyrosinase
Thite_2118071		IPR003953	Fumarate reductase/succinate dehydrogenase flavoprotein, N-terminal
Thite_2118076		IPR002018	Carboxylesterase, type B
Thite_2118088	Hypothetical protein		
Thite_2118104	arabinan endo-1,5-alpha-L-arabinosidase	IPR006710	Glycoside hydrolase, family 43
Thite_2118136	beta-glucosidase	IPR001764	Glycoside hydrolase, family 3, N-terminal
Thite_2118175	Peptidase M14, carboxypeptidase A	IPR000834	Peptidase M14, carboxypeptidase A
Thite_2118209	Hypothetical protein		
Thite_2118256	Hypothetical protein		
Thite_2118359	exo-beta-1,3-glucanase	IPR016160	Aldehyde dehydrogenase, conserved site
Thite_2118364	Hypothetical protein		
Thite_2118367	Glycosyltransferase Family 25 protein	IPR002654	Glycosyl transferase, family 25
Thite_2118436	Peptidase S10, putative carboxypeptidase Y	IPR001563	Peptidase S10, serine carboxypeptidase
Thite_2118697		IPR002018	Carboxylesterase, type B
Thite_2118884	Hypothetical protein		

Thite_2118894		IPR006620	Prolyl 4-hydroxylase, alpha subunit
Thite_2119018	Glycoside Hydrolase Family 16 protein	IPR008985	Concanavalin A-like lectin/glucanase
Thite_2119063	Hypothetical protein		
Thite_2119205	Hypothetical protein		
Thite_2119320		IPR002889	Carbohydrate-binding WSC
Thite_2119352		IPR001128	Cytochrome P450
Thite_2119465	Hypothetical protein		
Thite_2119483	Carbohydrate Esterase Family 4 protein	IPR001002	Chitin-binding, type 1
Thite_2119639		IPR011545	DNA/RNA helicase, DEAD/DEAH box type, N-terminal
Thite_2119706	alpha-mannanase	IPR005198	Glycoside hydrolase, family 76
Thite_2119804		IPR002889	Carbohydrate-binding WSC
Thite_2119832		IPR001604	DNA/RNA non-specific endonuclease
Thite_2119862		IPR004455	NADP oxidoreductase, coenzyme F420-dependent
Thite_2119883	Hypothetical protein		
Thite_2119941	Glycosyltransferase Family 90 protein	IPR006598	Lipopolysaccharide-modifying protein
Thite_2119967		IPR006222	Glycine cleavage T-protein, N-terminal
Thite_2120008		IPR001948	Peptidase M18, aminopeptidase I
Thite_2120068	Hypothetical protein		
Thite_2120233	beta-1,6-galactanase	IPR001547	Glycoside hydrolase, family 5
Thite_2120241		IPR002067	Mitochondrial carrier protein
Thite_2120310	Carbohydrate-Binding Module Family 52 protein		
Thite_2120324		IPR003737	N-acetylglucosaminyl phosphatidylinositol deacetylase
Thite_2120326	polysaccharide deacetylase	IPR002509	Polysaccharide deacetylase
Thite_2120373	cutinase	IPR011150	Cutinase, monofunctional
Thite_2120390	Hypothetical protein		
Thite_2120498	Hypothetical protein		
Thite_2120503	Hypothetical protein		

Thite_2120506		IPR001128	Cytochrome P450
Thite_2120540		IPR002872	Proline dehydrogenase
Thite_2120545	Peptidase A1; aspartic protease	IPR001461	Peptidase A1
Thite_2120602		IPR001117	Multicopper oxidase, type 1
Thite_2120607	Hypothetical protein		
Thite_2120753	alpha-mannanase	IPR008928	Six-hairpin glycosidase-like
Thite_2120825	Glycoside Hydrolase Family 79 protein		
Thite_2120847		IPR007219	Fungal specific transcription factor
Thite_2120863	Peptidase M28	IPR007484	Peptidase M28
Thite_2120876		IPR001971	Ribosomal protein S11
Thite_2120948	Glycoside Hydrolase Family 5 protein	IPR001547	Glycoside hydrolase, family 5
Thite_2121059		IPR013880	Yos1-like
Thite_2121102		IPR002114	Phosphotransferase system, HPr serine phosphorylation site
Thite_2121208	Hypothetical protein		
Thite_2121305	Hypothetical protein		
Thite_2121318		IPR007266	Endoplasmic reticulum oxidoreductin 1
Thite_2121448	Hypothetical protein		
Thite_2121514		IPR008313	Uncharacterised conserved protein UCP028846
Thite_2121649	Hypothetical protein		
Thite_2121650		IPR002921	Lipase, class 3
Thite_2121678	Hypothetical protein		
Thite_2121715		IPR000219	Dbl homology (DH) domain
Thite_2121854	Hypothetical protein		
Thite_2121909	Hypothetical protein		
Thite_2121983	chitinase	IPR015821	Glycoside hydrolase, family 18, N-terminal
Thite_2122016		IPR001041	Ferredoxin
Thite_2122060		IPR001199	Cytochrome b5
Thite_2122167		IPR001424	Superoxide dismutase, copper/zinc binding
Thite_2122212		IPR012312	Hemerythrin/HHE cation-binding motif
Thite_2122241	Peptidase G1, eqolisin	IPR000250	Peptidase G1, eqolisin
Thite_2122261		IPR006662	Thioredoxin-related

Thite_2122285	Glycosyltransferase Family 20 protein	IPR001830	Glycosyl transferase, family 20
Thite_2122411	Hypothetical protein		
Thite_2122499	Hypothetical protein		
Thite_2122548	acetyl esterase	IPR013830	Esterase, SGNH hydrolase-type
Thite_2122588	Hypothetical protein		
Thite_2122633		IPR013027	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
Thite_2122711		IPR000795	Protein synthesis factor, GTP-binding
Thite_2122788		IPR006025	Peptidase M, neutral zinc metallopeptidases, zinc-binding site
Thite_2122859	arabinofuranosidase	IPR000254	Cellulose-binding region, fungal
Thite_2122898		IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_2123310		IPR002048	Calcium-binding EF-hand
Thite_2123329		IPR003829	Pirin, N-terminal
Thite_2123413	Hypothetical protein		
Thite_2123443	Glycoside Hydrolase Family 30 protein	IPR001139	Glycoside hydrolase, family 30
Thite_2123444	Hypothetical protein		
Thite_2123482	Hypothetical protein		
Thite_2123523		IPR009020	Proteinase inhibitor, propeptide
Thite_2123754		IPR013810	Ribosomal protein S5, N-terminal
Thite_2123826	Hypothetical protein		
Thite_2123895		IPR008183	Aldose 1-epimerase
Thite_2124011		IPR006163	Phosphopantetheine-binding
Thite_2124025		IPR008928	Six-hairpin glycosidase-like
Thite_2124045		IPR014045	Protein phosphatase 2C, N-terminal
Thite_2124136		IPR000447	FAD-dependent glycerol-3-phosphate dehydrogenase
Thite_2124394		IPR002125	CMP/dCMP deaminase, zinc-binding
Thite_2124409		IPR017853	Glycoside hydrolase, catalytic core
Thite_2124432		IPR008972	Cupredoxin

Thite_2124512	Hypothetical protein		
Thite_2124570	Glycosyltransferase Family 4 protein	IPR001296	Glycosyl transferase, group 1
Thite_2124610	Hypothetical protein		
Thite_2124616	Hypothetical protein		
Thite_2124654	Glycoside Hydrolase Family 24 protein	IPR002196	Glycoside hydrolase, family 24
Thite_2124679	Hypothetical protein		
Thite_2124693		IPR013126	Heat shock protein 70
Thite_2124714		IPR008972	Cupredoxin
Thite_2124747	Glycosyltransferase Family 15 protein	IPR002685	Glycosyl transferase, family 15
Thite_2124772	mannanase	IPR001547	Glycoside hydrolase, family 5
Thite_2124787	Hypothetical protein		
Thite_2124841		IPR012913	Glucosidase II beta subunit-like
Thite_2124844	alpha-galactosidase	IPR000111	Glycoside hydrolase, clan GH-D
Thite_2125208	Hypothetical protein		
Thite_2125310	Hypothetical protein		
Thite_2125566		IPR002482	Peptidoglycan-binding LysM
Thite_2125607	Hypothetical protein		
Thite_2125664	arabinofuranosidase	IPR003305	Carbohydrate-binding, CenC-like
Thite_2125708		IPR006094	FAD linked oxidase, N-terminal
Thite_2126654	Hypothetical protein		
Thite_2126926	Hypothetical protein		
Thite_2127026	Hypothetical protein		
Thite_2127199		IPR011043	Galactose oxidase/kelch, beta-propeller
Thite_2127356	Hypothetical protein		
Thite_2127426	Hypothetical protein		
Thite_2127786	Hypothetical protein		
Thite_2128280	Hypothetical protein		
Thite_2128319	Carbohydrate-Binding Module Family 18 protein	IPR001002	Chitin-binding, type 1
Thite_2128403		IPR013090	Phospholipase A2, active site
Thite_2128443	Hypothetical protein		
Thite_2128521	Hypothetical protein		
Thite_2128592	Hypothetical protein		
Thite_2129271		IPR001542	Arthropod defensin
Thite_2129356		IPR014756	Immunoglobulin E-set
Thite_2129448	Hypothetical protein		

Thite_2129515	Hypothetical protein		
Thite_2129735		IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_2129788	Hypothetical protein		
Thite_2129842		IPR011058	Cyanovirin-N
Thite_2129895	Hypothetical protein		
Thite_2130104	Hypothetical protein		
Thite_2130213	Hypothetical protein		
Thite_2130234	Hypothetical protein		
Thite_2130237	Hypothetical protein		
Thite_2130248	Hypothetical protein		
Thite_2131275	Hypothetical protein		
Thite_2131280		IPR002925	Dienelactone hydrolase
Thite_2131399	Hypothetical protein		
Thite_2131460	Glycoside Hydrolase Family 43 protein	IPR006710	Glycoside hydrolase, family 43
Thite_2131928		IPR002227	Tyrosinase
Thite_2132235	Hypothetical protein		
Thite_2132746	Hypothetical protein		
Thite_2132777	Hypothetical protein		
Thite_2140512	Hypothetical protein		
Thite_2140739		IPR002347	Glucose/ribitol dehydrogenase
Thite_2141149		IPR002815	Spo11/DNA topoisomerase VI, subunit A
Thite_2141215		IPR002018	Carboxylesterase, type B
Thite_2141217		IPR006094	FAD linked oxidase, N-terminal
Thite_2141256		IPR000782	Beta-Ig-H3/fasciclin
Thite_2141257		IPR001466	Beta-lactamase
Thite_2141725	Hypothetical protein		
Thite_2141802	Hypothetical protein		
Thite_2141856	Hypothetical protein		
Thite_2142309	Peptidase G1, eqolisin	IPR000250	Peptidase G1, eqolisin
Thite_2142353	Peptidase G1, eqolisin	IPR000250	Peptidase G1, eqolisin
Thite_2142617	Hypothetical protein		
Thite_2142623	Hypothetical protein		
Thite_2142658	Hypothetical protein		
Thite_2142691	Hypothetical protein		
Thite_2142693	rhamnogalacturonan acetylesterase	IPR001087	Lipase, GDSL
Thite_2142696	Glycoside Hydrolase Family 61K protein	IPR000254	Cellulose-binding region, fungal
Thite_2142713	Hypothetical protein		
Thite_2143111		IPR001128	Cytochrome P450

Thite_2143262		IPR016864	Uncharacterised conserved protein UCP028035
Thite_2143588	endoglucanase	IPR000334	Glycoside hydrolase, family 45
Thite_2143700		IPR001189	Manganese and iron superoxide dismutase
Thite_2144067		IPR007312	Phosphoesterase
Thite_2144616	Peptidase S10, serine carboxypeptidase	IPR001563	Peptidase S10, serine carboxypeptidase
Thite_2144678		IPR013094	Alpha/beta hydrolase fold-3
Thite_2144687	Hypothetical protein		
Thite_2144766	acetylxylan esterase	IPR000675	Cutinase
Thite_2144774	Peptidase S10, serine carboxypeptidase	IPR001563	Peptidase S10, serine carboxypeptidase
Thite_2144808	Hypothetical protein		
Thite_2144836		IPR000172	Glucose-methanol-choline oxidoreductase, N-terminal
Thite_2144923		IPR005065	Platelet-activating factor acetylhydrolase, plasma/intracellular isoform II
Thite_2145295	Glycoside Hydrolase Family 43 protein	IPR006710	Glycoside hydrolase, family 43
Thite_2145367	Hypothetical protein		
Thite_2145487		IPR006076	FAD dependent oxidoreductase
Thite_2145534	Hypothetical protein		
Thite_2145578	Hypothetical protein		
Thite_2145591	beta-1,6-galactanase	IPR001547	Glycoside hydrolase, family 5
Thite_2145623	Hypothetical protein		
Thite_2145703		IPR003042	Aromatic-ring hydroxylase
Thite_2146087	Hypothetical protein		
Thite_2146577	Hypothetical protein		
Thite_2146687	Hypothetical protein		
Thite_2146815		IPR013078	Phosphoglycerate mutase
Thite_2147249	Hypothetical protein		
Thite_2147511	Hypothetical protein		
Thite_2147701		IPR002018	Carboxylesterase, type B
Thite_2148402	Hypothetical protein		
Thite_2148619	Hypothetical protein		
Thite_2149883	xylanase	IPR001137	Glycoside hydrolase, family 11

Thite_2151024		IPR000246	Peptidase T2, asparaginase 2
Thite_2152124	Hypothetical protein		
Thite_2152403		IPR000101	Gamma-glutamyltranspeptidase
Thite_2153165		IPR002347	Glucose/ribitol dehydrogenase
Thite_2153770		IPR002403	Cytochrome P450, E-class, group IV
Thite_2155501	Peptidase A1; aspartic protease	IPR001461	Peptidase A1
Thite_2169147		IPR001938	Thaumatin, pathogenesis-related
Thite_2169463	Hypothetical protein		
Thite_2169612		IPR001451	Bacterial transferase hexapeptide repeat
Thite_2169677		IPR015909	Protein Transporter, Pam16/TIM14
Thite_2169795	Hypothetical protein		
Thite_2170138		IPR015590	Aldehyde dehydrogenase
Thite_2170323	Peptidase C13; putative GPI-anchor transamidase	IPR001096	Peptidase C13, legumain
Thite_2170564		IPR008701	Necrosis inducing
Thite_2170708	Hypothetical protein		
Thite_2170732	Hypothetical protein		
Thite_2170913	Hypothetical protein		
Thite_2171055	Hypothetical protein		
Thite_2171281	Hypothetical protein		
Thite_2171393	Hypothetical protein		
Thite_2171499		IPR008427	Extracellular membrane protein, 8-cysteine region, CFEM
Thite_2171619	Hypothetical protein		
Thite_2171711		IPR007921	Cysteine, histidine-dependent amidohydrolase/peptidase
Thite_24856		IPR008263	Glycoside hydrolase, family 16, active site
Thite_27104	Hypothetical protein		
Thite_32558	Hypothetical protein		
Thite_33101	Hypothetical protein		
Thite_33488		IPR001117	Multicopper oxidase, type 1
Thite_34021	Hypothetical protein		
Thite_34437	pectin methyl esterase	IPR000070	Pectinesterase, catalytic

Thite_34523	rhamnogalacturonan lyase	IPR013784	Carbohydrate-binding-like fold
Thite_34861	Hypothetical protein		
Thite_35493	chitinase	IPR000254	Cellulose-binding region, fungal
Thite_38252		IPR006094	FAD linked oxidase, N-terminal
Thite_38647		IPR007312	Phosphoesterase
Thite_41497	Hypothetical protein		
Thite_42761		IPR002198	Short-chain dehydrogenase/reductase SDR
Thite_43138	Hypothetical protein		
Thite_43665	Glycoside Hydrolase Family 61U protein	IPR005103	Glycoside hydrolase, family 61
Thite_43879		IPR001128	Cytochrome P450
Thite_45518	Hypothetical protein		
Thite_46100	chitinase	IPR001002	Chitin-binding, type 1
Thite_46900	chitosanase	IPR009939	Fungal chitosanase
Thite_47394	Peptidase S8 and S53, subtilisin, kexin, sedolisin	IPR000209	Peptidase S8 and S53, subtilisin, kexin, sedolisin
Thite_47532		IPR014030	Beta-ketoacyl synthase, N-terminal
Thite_48148		IPR008313	Uncharacterised conserved protein UCP028846
Thite_48486	xylanase	IPR001000	Glycoside hydrolase, family 10
Thite_48594		IPR001117	Multicopper oxidase, type 1
Thite_49333	Hypothetical protein		
Thite_49921		IPR002018	Carboxylesterase, type B
Thite_50690	glucoamylase	IPR002044	Glycoside hydrolase, carbohydrate-binding
Thite_52337		IPR002347	Glucose/ribitol dehydrogenase
Thite_52349		IPR015366	Peptidase S53, propeptide
Thite_52670		IPR000073	Alpha/beta hydrolase fold-1
Thite_52693	chitinase	IPR001579	Glycoside hydrolase, chitinase active site
Thite_61645		IPR013027	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
Thite_62739		IPR002227	Tyrosinase
Thite_66812	Hypothetical protein		
Thite_71965		IPR002921	Lipase, class 3

Thite_72454		IPR000627	Intradiol ring-cleavage dioxygenase, C-terminal
Thite_73074		IPR001087	Lipase, GDSL
Thite_73871		IPR001148	Carbonic anhydrase, eukaryotic
Thite_74712	cellobiohydrolase	IPR001524	Glycoside hydrolase, family 6
Thite_76240	rhamnogalacturonase	IPR000254	Cellulose-binding region, fungal

extracellular proteins of *T. terrestris* .
xpression levels in glucose, alfalfa straw
ass spectrometry are indicated by

CAZy_module(s)	Exo-proteome	Alfalfa_RPKM	Barley_RPKM	Glucose_RPKM
	identified	1.96	2.69	0.28
GH7-CBM1	**identified**	960.69	1211.96	6.90
	identified	0.72	1.16	2.20
GH37	**identified**	236.40	91.45	103.91
	identified	96.19	30.96	20.82
GH7-CBM1	**identified**	315.35	222.21	2.02
	identified	2.53	1.39	1.10
CE5-CBM1	**identified**	38.25	127.39	0.17
GH61	**identified**	7.60	9.91	3.51
	identified	2.90	1.70	0.97
GH114	**identified**	68.46	51.10	58.15
	identified	22.11	252.52	0.50
	identified	9.10	14.59	23.07
	identified	473.46	140.30	207.07
	identified	811.07	516.78	174.69
	identified	410.11	658.79	316.89
GH47	**identified**	9.39	5.78	5.00
GH74-CBM1	**identified**	142.20	19.66	4.02
	identified	27.10	39.12	22.75
	identified	14.60	16.47	10.19
GH61	**identified**	1.37	9.56	0.90
	identified	37.36	61.87	33.20

GH18-CBM50	**identified**	821.37	1309.71	57.52
GH53	**identified**	23.00	14.19	4.64
GH2	**identified**	29.95	56.19	1.19
GH61-CBM1	**identified**	146.76	205.60	0.00
	identified	382.38	683.00	308.45
GH10-CBM1	**identified**	158.55	377.89	0.63
	identified	52.32	16.32	3.00
CBM1	**identified**	123.77	100.58	0.41
	identified	225.32	415.96	175.96
CE15	**identified**	18.25	5.34	1.42
GH115	**identified**	87.39	101.75	4.44
CBM1-CE16	**identified**	184.06	47.81	1.00
GH10	**identified**	49.78	304.25	1.02
	identified	803.89	307.45	258.02
GH61-CBM1	**identified**	81.83	99.11	0.14
GH3	**identified**	45.29	32.36	58.60
GH20	**identified**	38.11	28.91	47.36
	identified	5.99	5.69	4.68
	identified	0.91	10.91	0.91
GH43	**identified**	20.67	20.20	13.72
GH61	**identified**	72.87	416.18	0.14
	identified	21.55	27.69	13.91
CBM1	**identified**	80.99	50.81	6.66
	identified	0.47	37.60	0.00

	identified	5.27	26.29	0.28
GH16	**identified**	433.75	476.01	387.09
	identified	26.36	46.36	24.28
	identified	285.54	510.28	325.49
GH61-CBM1	**identified**	218.69	169.64	14.08
GH18	**identified**	2.80	4.76	0.96
GH31	**identified**	2.54	2.16	0.06
GH7	**identified**	18.39	68.08	0.11
CBM1	**identified**	269.16	126.45	0.96
CBM1-GH5	**identified**	541.83	151.86	3.67
		0.09	0.75	0.34
		2.25	3.00	0.31
		1.00	0.92	2.13
		0.61	0.70	0.49
		3.15	2.49	3.07
		0.09	0.13	0.00
		5.89	5.31	0.97
		35.43	21.24	29.75
		3.85	9.74	0.34
		1.04	1.58	1.51
		0.69	1.11	0.21
		0.19	0.14	0.17
		4.60	4.57	4.21
		0.21	0.31	0.81
		0.00	0.14	0.00
GH61		28.30	16.86	11.52
		2.52	2.16	4.87
		4.68	5.59	2.68
		0.29	0.59	2.33
		0.22	0.27	0.31
		0.00	0.07	0.00
		0.79	7.99	0.00
		0.05	0.17	0.00
		0.33	0.57	0.19
GH5		2.86	3.72	7.30
		0.00	0.00	0.00

		0.37	1.49	0.30
GH55		0.83	0.77	1.83
		2.80	3.03	3.48
		11.65	7.75	5.97
		0.32	0.43	1.48
GH10		0.17	0.14	0.00
		9.72	11.12	1.53
CBM50-CBM50-CBM50-CBM50		0.08	0.04	0.24
		0.09	0.00	0.11
		0.12	0.13	0.00
GH28		36.68	7.92	0.70
		0.03	0.18	0.10
		6.87	1.31	5.39
		0.06	0.15	0.25
		1.99	0.94	0.00
CE8		13.32	2.11	0.26
		13.26	15.14	16.48
CBM1-GH5		52.08	82.70	1.60
		0.00	0.15	0.12
		0.00	0.00	0.00
		0.10	0.05	0.00
		1.27	1.26	0.45
		0.04	0.00	0.00
		0.02	0.00	0.31
GH2		1.32	0.63	0.53
GH5		1.84	3.66	1.50
		0.11	0.13	0.00

		0.03	0.03	0.00
		24.28	13.62	10.17
		0.48	0.88	0.56
GH3		3.70	2.00	1.74
		18.52	7.30	17.36
		0.70	1.28	0.00
GH55		0.02	0.00	0.00
		0.00	0.05	0.00
		22.13	26.55	1.52
		0.16	0.52	0.09
		618.28	585.81	557.67
CBM1-GH6		654.24	658.90	3.11
		0.68	0.76	0.87
GH61		1.94	27.81	0.00
		9.47	8.65	6.29
		4.40	14.70	0.28
		3.40	3.34	4.91
GH7		0.55	0.50	0.70
		0.19	0.96	0.00
		0.11	0.04	0.14
		0.02	0.02	0.00
		0.77	0.90	6.90
GH13-CBM20		7.24	6.91	4.19
		0.06	0.56	0.00
		1.67	1.88	2.03
CBM50-CBM50-CBM18-GH18		0.04	0.14	0.11
		0.11	1.05	0.13
		0.18	0.22	0.00
		0.55	0.07	0.00
		0.64	0.53	1.45
		4.38	5.76	4.34
GH28		0.83	1.10	0.23

GH63		3.88	2.68	3.49
		1.07	1.21	1.86
		0.46	0.21	0.00
CBM18		0.25	0.33	0.43
		6.56	6.02	1.02
		0.12	0.00	0.74
		7.40	29.30	0.44
		0.04	0.04	0.12
		2.26	1.94	1.53
		0.12	0.00	0.00
		0.62	0.18	0.00
		0.18	0.58	0.31
		18.80	22.85	18.74
		1.09	2.08	1.28
		0.26	0.15	0.46
		0.81	0.48	0.62
		1.07	2.32	0.87
CBM18-CBM18-GH18		0.07	0.03	0.04
GH15-CBM20		5.89	1.68	0.80
		1.89	3.27	0.25
GH28		87.74	54.19	3.67
		1.33	0.76	2.04
GH61		52.26	2.81	0.85
		56.85	83.30	61.35
		40.91	77.35	18.19
		0.15	0.00	0.00
		0.23	0.72	0.25
		0.28	0.00	0.14
		0.69	0.80	0.22

		30.25	15.59	26.74
		0.46	0.22	0.22
		0.75	3.31	0.27
		2.69	2.93	2.41
		1.39	0.88	0.66
		7.64	3.57	6.30
GH28		4.49	4.07	5.16
		1.42	0.52	0.76
		0.33	0.21	0.27
		0.74	0.66	1.91
		0.17	0.24	0.64
		5.80	6.11	5.84
		11.23	9.13	1.11
		0.83	0.99	1.58
GH61		0.07	0.00	0.00
		0.03	0.00	0.00
		0.61	0.63	1.18
		1.44	2.97	1.15
		2.18	2.30	2.27
GH43		6.46	4.09	6.53
		0.85	1.49	0.29
		2.02	1.95	4.08
GH54-CBM42		1.82	2.61	0.00
		0.13	0.19	0.31
		0.05	0.11	0.48
GH11		0.28	1.33	0.00
		0.29	0.34	0.30
CE3		1.06	2.01	1.09
GH16-CBM1		0.19	0.07	0.00
		3.17	42.77	0.69
GH61		2.18	1.10	3.55
		0.32	0.73	2.06

		0.48	0.31	0.78
GH3		3.69	0.64	0.42
GH35		1.49	0.66	0.78
		0.38	0.32	0.34
PL4		1.35	0.50	0.09
		7.23	8.23	5.95
GH2		0.04	0.00	0.11
		2.52	1.86	3.13
		1.30	2.89	1.52
		0.63	0.24	0.90
GH35		0.96	0.62	0.21
		8.96	34.44	0.00
		5.25	7.57	5.46
		0.19	0.23	0.00
		1.72	3.19	1.53
		0.15	0.33	0.00
		5.00	5.47	3.82
GH2		1.04	0.30	0.56
		3.48	8.01	0.27
		42.60	65.96	2.39
GH11		0.00	0.06	0.00
		1.22	1.12	0.71
		0.06	0.31	0.41
		0.56	0.99	0.84
		0.07	0.28	0.94
		0.10	0.05	0.08
		0.79	1.23	1.17
		0.90	1.54	2.78
		0.17	0.08	0.00

		0.18	0.51	0.19
		0.10	0.10	0.00
PL7		0.00	0.00	0.00
CBM18-CBM18-GH18		0.18	0.35	0.10
		1.22	1.21	1.16
GH61		0.13	0.11	0.00
		2.42	2.69	3.34
		0.59	1.93	0.16
		0.40	0.48	0.86
		2.02	1.21	0.00
CE15		0.47	0.78	0.49
		0.11	0.10	0.00
		0.35	0.34	0.34
		0.00	0.26	0.00
		14.72	5.95	11.92
		0.10	0.81	0.72
		0.79	2.60	0.12
CE3		5.06	2.32	8.69
		126.15	149.53	101.95
		0.16	0.18	0.82
		2.29	2.67	0.80
		0.08	1.12	0.00
		0.00	0.00	0.00
		0.07	0.38	0.00
GH61		0.03	0.42	0.00
GH16		365.09	222.83	147.19
		18.37	22.19	15.26
		64.15	88.31	56.53
		34.54	42.51	28.45

		0.28	0.22	0.18
		1.58	2.78	1.83
		158.04	376.64	128.49
		129.29	254.55	100.80
		35.60	51.12	20.63
		69.53	68.82	75.02
		0.75	0.94	0.92
		1.06	1.89	3.35
		0.10	0.20	0.65
		0.10	0.00	0.20
		0.18	0.00	0.00
		0.89	0.52	0.52
		0.58	0.39	0.24
		0.16	3.26	0.00
		0.07	0.19	0.00
		0.16	0.19	0.56
		0.34	7.49	0.00
		0.02	0.00	0.00
GH61		0.75	0.60	1.90
		0.07	0.07	0.00
		0.33	0.44	0.00
		0.21	0.44	0.15
		0.64	0.00	0.00
		0.02	0.23	0.00
		0.04	0.14	0.00
CBM50-CBM18		0.07	0.59	0.00
		0.16	0.43	0.00
		3.16	2.36	3.44
		0.67	0.27	0.74
		1.62	1.16	1.23
		0.04	0.16	0.00
		1.11	16.43	0.31
		2.49	0.79	2.26
		2.89	2.60	0.92
		0.16	0.26	0.00
		5.55	19.97	0.41

GT1		6.88	1.76	2.70
		0.24	0.19	0.39
		0.18	0.20	0.27
		0.41	0.77	0.47
		1.26	1.47	2.24
		0.29	0.48	0.17
		4.19	5.29	1.79
		0.63	0.59	0.47
GT32		47.27	32.55	78.97
		22.31	31.33	17.81
		316.86	395.08	314.92
		1.12	1.62	1.34
		22.61	11.88	22.24
		19.72	27.76	9.38
		28.97	296.63	3.49
		5.73	8.08	12.74
		9.89	15.16	9.94
		22.49	16.72	29.43
		17.16	12.32	15.00
		42.37	78.17	9.60
		54.37	25.71	56.11
GH61		2.89	3.86	2.40
		30.34	6.73	21.50
		7.10	12.54	10.45
		10.17	18.23	10.90
		179.09	206.54	165.07
		87.51	38.28	26.92
GH3		9.60	8.00	3.22
		29.53	37.29	39.44
		17.72	33.08	18.37
GH31		88.20	64.62	96.67
		27.45	12.33	21.78
		39.24	56.88	36.39
		7.72	10.47	14.36

		39.18	60.82	45.40
		173.98	243.18	163.35
		4.10	2.51	2.25
		6.03	13.37	7.67
GH5		6.53	2.83	0.63
		53.26	32.10	14.21
		14.63	43.55	1.95
		9.85	28.05	4.51
GH11		2.03	30.17	0.25
		225.57	243.52	210.55
		0.80	3.19	1.55
		9.81	11.57	17.31
GH93		1.77	89.86	0.73
GH47		26.37	25.23	28.42
		20.44	40.01	120.65
GH11-CBM1		88.42	684.52	0.97
		42.63	75.35	2.56
		3.93	6.23	0.10
		12.73	5.11	8.19
		112.69	279.34	211.50
		13.54	33.01	10.74
		60.69	32.16	55.77
		242.60	239.71	310.85
GH63		49.69	51.99	52.30
		17.60	27.22	9.39
GH10		10.92	3.39	5.25
		66.80	22.92	45.50

		10.58	15.71	21.93
		4.70	5.39	11.16
		43.63	20.60	34.54
		9.67	8.84	10.60
		21.37	24.81	0.14
		22.73	12.48	0.91
GH2		21.22	24.08	0.27
		8.30	7.08	2.82
GH55		111.51	54.88	25.43
		4.21	3.42	7.52
GH25		4.93	12.19	4.51
		29.18	38.04	27.77
		17.57	15.47	18.75
		10.39	5.05	7.27
		7.05	3.81	3.34
GH81		8.99	9.43	5.87
		4.70	6.70	1.59
		7.25	12.26	12.67
		1.37	2.41	1.79
		0.73	0.90	3.48
		21.80	17.77	23.69
		4.90	28.04	9.50
		8.13	13.17	2.01
		1.48	2.10	3.26
		190.65	213.17	195.38
		336.47	441.49	449.02
		4.10	3.49	5.01
		2.41	10.12	0.55
		32.61	116.03	12.20
GT32		53.50	67.50	52.97

		7.37	58.20	11.65
GT69		4.86	5.09	6.81
GT25		2.88	10.33	0.39
GH62		12.66	172.58	0.37
		1.40	7.27	2.11
GH45-CBM1		71.65	139.82	2.25
		2.60	4.52	2.59
		15.55	10.48	5.76
		94.12	109.76	3.10
		3.01	3.49	2.45
		4.42	2.84	6.71
		63.87	98.40	94.81
CBM18-CBM18-CBM18		2.79	8.09	1.04
		3.25	23.87	7.27
		143.32	209.43	80.61
CBM1-CE2		18.43	2.20	0.89
		27.42	29.37	40.70
		22.94	35.03	20.72
		58.76	35.17	12.36
		19.47	8.20	8.54
		4.35	4.16	23.14
GH47		9.97	13.49	16.56
GT69		9.83	7.79	9.31
		16.48	20.63	15.05
		15.15	14.61	15.96
		7.50	16.10	12.89
		9.78	10.68	20.95
		38.00	12.98	13.21
		60.81	95.31	74.40
		26.93	21.30	17.98
		42.45	48.61	53.47
		139.28	655.13	4.61
		34.35	41.45	32.05
		5.34	11.51	4.23

		15.76	20.77	9.27
		78.32	64.29	25.97
		1.43	1.79	3.93
		1.09	1.39	15.14
GH16		3.02	3.79	6.01
		8.17	4.98	5.54
		86.43	51.53	91.85
		4.48	3.96	7.74
		45.82	14.82	38.54
		1.01	0.76	10.34
		2.56	1.34	1.64
		7.01	4.91	10.11
		4.84	10.90	12.76
		32.21	37.42	30.34
		24.43	28.41	26.35
		128.70	114.18	204.14
GT17		28.86	15.81	24.42
		77.92	180.27	7.77
		2.43	3.37	2.95
		20.89	15.34	11.35
GH16		3.35	2.16	1.05
GH5		5.58	4.86	2.37
GT69		27.69	35.43	33.89
		179.59	362.88	34.11
		37.91	55.68	24.81
		26.89	20.03	18.56
CE3		30.48	14.86	57.69
GH28		1.54	1.38	2.18
		26.15	31.74	24.60
		37.92	98.94	24.52

GT32		38.12	37.99	32.77
GH92		16.79	8.15	6.07
		412.23	652.17	107.78
		11.84	28.20	3.93
		20.99	21.08	11.64
		12.12	11.80	12.29
		68.77	33.62	62.84
		51.70	50.94	40.37
		5.65	9.35	3.01
		3.73	5.05	5.23
		48.66	39.78	47.04
		53.54	26.03	53.95
		2.30	5.20	8.76
		14.89	15.34	20.32
GH39		2.41	9.12	0.39
		320.91	489.98	238.48
		2.49	3.93	3.78
		3.03	2.72	6.52
		13.95	17.55	13.72
		16.10	21.69	16.92
		9.10	25.47	3.61
GT71		10.00	1.62	9.84
		580.89	268.52	43.76
		10.55	27.65	1.88
		3.62	7.66	0.34
		1.38	2.20	0.80
		2.35	3.85	1.94

		8.32	4.17	2.34
GH93		5.47	21.50	2.53
		9.60	10.11	24.71
CE8-GH28		24.80	7.99	0.57
		13.91	2.99	16.15
		11.10	32.30	1.39
GH61		42.57	29.29	0.69
		15.26	11.44	2.21
GT24		34.11	26.51	31.97
		2.29	1.53	2.83
		13.53	8.40	19.47
		56.79	89.67	59.39
		193.63	317.31	167.12
GH18		6.36	8.96	7.14
		42.78	28.47	18.40
		8.85	16.99	4.77
		46.47	59.18	18.31
		3.62	3.72	6.89
		43.02	69.51	34.55
		6.58	5.27	8.88
GT31		16.64	19.88	26.78
		2.69	6.46	7.09
		7.73	20.78	4.22
		1.16	1.67	1.12
CBM1-GH5		81.91	10.93	0.00
		16.44	25.31	13.30
		69.98	94.24	60.44
GH47		106.10	94.71	73.80

		19.05	33.71	25.06
		31.75	34.27	5.43
		12.33	13.75	6.33
GH30		12.89	5.00	3.16
GH10-CBM1		3.98	118.23	0.37
		2.41	3.73	3.19
GH12		7.21	48.11	0.40
CE1		8.64	312.10	0.00
		2.81	97.02	26.34
		43.56	96.54	0.79
		151.74	113.51	173.22
CE4		18.48	78.41	2.03
		9.04	7.45	9.82
		11.10	17.92	4.96
GT90		26.73	28.28	26.79
GT31		82.73	114.36	104.53
		11.02	76.56	1.27
		75.23	24.12	31.02
		10.56	21.10	0.33
		1.80	6.28	0.77
GH43		15.18	23.86	0.22
GH3		32.53	15.99	0.59
		20.80	7.86	11.33
		61.49	78.65	294.56
		282.95	994.56	1.10
GH55		24.86	19.37	43.73
		1.82	2.70	3.80
GT25		23.81	11.34	26.45
		85.64	187.95	64.11
		17.36	16.42	5.35
		118.89	77.16	136.44

		25.95	33.89	22.98
GH16		5.75	6.49	3.81
		0.84	0.86	2.97
		12.38	19.59	17.57
		121.02	65.58	68.45
		14.12	21.64	24.04
		14.15	16.89	16.00
CBM18-CE4		1.06	3.48	1.60
		37.22	18.48	20.28
GH76		10.63	10.43	19.18
		477.56	795.25	66.36
		78.07	45.97	47.83
		52.07	26.15	71.15
		1.04	1.25	4.50
GT90		31.70	27.60	43.85
		45.33	18.46	40.66
		80.82	25.76	37.29
		78.94	85.71	88.81
GH5		12.56	11.11	5.40
		106.94	128.14	173.95
CBM52		16.01	33.40	8.13
		29.02	32.64	22.53
CE4		16.57	17.55	15.38
CE5		17.91	108.12	1.24
		26.39	162.05	0.97
		2.00	15.32	1.22
		14.31	40.07	5.86

		14.32	8.57	55.95
		27.71	41.02	26.79
		1.39	1.44	3.56
		0.60	2.50	6.12
		9.11	14.42	12.68
GH76		10.11	2.37	10.22
GH79		27.49	23.26	55.75
		6.13	9.06	10.69
		25.34	8.87	18.03
		182.01	212.13	203.79
GH5		36.92	20.16	16.98
		71.84	115.58	87.29
		11.24	17.33	10.24
		37.35	42.15	24.64
		18.38	30.14	16.22
		39.01	78.65	38.01
		4.24	3.60	9.90
		17.56	11.28	9.50
		45.30	69.81	109.86
		33.68	23.36	20.14
		1.51	6.29	0.45
		3.77	3.68	7.45
		65.17	84.48	40.00
		61.10	78.93	140.71
GH18		0.29	0.56	1.76
		313.79	398.56	236.20
		273.01	261.57	115.27
		14.78	16.89	6.84
		51.39	80.34	36.47
		93.53	117.72	25.73
		118.66	139.33	85.96

GT20		43.01	49.99	16.87
		1.37	2.47	0.31
		27.57	36.30	24.74
CE16		109.83	407.61	2.99
		1.92	2.10	3.21
		9.64	5.93	11.72
		30.97	7.35	30.06
		154.16	103.93	87.29
GH62-CBM1		22.07	62.37	7.37
		13.02	25.23	0.49
		47.06	52.83	47.77
		17.32	24.99	25.85
		4.40	12.70	0.00
GH30		5.70	35.52	1.11
		10.60	5.09	4.97
		0.00	0.00	3.53
		77.74	122.78	78.69
		166.78	40.40	126.53
		11.64	8.61	17.14
		38.06	110.82	35.13
		11.05	3.45	14.04
		9.27	4.73	7.17
		51.98	16.72	43.85
		131.25	38.26	38.49
		44.44	58.45	37.73
		533.57	463.56	740.31
		65.61	49.83	99.12

		9.48	1.00	7.91
GT4		23.28	12.80	22.07
		231.57	140.95	106.28
		8.25	9.87	9.79
GH24		9.16	17.29	5.43
		1.53	1.19	1.34
		84.19	49.65	51.86
		15.26	28.31	26.79
GT15		16.71	7.42	12.30
GH5		30.82	30.24	1.80
		3.19	5.18	1.44
		35.10	22.22	32.79
GH27		6.35	11.70	7.07
		2.11	0.81	3.96
		0.24	1.02	1.00
		2.61	0.89	1.72
		0.22	0.00	0.00
GH51		5.79	139.17	0.24
		0.60	0.34	0.00
		3.41	4.16	7.00
		0.04	0.34	0.26
		0.65	0.62	0.48
		0.19	0.14	0.37
		1.64	0.67	1.61
		0.53	0.90	0.78
		0.27	0.11	0.00
		3.46	10.48	1.76
CBM18		0.06	0.40	0.39
		0.78	2.44	0.20
		0.00	0.00	0.69
		0.61	0.30	0.29
		0.00	0.00	0.00
		5.13	4.38	1.40
		2.45	5.52	0.49
		0.00	0.00	0.00

		0.00	0.00	0.00
		0.67	0.33	0.17
		11.63	9.92	4.02
		1.68	1.55	4.48
		0.00	0.08	0.00
		0.94	5.14	0.14
		0.08	0.42	0.00
		0.26	0.51	0.00
		0.20	0.00	0.00
		0.08	0.11	0.10
		0.05	0.25	0.17
		0.25	2.14	0.20
		2.83	12.28	2.31
GH43		0.40	0.15	0.46
		5.55	5.06	12.99
		0.14	0.63	0.46
		1.16	2.08	0.91
		0.65	0.25	1.94
		4.58	14.91	5.01
		4.03	2.40	5.94
		3.42	4.09	2.93
		0.55	0.41	0.40
		17.14	5.06	9.91
		0.04	0.00	0.13
		0.06	0.49	0.16
		0.17	0.44	0.00
		0.13	0.00	0.28
		6.04	5.96	4.62
		4.00	3.83	4.51
		1.63	4.94	1.29
		0.00	0.08	0.00
		0.07	0.11	0.00
		0.93	1.52	3.82
		1.00	1.97	0.70
CE12		2.17	0.43	0.78
GH61-CBM1		0.46	0.61	0.70
		0.00	0.00	0.00
		0.05	0.15	0.00

		0.19	0.03	0.33
GH45		2.65	1.87	4.93
		176.23	134.51	140.30
		2.77	2.76	0.83
		2.13	1.07	2.27
		1.44	2.35	0.00
		0.43	1.04	0.20
CE5		0.32	12.68	0.00
		0.15	0.11	0.26
		0.04	0.02	0.07
		0.19	0.17	0.00
		1.23	1.04	0.26
GH43-CBM35		3.77	3.92	2.43
		3.39	3.86	2.39
		13.75	14.91	10.65
		1.33	1.48	0.24
		1.55	1.13	2.75
GH5		0.22	0.38	0.48
		3.23	1.87	3.31
		0.44	0.52	0.24
		0.47	0.34	1.58
		0.31	0.38	0.15
		1.16	1.50	0.89
		1.70	3.37	0.44
		1069.96	483.92	550.43
		0.55	0.52	2.05
		0.00	0.00	0.00
		1.97	1.63	0.77
		11.67	1.95	6.99
GH11		176.85	5087.01	1.56

		7.77	3.55	5.53
		50.16	58.05	25.63
		45.13	45.97	41.30
		106.14	82.36	70.85
		0.10	0.04	0.08
		10.98	4.93	22.02
		1112.85	1345.17	1507.65
		17.49	36.21	50.73
		268.18	377.48	313.42
		95.98	111.95	72.84
		111.23	103.21	61.21
		40.54	46.86	40.46
		74.79	86.98	79.61
		2.19	2.61	0.67
		44.87	166.19	8.67
		237.79	272.26	146.31
		91.84	110.94	104.87
		209.89	383.71	48.11
		604.86	744.32	911.20
		782.73	958.52	908.83
		15.65	39.14	178.93
		389.10	79.82	105.50
		116.71	149.77	94.15
		0.97	1.90	0.75
		279.39	282.06	261.83
		0.00	0.00	0.00
		0.00	0.12	0.00
		1.08	0.72	1.06
		2.98	3.81	3.70
CE8		39.62	5.59	0.46

PL4		4.55	2.90	0.45
		3.83	3.76	1.67
GH18-CBM1		0.37	0.62	0.26
		19.51	15.39	9.10
		2.70	4.56	0.11
		1.47	3.32	2.07
		0.00	0.06	0.18
		0.26	0.64	0.00
GH61		0.08	0.22	0.00
		0.56	0.62	0.10
		0.09	0.21	0.15
CBM50-CBM50-CBM18-GH18		178.72	60.53	3.60
GH75		0.29	0.44	0.50
		1.48	0.83	0.26
		0.20	0.30	0.02
		2.12	1.12	0.58
GH10		1.55	5.74	0.31
		0.77	0.91	0.72
		0.17	1.62	0.15
		0.14	0.29	0.10
GH15-CBM20		3.33	1.48	1.95
		1.22	1.66	2.73
		0.93	2.67	1.28
		22.83	13.98	25.81
GH18		2.21	0.97	1.18
		543.53	388.59	375.99
		5.26	8.55	2.89
		0.34	0.65	1.43
		35.28	48.46	26.22

		14.29	1.91	0.80
		23.70	19.03	1.48
		4.36	5.75	14.83
GH6		2.23	2.43	3.69
GH28-CBM1		26.34	16.07	0.19