Genome sequencing of thermofilic fungus *Humicola grisea* var. *thermoidea* and effect of pH in sugarcane biomass transformation

Andrei Stecca Steindorff[1], Eduardo Fernandes Formighieri[2], Fabrícia Paula de Faria[3], Marcio José Poças-Fonseca[4], João Ricardo Moreira de Almeida[5]

Abstract

The thermophilic fungus *Humicola grisea* var. *thermoidea* produces an efficient hydrolytic system with a wide range of thermostable carbohydrate active enzymes (CAZy). In this respect, it represents a promising agent for agricultural residue bioconversion. It can also be employed as a gene donor for the heterologous production of industrial enzymes and to complement enzymatic cocktails. In this work, we sequenced and annotated the genome from *Humicola grisea* var. thermoidea and analyzed the transcriptome of this fungus grown in different pH during sugarcane bagasse degradation normalized with glucose as sole carbon source. Genome features corresponds to what we expect for a thermofilic *Sordariomycetes.* In our transcriptomic data, categories related to primary metabolism and polysaccharide deconstruction, especially hydrolases, were significantly up-regulated in pH 5. In pH 8, up-regulated categories were similar to pH 5 with an addition of kinase category, on the other hand, down-regulated categories show a larger diversity of functions – RNA metabolism, transmembrane transporter, oxidoreductase and electron carrier. In conclusion, our data reveals that pH has a great importance in biomass transformation modulating several pathways and will provide support for several biotechnological applications.

Keywords: *Humicola grisea*. transcriptome. genome sequencing. sugarcane bagasse. pH.

Introduction

The thermophilic fungus *Humicola grisea* var. *thermoidea* produces an efficient hydrolytic system with a wide range of thermostable carbohydrate active enzymes (CAZy). In this respect, it represents a promising agent for agricultural residue

[1] Biólogo, Doutor em Biologia Molecular, Consultor na Embrapa Agroenergia, andrei.steindorff@colaborador.embrapa.br
[2] Agrônomo, Doutor em Biologia Funcional e Molecular, Pesquisador da Embrapa Agroenergia, eduardo.formighieri@embrapa.br
[3] Bióloga, Doutora em Biologia Molecular, Professora da Universidade Federal de Goiás, fabriciapfaria@hotmail.com
[4] Biólogo, Doutor em Biologia Molecular, Professor da Universidade de Brasília, mpossas@unb.br
[5] Biólogo, Doutor em Microbiologia Aplicada, Pesquisador da Embrapa Agroenergia, joao.almeida@embrapa.br

bioconversion. It can also be employed as a gene donor for the heterologous production of industrial enzymes and to complement enzymatic cocktails (MELLO-DE-SOUZA et al., 2011).

So far, no genome from *Humicola* genus has been published and relatively few genes encoding *Humicola grisea* var. *thermoidea* have already been identified and expressed in heterologous systems. Despite the biotechnological potential of *H. grisea*, little is known about the regulation of its gene expression. In 2011 Mello-de-Sousa and collaborators demonstrated that *H. grisea* cbh1.2 gene expression, as well as the expression of several other genes of this fungal cellulolytic system, is regulated through a crosstalk between carbon source- and pH-dependent regulatory mechanisms.

In this work we sequenced and annotated the *Humicola grisea* var. Thermoidea genomeand analyzed the transcriptome of this fungus grown in different pH during sugarcane bagasse degradation normalized with glucose.

Material and Methods

Pre-cultivation

H. grisea was maintained at 42 °C on 4.0% (w/v) oatmeal (Quaker) solid medium. For mycelium obtainment, 10^6 spores/mL were inoculated in 50 mL of Pontecorvo's minimal medium (MM), at pH 6.8 (non-buffered), enriched with 0.25% (w/v) yeast extract, 0.1% (w/v) peptone and supplemented with 1% (w/v) glucose. Incubation proceeded for 24 h (42 °C/120 rpm).

DNA isolation

Harvested mycelia from pre-cultivation step were immediately ground in liquid nitrogen to a fine powder. DNA was isolated using the DNAzol reagent (Invitrogen) according to manufacturer's instructions.

Cultivation and RNA isolation

For the transcriptome experiment, grown mycelium was filtered, washed and transferred to fresh 50 mL MM, supplemented with 1% (w/v) of glucose or 0.1% ball-milled, steam-exploded sugarcane bagasse (SCB) as carbon sources. Culture media pH was adjusted to 5.0 or to 8.0 (buffered with 100 mM sodium citrate). Cultures were incubated for 6 and 12 h (42°C 120 rpm). Three biological replicates were made for each culture condition. Harvested mycelia were immediately ground in liquid nitrogen to a fine powder. RNA was isolated using Trizol reagent (Invitrogen) according to manufacturer's instructions.

Genome assembly and annotation

H. grisea genomic DNA (gDNA) was isolated and sent for sequencing through two strategies: short inserts (Standard Illumina Hiseq2000 paired-end 2x150bp) and mate-pairs (Illumina Hiseq2000 2x100bp with insert size of 3 Kb). FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to evaluate the libraries quality before and after trimming. For quality trimming and sequence filtering, the software NGS QC Toolkit (version 2.3.3, www.nipgr.res.in/ngsqctoolkit.html) was employed to remove sequencing adapters residues and low-quality reads.

The assembly was performed using AllPaths-LG (software.broadinstitute.org/allpaths-lg) using a maximum coverage of 100X for each library. Genome structural and functional annotation were performed with the MAKER pipeline (www.yandell-lab.org/software/maker.html). Functional annotation of the predicted genes was made using InterProScan v.5.21.60. CAZymes were predicted based on dbCAN HMMs (csbl.bmb.uga.edu/dbCAN). Transporters were predicted based on Transporter Classification DataBase – TCDB (http://www.tcdb.org/) and transcription factors based on DBD - Transcription factor prediction database (http://www.transcriptionfactor.org).

Transcriptome analysis

Illumina Hiseq2000 100 bp paired-end reads were used for transcript quantification. Quality-filtered reads were mapped to the *H. grisea* assembled in this work, using the TopHat2 v2.0.4 aligner (ccb.jhu.edu/sotware/tophat). Htseq version 0.6.0 was used to count reads mapped to *T. reesei* transcripts. The R package DESeq2 version 1.6.3 was used to perform the differential expression analysis, using the raw number of reads mapped to each gene in each sample to perform statistical tests, based on the negative binomial distribution, which indicate whether a gene is differentially expressed in a condition relative to each other. Therefore, the DESeq2 package was utilized for normalization, using the median log deviation, and for the differential expression analysis, applying an adjusted $p \leq 0.05$ as threshold. Functional enrichment analysis of differentially expressed genes based on Gene Ontology (GO) terms was performed using the R package GO_MWU (https://github.com/z0on/GO_MWU).

Results and discussion

Humicola grisea genome features

We sequenced the genome of the *H. grisea* var. Thermoidea in an Illumina-based whole-genome shotgun sequencing approach delivering 9,460,608 paired reads with an approximate insert size of 350 bp. The acquired sequence reads were assembled into 509 scaffolds using AllPaths-LG. The resulting genome features are shown in Table 1.

Table 1. Genome features of *Humicola grisea* var. Thermoidea genome.

Features	H. grisea genome
Genome size	29.16 Mb
Scaffolds	509
N50	156 Kb
GC content	55,8%
Genes	7120
Genes with PFAM domains	78.56%
Secreted proteins (SignalP 4.1)	650
Completeness (BUSCO)	98.20%

Trancriptome analysis in different pH

In this study, an RNA-seq approach was used to map genes differentially expressed during *H. grisea* growth on a medium containing sugarcane bagasse in pH 5 and pH 8 and normalized with the same conditions but using glucose as sole carbon source. A total of 323,849,916 sequence reads were obtained after quality trimming and aligned onto the reference genome. Figure 1 shows the PCA of samples and replicates based on expression patterns using DESeq2 package. The PCA reveals the good quality of the data and that pH has a crucial role on determining the gene during sugarcane bagasse deconstruction.

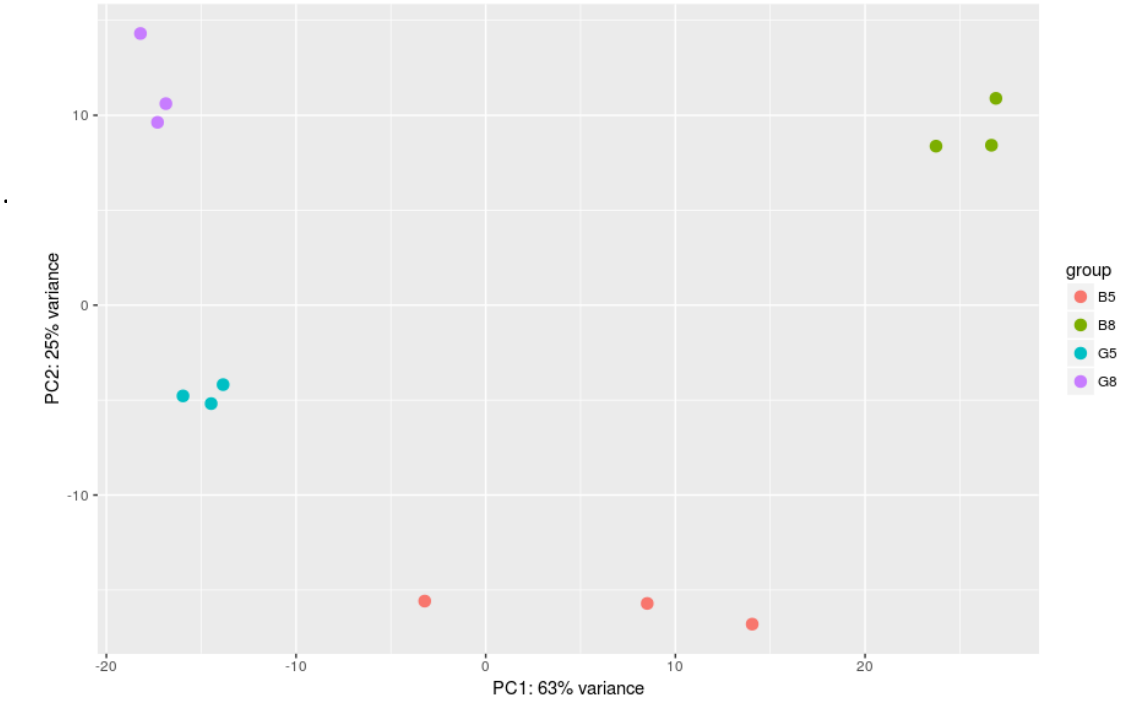

Figure 1. Principal component analysis (PCA) log transformed counts from biological replicates.

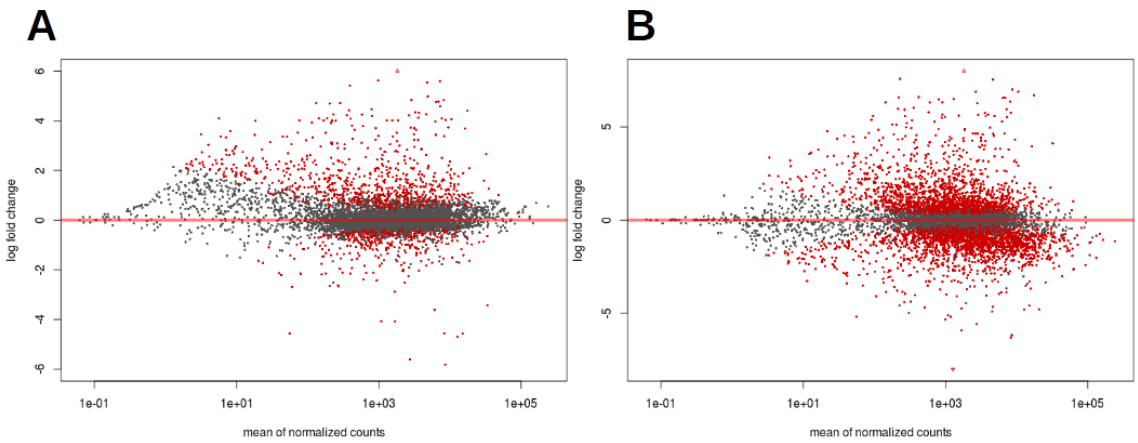

Figure 2. MA plot showing the log2 fold changes over the mean of normalized counts. Differentially expressed genes are shown in red (adjusted p-value < 0.05) **A.** Expression data of *Humicola grisea* grown in 0.1% of sugar cane bagasse in pH 5 during 6-12h hours of cultivation, normalized with same media using 1% glucose as carbon source. **B.** Expression data of *Humicola grisea* grown in 0.1% of sugar cane bagasse in pH 8 during 6-12h hours of cultivation, normalized with same media using 1% glucose as carbon source.

The MAplots in figure 2 show the distribution of expression *H. grisea* in pH 5 (Figure 2A) and pH 8 (Figure 2B). The growth in pH 8 clearly shows more genes differentially expressed compared with pH 5 (4218 genes and 1353 genes, respectively). Another difference is the distribution between up- and down-regulated genes. In pH 5, 842 genes were up-regulated and 511 down-regulated. On the other hand, in pH 8, 1927 genes were up-regulated and 2291 genes were down-regulated (Figure 3). This difference could be related to a larger diversity of pathways required to grow in pH 8, while pH 5 is a more “natural” environment to H. grisea grow and secreted its enzyme arsenal.

To broadly compare gene expression patterns between pHs, functional categories were assigned to the differentially expressed genes according to Gene Onthology (GO) guidelines using the R package GO_MWU. To enrich the category analysis for up and down regulated genes at each pH, and exact Fisher test ($p < 0.05$) was performed (Figure 3). In pH 5, as expected, categories related to primary metabolism and polysaccharide deconstruction, especially hydrolases, were significantly up-regulated. In pH 8, up-regulated categories were similar to pH 5 (an exception is presence of kinase category), on the other hand, down-regulated categories show a larger diversity of functions – RNA metabolism, transmembrane transporter, oxidoreductase and electron carrier.

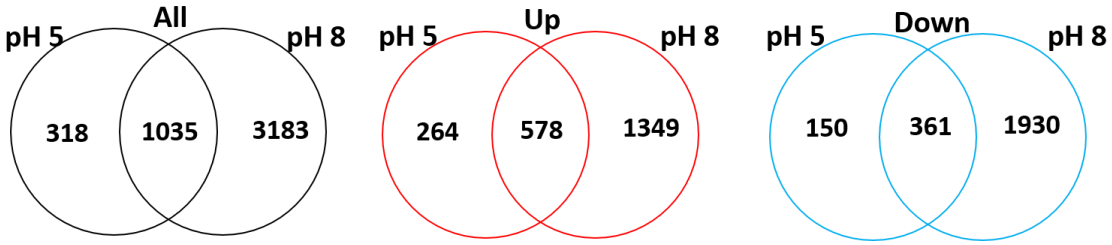

Figure 3. Venn diagram of differentially expressed genes in pH 5 and pH 8. All differentially expressed genes in black, up-regulated in red and down-regulated in blue.

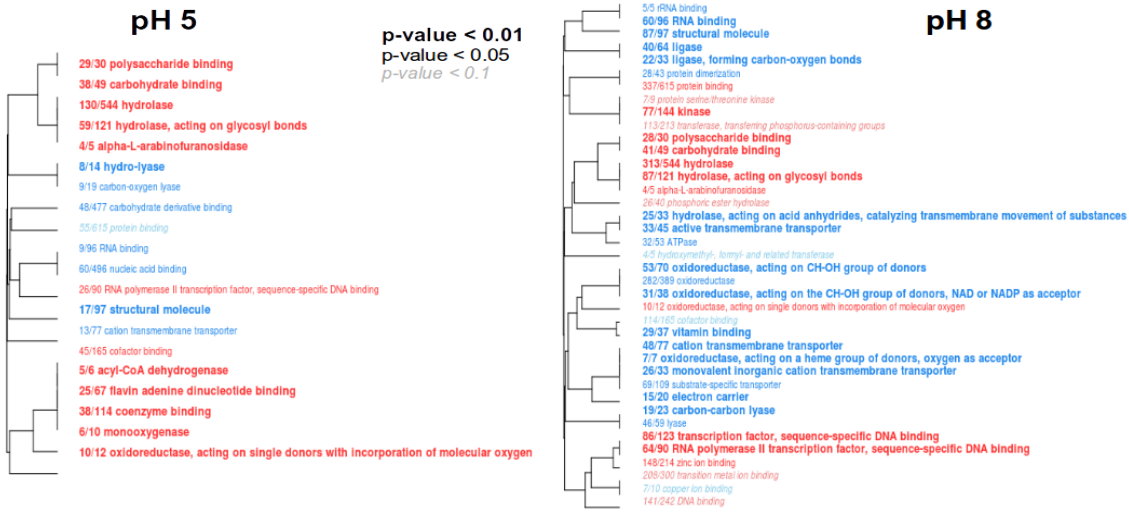

Figure 4. Enrichment of Gene Onthology (GO) categories (Molecular Function) of differentially expressed genes on pH 5 and pH 8. The tree clusters show related categories and the number prior to description represents the number of genes on that category and differentially expressed compared with the whole genome content. In red: up-regulated categories; In blue: down-regulated categories.

Among the top 10 up-regulated genes in both conditions shows a strong presence of carbohydrate active enzymes (CAZy) – celulases, xylanases, mannanases and esterases. An interesting result is the presence of the gene PID9 (pisatin demethylase), differentially expressed in both conditions but more expressed in pH 8, usually produced by phytopathogens to detoxify phytoalexins produced by plants (COLEMAN et al., 2011). This result suggests that *H. grisea* express this gene for sugarcane bagasse detoxification.

Table 2. Expression of top 10 genes differentially expressed and up-regulated on pH 5 and 8.

Top 10 up-regulated pH 5				Top 10 up-regulated pH 8			
Putative function (blast best-hit)	**log_2FC pH 5**	**log_2FC pH 8**	**CAZy annotation**	**Putative function (blast best-hit)**	**log_2FC pH 5**	**log_2FC pH 8**	**CAZy annotation**
AXE1 Acetylxylan esterase	7.04	9.05	CE5-CBM1	MANA endo-1,4-beta-mannosidase	6.68	10.12	GH26-CBM35
MANA endo-1,4-beta-mannosidase	6.68	10.12	GH26-CBM35	XYLB Aryl-alcohol dehydrogenase	5.49	10.07	-
CEL1 Cellulose-growth-specific protein	6.55	6.66	AA9	AXE1 Acetylxylan esterase	7.04	9.05	CE5-CBM1
BXLB exo-1,4-beta-xylosidase	6.44	5.59	GH3	EGLD endo-beta-1,4-glucanase D	4.86	8.69	AA9
EGLD endo-beta-1,4-glucanase D	6.05	1.24	AA9	YPS3 Aspartic proteinase yapsin-3	5.12	8.54	-
GRIF Grixazone synthase	5.86	4.81	-	no hit	4.44	8.06	GH43-CBM35
EGLD endo-beta-1,4-glucanase D	5.78	5.84	CBM1-AA9	GUX1 Exoglucanase 1	5.57	7.78	GH7
YF36 Uncharacterized hydrolase	5.66	4.32	-	CUTI Cutinase	4.79	7.76	CE5
PME Pectinesterase	5.61	5.20	CE8	EGLD endo-beta-1,4-glucanase D	2.66	7.74	AA9
GUX1 Exoglucanase 1	5.57	7.78	GH7	PID9 Pisatin demethylase	4.61	7.68	-

Conclusions

In conclusion, our transcriptomic data reveals that pH has a great importance in biomass transformation modulating several pathways, especially in basic pH, however, enzymes belonging to CAZy group are still important in such condition. Our genomic data will provide support for other works involving biomass degradation improvement and heterologous expression.

Acknowledgements

This work was supported by Embrapa Agroenergia and a special thanks to Lunalva P. Sallet for genomic DNA isolation.

References

COLEMAN, J. F.; WASMANN, C. C.; USAMI, T.; WHITE, G. J.; TEMPORINI, E. D.; MCCLUSKEY, K.; VANETTEN, H. D. Characterization of the encoding pisatin demethylase (FoPDA1) in Fusarium oxysporum. **Molecular Plant-Microbe Interactions**, v. 24, n. 12, p. 1482-1491, 2011.

MELLO-DE-SOUSA, T. M.; SILVA-PEREIRA, I.; POÇAS-FONSECA, M. J. Carbon source and pH-dependent transcriptional regulation of cellulase genes of Humicola grisea var thermoidea grown on sugarcane bagasse. **Enzyme Microbial Technology**, n. 48, edição 1, p. 19-26, 2011.