

“Integrative Genomic Analysis for the Bioprospection of Regulators and Accessory Enzymes Associated with Cellulose Degradation in a Filamentous Fungus (*Trichoderma harzianum*)”

Jaire A. Ferreira Filho^{1,2}, Maria Augusta C. Horta^{1,3}, Clelton A. dos Santos¹, Deborah A. Almeida^{1,2}, Natália F. Murad¹, Juliano S. Mendes¹, Danilo A. Sforça¹, Claudio Benício C. Silva¹, Aline Crucello¹, Anete P. de Souza^{1,4,*}

¹Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil

²Graduate Program in Genetics and Molecular Biology, Institute of Biology, UNICAMP, Campinas, SP, Brazil

³Holzforshung München, TUM School of Life Sciences Weihenstephan, Technische Universität München, 85354 Freising, Germany

⁴Department of Plant Biology, Institute of Biology, UNICAMP, Campinas, SP, Brazil

Supplementary data

This PDF file includes:

Figs S1 to S3

Tables S2, S3, S8 and S9

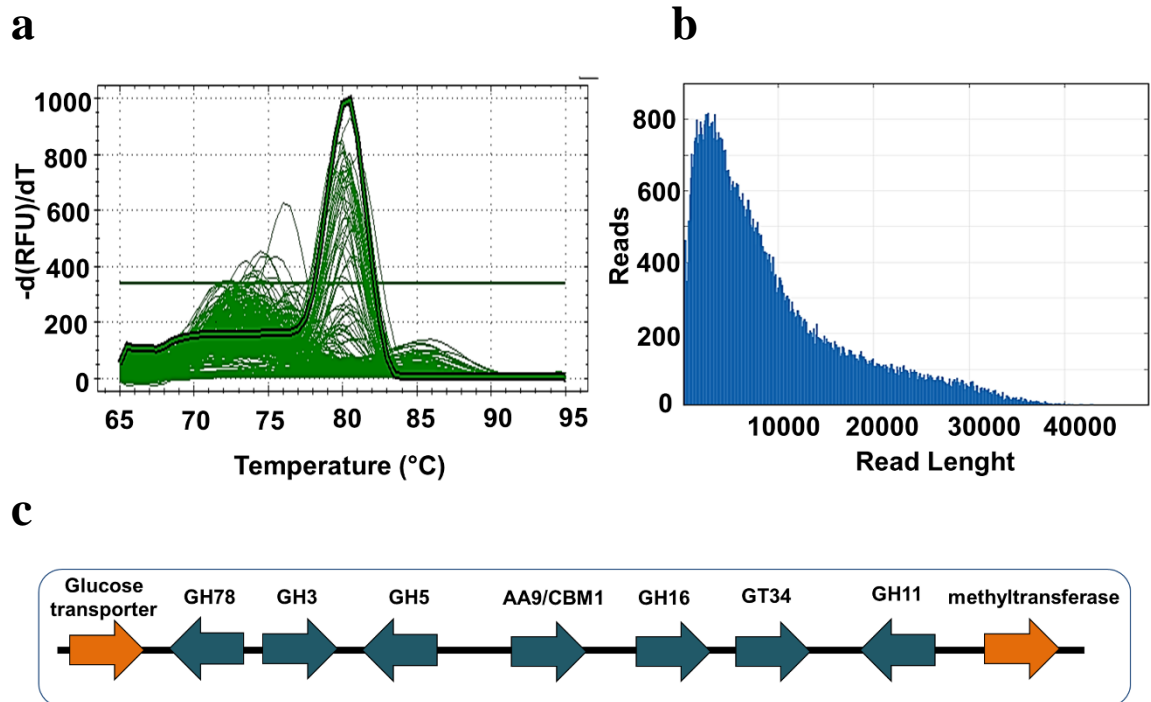


Fig. S1. Screening genes of interest in the genomic library of *T. harzianum* IOC3844 by qPCR (a); reads size sequenced using PACBio technology (b); genes cluster in a genomic region of *T. harzianum* (c).

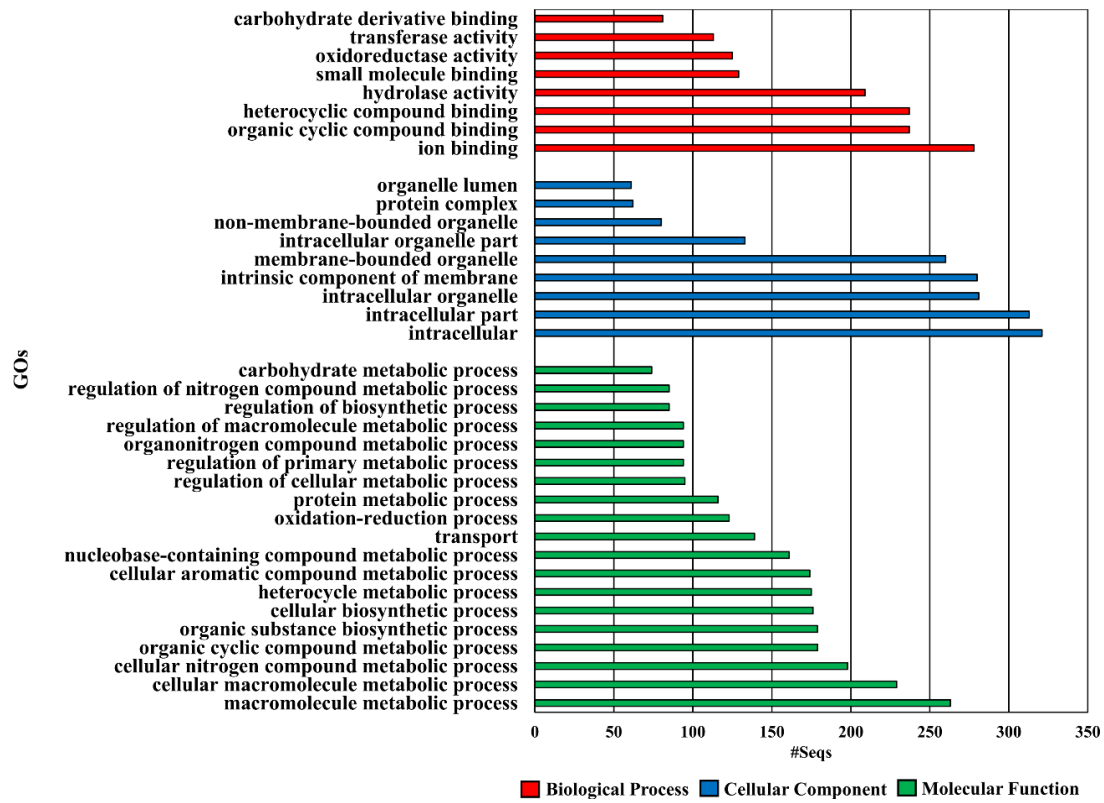


Fig. S2. Distribution of the main GO terms of the annotated genes in *T. harzianum* IOC3844.

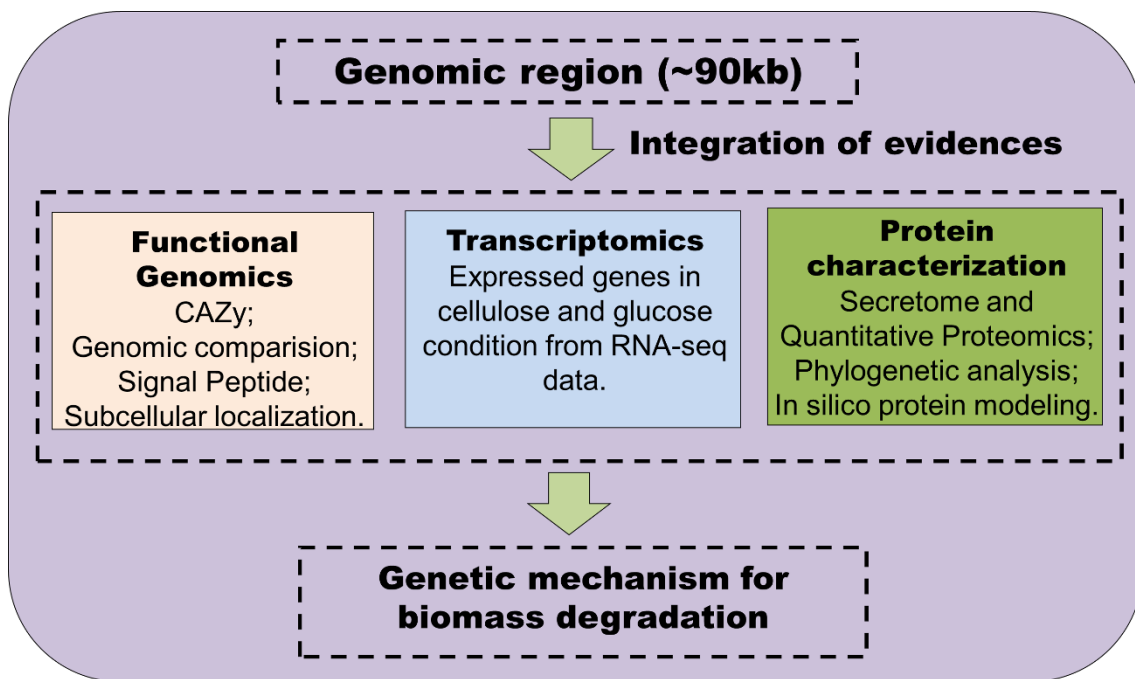


Fig. S3. Pipeline approach for the analyzes used in this work of genes and genomic study in *T. harzianum*.

Supplementary Table S2. Assembly parameters of a set of sequenced genomic region using PACBio technology.

Parameters	Results
Polymerase Read Bases	1151876133
Seed Bases	150009116
Pre-Assembly Yield	0.296
Pre-Assembly Reads Length	5528
Length Cutoff	12815
Pre-Assembly bases	44419992
Pre-Assembly Reads	8035
Pre-Assembly N50	8199
Mean Read Score	0.667
Number of Reads	100732
Mean Read Length	11493
Number of bases	1157715225
N50 Read Length	17608
Polymerase Read Quality	0.849

Supplementary Table S3. Comparison of genomic data among different species of *Trichoderma* spp.

Organism/Name	Strain	Genome type	Size (Mb)	GC%	Scaffolds	CAZymes
<i>T. harzianum</i>	IOC3844	BACs	5.00	49.65	63	222
<i>T. harzianum</i>	T6776	Draft	40.98	48.50	1572	430
<i>T. reesei</i>	QM6a	Complete genome	33.39	52.80	7	327
<i>T. atroviride</i>	IMI 206040	Draft	36.14	49.70	29	422
<i>T. virens</i>	Gv29-8	Draft	39.00	49.20	93	441

Supplementary Table S8. Description of the species used for the phylogenetic analysis of the transcription factor CLR2.

Group	Sequence ID	Species
1	XP_012182108.1	<i>Fibroporia radiculosa</i>
	KZT72632.1	<i>Daedalea quercina</i> L-15889
	PCH41399.1	<i>Wolfiporia cocos</i> MD-104 SS10
	XP_024336022.1	<i>Postia placenta</i> MAD-698-R-SB12
	OCH94255.1	<i>Obba rivulosa</i>
	PSR73669.1	<i>Phlebia centrifuga</i>
	XP_007392271.1	<i>Phanerochaete carnosae</i> HHB-10118
	XP_008035817.0	<i>Trametes versicolor</i> FP-101664 SS1
	XP_007364283.1	<i>Dichomitus squalens</i> LYAD-421
	OJA15282.1	<i>Rhizopogon vesiculosus</i>
	OAX40653.1	<i>Rhizopogon vinicolor</i> AM-OR11-026
	KIJ70318.1	<i>Hydnomerulius pinastri</i> MD-312
	KIK30790.1	<i>Pisolithus microcarpus</i> 441
	KIO13022.1	<i>Pisolithus tinctorius</i> Marx 270
	2	OXG14340.1
OWZ51650.1		<i>Cryptococcus neoformans</i> var. <i>grubii</i> 125.91
OXG77158.1		<i>Cryptococcus neoformans</i> var. <i>grubii</i> Br795
OWZ35064.1		<i>Cryptococcus neoformans</i> var. <i>grubii</i> AD1-83a
XP_012051901.1		<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99
OWT36835.1		<i>Cryptococcus neoformans</i> var. <i>grubii</i> Bt1
OXG13205.1		<i>Cryptococcus neoformans</i> var. <i>grubii</i> Tu401-1
OWZ76252.1		<i>Cryptococcus neoformans</i> var. <i>grubii</i> Bt85
OXM75578.1		<i>Cryptococcus neoformans</i> var. <i>grubii</i> Bt63
XP_018260593.1		<i>Kwoniella dejecticola</i> CBS 10117
XP_018999125.1		<i>Kwoniella mangroviensis</i> CBS 8507
XP_019045940.1		<i>Kwoniella bestiolae</i> CBS 10118
XP_019014217.1		<i>Kwoniella pini</i> CBS 10737
XP_018266380.1		<i>Kwoniella dejecticola</i> CBS 10117
OCF60695.1		<i>Kwoniella mangroviensis</i> CBS 10435
OCF74556.1		<i>Kwoniella mangroviensis</i> CBS 8886
XP_019001153.1		<i>Kwoniella mangroviensis</i> CBS 8507
XP_021869321.1		<i>Kockovaella imperatae</i>
XP_019045848.1		<i>Kwoniella bestiolae</i> CBS 10118
XP_018260036.1		<i>Kwoniella dejecticola</i> CBS 10117
XP_003196556.1		<i>Cryptococcus gattii</i> WM276

2	OWZ29927.1	<i>Cryptococcus neoformans</i> var. <i>grubii</i> c45
	XP_019010711.1	<i>Kwoniella pini</i> CBS 10737
	POY75593.1	<i>Rhodotorula taiwanensis</i>
	XP_018280147.1	<i>Cutaneotrichosporon oleaginosum</i>
	EKD01798.1	<i>Trichosporon asahii</i> var. <i>asahii</i> CBS 8904
	XP_014181811.1	<i>Trichosporon asahii</i> var. <i>asahii</i> CBS 2479
3	CDM38126	<i>Penicillium roqueforti</i> FM164
	XP_660973.1	<i>Aspergillus nidulans</i> FGSC A4
4	PKK47514.1	<i>Trichoderma harzianum</i> CL102_9848
	Th_IOC3844	<i>Trichoderma harzianum</i> IOC3844
	XP_024778108.1	<i>Trichoderma harzianum</i> CBS 226.95
	KKP03054.1	<i>Trichoderma harzianum</i>
	XP_013955330.1	<i>Trichoderma virens</i> Gv29-8
5	ETS03553.1	<i>Trichoderma reesei</i> RUT C-30
	XP_006964036.1	<i>Trichoderma reesei</i> QM6a
	XP_013941191.1	<i>Trichoderma atroviride</i> IMI 206040
	XP_024764890.1	<i>Trichoderma asperellum</i> CBS 433.97
	KOS18048.1	<i>Escovopsis weberi</i>
6	CZT51594.1	<i>Rhynchosporium secalis</i>
	CZT06835.1	<i>Rhynchosporium agropyri</i>
	CZT07141.1	<i>Rhynchosporium commune</i>
	PMD16076.1	<i>Pezoloma ericae</i>
	PMD30568.1	<i>Hyaloscypha variabilis</i> F
	XP_024739234.1	<i>Meliniomyces bicolor</i> E
	XP_018061285.1	<i>Phialocephala scopiformis</i>
	CZR63207.1	<i>Phialocephala subalpina</i>
	XP_007288749.1	<i>Marssonina brunnea</i>
	PQE16296.1	<i>Rutstroemia</i> sp. NJR-2017a WRK4
	PQE03599.1	<i>Rutstroemia</i> sp. NJR-2017a BBW
	ESZ95397.1	<i>Sclerotinia borealis</i> F-4128
	EMR81895.1	<i>Botrytis cinerea</i> BcDW1
	XP_024552100.1	<i>Botrytis cinerea</i> B05.10
	CCD51977.1	<i>Botrytis cinerea</i> T4
	KFX95591.1	<i>Pseudogymnoascus</i> sp. VKM F-3557
	OBT56849.1	<i>Pseudogymnoascus</i> sp. 24MN13
	KFZ12187.1	<i>Pseudogymnoascus</i> sp. VKM F4519 (FW-2642)
	XP_018129406.1	<i>Pseudogymnoascus verrucosus</i>
	7	KUI68282.1

7	KUI55118.1	<i>Valsa mali</i> var. <i>pyri</i>
	POS76166.1	<i>Diaporthe helianthi</i>
	KKY36004.1	<i>Diaporthe ampelina</i>
	PSS02243.1	<i>Coniella lustricola</i>
8	XP_016589835.1	<i>Sporothrix schenckii</i> 1099-18
	KIH94787.1	<i>Sporothrix brasiliensis</i> 5110
	EPE08882.1	<i>Ophiostoma piceae</i> UAMH 11346
9	XP_009849767.1	<i>Neurospora tetrasperma</i> FGSC 2508
	XP_962712.2	<i>Neurospora crassa</i> OR74A
	XP_003347695.1	<i>Sordaria macrospora</i> k-hell
	XP_006693821.1	<i>Chaetomium thermophilum</i>
	KXX83256.1	<i>Madurella mycetomatis</i>
	XP_003660436.1	<i>Thermothelomyces thermophila</i>
10	KLU85969.1	<i>Magnaporthiopsis poae</i> ATCC 64411
	XP_009224916.1	<i>Gaeumannomyces tritici</i> R3-111a-1
	XP_003714853.1	<i>Magnaporthe oryzae</i> 70-15
11	OTA63138.1	<i>Hypoxylon</i> sp. EC38
	OTA97472.1	<i>Hypoxylon</i> sp. CO27-5
	OTB08885.1	<i>Hypoxylon</i> sp. CI-4A
	OTB14150.1	<i>Daldinia</i> sp. EC12
	KXJ95218.1	<i>Microdochium bolleyi</i>
12	ELA25990.1	<i>Colletotrichum fructicola</i> Nara gc5
	ENH84300.1	<i>Colletotrichum orbiculare</i> MAFF 240422
13	OLN97553.1	<i>Colletotrichum chlorophyti</i>
14	KXH28512.1	<i>Colletotrichum nymphaeae</i> SA-01
	KXH27412.1	<i>Colletotrichum simmondsii</i>
	XP_022478568.1	<i>Colletotrichum orchidophilum</i>
	KXH59612.1	<i>Colletotrichum salicis</i>
	XP_008097168.1	<i>Colletotrichum graminicola</i> M1.001
	KDN69204.1	<i>Colletotrichum sublineola</i>
	KZL75422.1	<i>Colletotrichum tofieldiae</i>
	OHW92794.1	<i>Colletotrichum incanum</i>
	KZL81134.1	<i>Colletotrichum incanum</i>

Supplementary Table S9. Description of the genes found in the coregulation networks.

Network ID	Protein ID	Protein name
1	KKP06817.1	Hypotetical protein
2	KKP05702.1	Hypotetical protein
3	KKP00617.1	Hypotetical protein
4	KKP03936.1	Kinase DC2
5	KKP07737.1	Translation initiation factor 3 subunit k
6	KKP01671.1	Hypotetical protein
7	KKO98378.1	Translation initiation factor subunit 1
8	KKP03537.1	Murein transglycosylase
9	KKO96961.1	Hypotetical protein
10	KKP00812.1	Hypotetical protein
11	KKP06268.1	Hypotetical protein
12	KKP01476.1	Peptidyl-prolyl-cis-trans isomerase sspl
13	KKO97887.1	Hypotetical protein
14	KKO99717.1	Hypotetical protein
15	KKP00810.1	Hypotetical protein
16	KKP07167.1	Hypotetical protein
17	KKP02416.1	Hypotetical protein
18	KKO98063.1	Vacuolar ATP synthase subunit D
19	KKP03148.1	Hypotetical protein
20	KKP07719.1	Hypotetical protein
21	KKO97889.1	Hypotetical protein
22	KKP00888.1	ATP-dependent RNA helicase DED1
23	KKP02267.1	Ubiquitin-protein ligase E3 C
24	KKP02972.1	Hypotetical protein
25	KKO98726.1	DUF718
26	KKP02131.1	Hypotetical protein
27	KKO96618.1	Hypotetical protein
28	KKO98708.1	Hypotetical protein
29	KKP02590.1	Proteasome subunit alpha type-2
30	KKO99933.1	Hypotetical protein
31	KKP07598.1	Undercaprenyl diphosphate synthase
32	KKP00804.1	14-3-3 family protein
33	KKP02280.1	ATP-dependent RNA helicase SUB2
34	KKP03054.1	Cellulose Degradation regulator 2 – CLR2
35	KKO97002.1	Hydroxyacylglutathione hydrolase
36	KKP01112.1	Hypotetical protein