In silico characterization of hypothetical proteins from *Paracoccidioides lutzii*

P.F.F. Silva[1], E. Novaes[2], M. Pereira[1], C.M.A. Soares[1], C.L. Borges[1] and S.M. Salem-Izacc[1]

[1]Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Goiás, Goiânia, GO, Brasil
[2]Escola de Agronomia, Universidade Federal de Goiás, Goiânia, GO, Brasil

Corresponding author: S.M. Salem-Izacc
E-mail: silviaizacc@gmail.com

ABSTRACT. Nearly 60% of *Paracoccidioides lutzii* genes encode products annotated as hypothetical or predicted proteins (HPs). In this study, we describe the global detection and functional inference of HPs, using computational methods based on sequence similarity, identification of targeting signals, presence of known protein domains, and use of the Gene Ontology functional classification scheme. Our analysis enabled a high-throughput characterization of predicted cellular localization and presence of protein domains, clustering HPs into different functional categories including metabolism, localization, cell cycle, response to stimulus, and signaling. To investigate *P. lutzii* HP expression profiles, we used data obtained from the expressed sequence tag database (dbEST). These analyses revealed 2364 HPs expressed in different situations, namely in mycelial and yeast forms, during the transition from mycelium to yeast, and under conditions mimicking infection. Based on this transcriptomic data, we performed a functional enrichment analysis according to the domains present in the HPs expressed in each condition. The most overrepresented functional domains were those involved in the regulation of gene expression, suggesting important and as yet undescribed roles

for these HPs in the adaptation of *P. lutzii* to environmental conditions. In addition, the expression profiles of six randomly selected HPs were analyzed by quantitative real-time polymerase chain reaction in order to verify their expression in the complementary DNA libraries analyzed in this investigation. The approach used in this study should improve functional characterization of *P. lutzii* HPs.

Key words: *Paracoccidioides*; Hypothetical proteins; *In silico* analysis

INTRODUCTION

Despite the great number of genomic and transcriptomic sequences deposited in databases over recent years, a large proportion of the genes identified encode a group of proteins with unknown biochemical, biophysical, or cellular functions, annotated as predicted or hypothetical proteins (HPs). HPs are proteins predicted from nucleic acid sequences, based on computational identification of gene structure, but for which no evidence of *in vivo* expression exists. These proteins may represent up to half of the proteome for several organisms (Loewenstein et al., 2009; Shahbaaz et al., 2013; Ijaq et al., 2015).

Paracoccidioides brasiliensis and *P. lutzii* are thermodimorphic pathogens that cause paracoccidioidomycosis, a mycosis endemic to Latin America. The genus *Paracoccidioides* comprises the cryptic species *P. brasiliensis* S1, PS2, and PS3, and the closely related *P. lutzii.* Until recently, *P. brasiliensis* and *P. lutzii* were considered a single species, identified using the former's name (Matute et al., 2006; Teixeira et al., 2009). Infection results from inhalation of hyphal fragments or spores released from mycelium (M), resulting in exposure to elevated temperatures and triggering a morphological switch to the yeast (Y) form (Restrepo, 1985; Franco et al., 1987). The pathogenicity of these fungi seems to be related to this morphological change, since strains unable to differentiate are often avirulent (Medoff et al., 1987; Borges-Walmsley et al., 2002).

Nearly 60% of *Paracoccidioides* genes are annotated as encoding HPs (Desjardins et al., 2011). The characterization of proteins of unknown function poses a challenge both to functional genomics and to biology in general, and if achieved, could help to fill the gaps in attempts to attain complete genomic and proteomic data. Several computational approaches are available for predicting protein function (Doerks et al., 2012; Qiao et al., 2013; Ijaq et al., 2015). This work describes an *in silico* characterization of unclassified proteins from *P. lutzii* using analysis of predicted cellular localization, conserved protein domains, and functional classification according to the Gene Ontology (GO) scheme. In addition, we measured the expression of HPs in complementary DNA (cDNA) libraries constructed from *P. brasiliensis* and *P. lutzii* M and Y forms (Goldman et al., 2003; Felipe et al., 2005), during the transition from M to Y (Bastos et al., 2007), and under conditions mimicking infection (Bailão et al., 2006; Bailão et al., 2007; Costa et al., 2007). We then performed functional enrichment analysis of the HPs expressed in these different transcriptomes to identify overrepresented protein domains.

MATERIAL AND METHODS

Determination of *Paracoccidioides* HPs

A list of HPs was obtained from the *P. brasiliensis* genome database (http://www.broadin-

stitute.org/annotation/genome/paracoccidioides_brasiliensis/MultiDownloads.html). For the present study, HPs were considered as all proteins annotated as "hypothetical protein", "conserved hypothetical protein", or "predicted protein".

Identification of HPs expressed in *Paracoccidioides* cDNA libraries

To compile a list of expressed HPs, we extracted all *Paracoccidioides* expressed sequence tags (ESTs) deposited in the dbEST database (http://www.ncbi.nlm.nih.gov/dbEST) as of September 19, 2013. These ESTs were then used as queries in blastx searches (Altschul et al., 1990) against the *P. lutzii* protein database available on the Broad Institute website (Desjardins et al., 2011). The following cutoff values were used to identify expressed HPs: E-value, 10^{-6}; identity, 60%; and alignment length, 50 amino acids.

Prediction of cellular localization

Cellular localization of HPs was predicted using TargetP (Emanuelsson et al., 2007), under the default settings. TargetP suggests the subcellular location of eukaryotic proteins based on the predicted presence of N-terminal presequences.

GO term and protein domain annotation

We assigned GO terms and protein domains to HPs encoded in the *P. lutzii* genome using the Blast2GO program (Conesa et al., 2005). Deduced HP sequences were compared to those in the National Center for Biotechnology Information (NCBI) non-redundant protein database, with a cutoff value of 10^{-6}. The pipeline began with a BLAST search against this database, following which the program consolidated annotations of BLAST hits to attribute a putative function to each protein based on GO data.

Enrichment analysis for protein domains

In order to assign a biological meaning to the HPs expressed in the different cDNA libraries, we performed an enrichment analysis of the protein domains identified in each condition. This was carried out online by inserting the list of expressed HP locus IDs for each condition into the Broad Institute's functional enrichment analysis tool (http://www.broadinstitute.org/annotation/genome/paracoccidioides_brasiliensis/BatchSelect.html?target=GeneEnrichment.html). We imposed a corrected P value cutoff of 0.2.

P. lutzii growth conditions

P. lutzii strain Pb01 (ATCC MYA-826) was used in quantitative real-time polymerase chain reaction (qPCR) experiments in this study. Cells were cultured in liquid Fava Netto's medium (0.3% proteose peptone, 1% peptone, 0.5% meat extract, 0.5% yeast extract, 4% glucose, and 0.5% NaCl, pH 7.2) at 36 and 22°C, for Y and M forms, respectively. The M to Y transition was performed in the same medium by raising the incubation temperature from 22° to 36°C over 22 h.

Incubation of *P. lutzii* Y cells with human blood or plasma

P. lutzii Y cells were incubated with either human blood or plasma, as described by Bailão et al. (2006, 2007). Briefly, human blood from five healthy donors was collected by venipuncture using heparinized syringes and centrifuged at 1000 *g*. Fungal *c*ells were harvested from 7-day-old cultures and washed once with phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 1.4 mM NaH_2PO_4, 4.3 mM Na_2HPO_4, pH 7.4) before being inoculated (at 5 x 10^6 cells/mL) into 5 mL of human blood or plasma at 36°C for 10 min with shaking. They were then collected by centrifugation for 5 min at 1500 *g*, and washed five times with PBS. The use of human blood was approved by the Research Ethics Committee of Universidade Federal de Goiás. Written informed consent was obtained from all donors. All specimens were handled and made anonymous according to ethical and legal standards.

RNA extraction

Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer protocol. RNA samples were isolated from Y and M cells, those undergoing M to Y differentiation, and Y cells having been incubated with human blood or plasma for 10 min.

qPCR assay

RNA samples were reverse transcribed using a High Capacity RNA-to-cDNA kit (Applied Biosystems, Foster City, CA, USA). cDNA samples were then diluted 1:5 in water, and qPCR was performed using SYBR green PCR master mix and the StepOnePlus Real-Time PCR System (both Applied Biosystems). Specific primers (**Table S1**) were designed for each selected gene using the NCBI Primer-BLAST program (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Reaction specificity was evaluated by analyzing the dissociation curves of amplified products. qPCR was performed in triplicate for each cDNA sample. Measurements were normalized using *Paracoccidioides* beta-tubulin as an endogenous control and relative transcript expression was estimated with the comparative cycle threshold (ΔΔCt) method (Livak and Schmittgen, 2001).

RESULTS AND DISCUSSION

Overview of HPs encoded in the *Paracoccidioides* genome and transcriptomes

Annotation of the genomes of three *Paracoccidioides* strains revealed that approximately 60% of their genes encode proteins of unknown function, generally annotated as HPs. Almost 50% of these HPs are conserved, i.e. they have also been identified in other organisms, while the remainder appear to be restricted to *Paracoccidioides* (Table 1).

We also utilized the *Paracoccidioides* ESTs deposited in the dbEST. Our analysis revealed that the percentage of peptide-encoding transcripts identified as HPs in the cDNA libraries ranged from 24 to 46% (Table 2), indicating that a substantial number of HPs are expressed. Altogether, we detected 2364 unique transcripts annotated as HPs in at least one of the libraries. In an effort to attribute biological meaning to the expression data, the libraries were arranged into three conditions: "M/Y", M and Y cells (Goldman et al., 2003; Felipe et al., 2005); "transition", cells undergoing

morphological transition from M to Y form (Bastos et al., 2007); and "infection", cells under conditions mimicking infection (Bailão et al., 2006, 2007; Costa et al., 2007). Of the 2364 unique HP transcripts, 2324 were able to be placed into one or more of these three conditions. Figure 1 shows a Venn diagram depicting the number of HPs identified in each situation. Only 25 were expressed under all conditions, demonstrating that these proteins are regulated by environmental factors. Interestingly, 117 HPs were exclusively expressed in an infection-like context, indicating important roles during this process, and emphasizing the need to establish their functions.

Table 1. Summary of the proteins annotated as hypothetical encoded in the genomes of *Paracoccidioides* species and strains.

Species (strain)	Total proteins	Number of predicted proteins (%)	Number of hypothetical proteins (%)	Number of conserved hypothetical proteins (%)	Total number of HPs (%)
P. lutzii (Pb01)	9132	2116 (23.2%)	592 (6.5%)	2928 (32.1%)	5636 (61.7%)
P. brasiliensis (Pb03)	7876	1922 (24.4%)	280 (3.6%)	2540 (32.2%)	4742 (60.2%)
P. brasiliensis (Pb18)	8741	1997 (22.8%)	401 (4.6%)	2916 (33.4%)	5314 (60.8%)

HP = hypothetical protein.

Table 2. Survey of the hypothetical protein transcripts expressed in *Paracoccidioides* transcriptomes.

Condition	Library ID (cells from which mRNA was extracted)	Strain	Number of ESTs in the library	Number of peptides with a blastx hit[*]	Number of HP peptides with a blastx hit[*]
M/Y	[a]LIBEST_012137 (M and Y)	Pb01	19,898	4075	1872 (45.94%)
	[b]LIBEST_010891 (Y)	Pb18	13,490	1868	715 (38.28%)
Transition	[c]LIBEST_020653 (22 h of M→Y transition)	Pb01	1107	477	176 (36.90%)
Infection	[d]LIBEST_020789 (Y recovered from the livers of infected mice)	Pb01	4934	1019	408 (40.04%)
	[e]LIBEST_019382 (Y incubated with human blood or plasma)	Pb01	1486	151	36 (23.84%)
-	All five libraries	-	40,915	5012	2364 (47.17%)

[a]Felipe et al., 2005; [b]Goldman et al., 2003; [c]Bastos et al., 2007; [d]Costa et al., 2007; [e]Bailão et al., 2006, 2007. *Result of a blastx search against the *Paracoccidioides* protein database. M = mycelium; Y = yeast; EST = expressed sequence tag; HP = hypothetical protein.
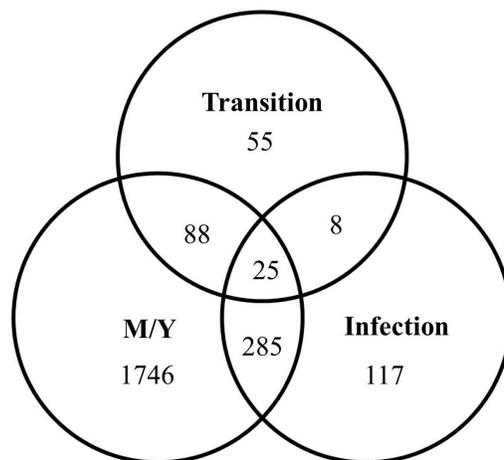

Figure 1. Venn diagram representing the number of hypothetical proteins expressed in each condition. M = mycelium; Y = yeast.

Protein length distribution

All HPs and annotated proteins encoded in the *P. lutzii* genome were plotted according to their length, and the percentage of expressed HPs in each length category was determined. As shown in Figure 2, more than 90% of the proteins consisting of 1 to 100 amino acid residues were annotated as HPs, however, only 11% of these had corresponding transcripts in the cDNA libraries. Of those proteins between 201 and 300 amino acids in length, more than 70% were identified as HPs, but only 31% of these were expressed. In contrast, the percentage of the largest proteins (longer than 301 residues) annotated as HPs was much lower (between 36 and 58%). However, a much larger proportion of these, from 47 to 67%, were expressed. These results suggest that most of the small proteins annotated as HPs may in fact correspond to incorrectly annotated genes and their legitimacy needs further confirmation. Our data corroborates earlier findings that open reading frame (ORF) calls relating to small peptides may simply indicate random sequences coincidentally forming ORFs (Qiao et al., 2013). These small sequences might also represent ancient functional genes having undergone pseudogenization.
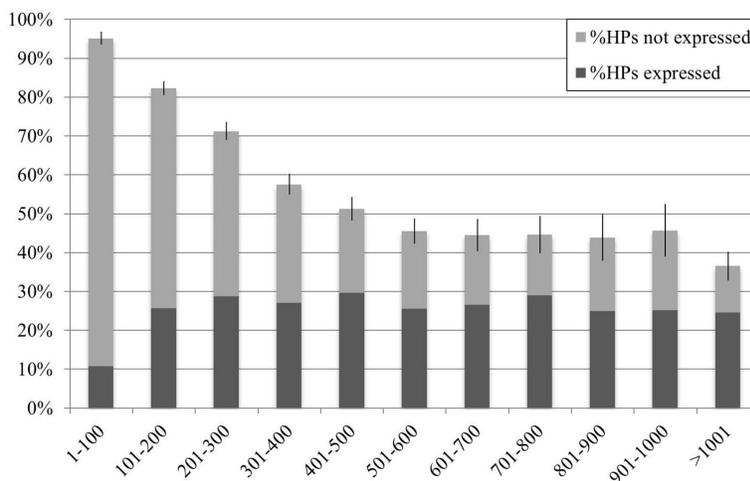

Figure 2. Length distribution of *Paracoccidioides lutzii* hypothetical proteins (HPs). The error bars correspond to the 95% confidence intervals.

Targeting signals and characterized protein domains

The predicted existence of targeting signals and specific domains in a protein suggests that it plays a functional role. Although these elements may also be found in pseudogenes as a result of gene duplication, their presence in a predicted amino acid sequence suggests that the protein is indeed expressed (Desler et al., 2009). Table 3 shows a summary of the predicted cellular localization of *P. lutzii* HPs. Most appear to be localized in the cytoplasm (65 to 70%), while around 20% contain a mitochondria-targeting peptide, and around 11% carry a secretory pathway signal peptide. The proportion of HPs allocated to each cellular compartment was similar between those identified by genome annotation and those expressed in the cDNA libraries.

Table 3. Predicted cellular localization of *Paracoccidioides lutzii* hypothetical proteins according to the TargetP program.

Predicted cellular localization	Number of HPs encoded in the Pb01 genome (%)*	Number of HPs expressed in the cDNA libraries (%)**
Mitochondrion	1101 (19.53%)	525 (22.21%)
Secretory pathway	637 (11.30%)	275 (11.63%)
Any other location	3898 (69.16%)	1564 (66.16%)

*Percentage based on the number of proteins annotated as HPs in the Pb01 genome (5636). **Percentage based on the number of unique HPs expressed (2364). HP = hypothetical protein; cDNA = complementary DNA.

Identifying conserved domains within a predicted polypeptide sequence can provide insight into the function, regulation, and/or localization of the putative protein. In total, 1984 HPs in the *P. lutzii* genome included at least one domain (**Table S2**). In order to discern the biological implications of the HP domains identified in the cDNA libraries, we performed a functional enrichment analysis based on the best-represented protein family (Pfam) domains (Finn et al., 2014).

Functional enrichment analysis

Our functional enrichment analysis correlated the protein domains identified by Pfam with the expression of HPs in each condition (Table 4).

Table 4. Functional enrichment of the hypothetical proteins expressed under various conditions.

Condition	Pfam term	Brief description	Corrected P value
M/Y	zf-CCHC	DNA-dependent transcription regulation	0.00
	Chromo	Structural domain associated with chromatin remodeling	0.01
	BolA	Involved in the elongation and septation process during cell division	0.19
	ChaC	Involved in active Ca^{2+}/ H$^+$ transport	0.19
	SAP18	Involved in RNA processing and apoptosis.	0.19
	SRP19	Involved in translation and translocation of proteins	0.19
	SRPRB	Beta-subunit of the signal recognition particle receptor	0.19
	DHFR_1	Reduction of dihydrofolic acid to tetrahydrofolic acid, using NADPH as electron donor	0.19
	Mit_rib_S27	Involved in protein synthesis	0.19
	Skp1_POZ	Involved in the ubiquitin protein degradation pathway	0.19
	Ribosomal_S28e	Plays an important role in the elongation step of protein synthesis	0.19
	Ubiq_cyt_C_chap	Required for assembly of coenzyme QF-2-cytochrome C reductase	0.19
Transition	zf-CCHC	Transcription factor	0.00
Infection	R3H	DNA-binding fator	0.04
	Thioredoxin	Class of redox proteins, plays a role in response to reactive oxygen species	0.18

M = mycelium; Y = yeast; Pfam = protein family; NADPH = reduced nicotinamide adenine dinucleotide phosphate.

The functional domains overrepresented during the M to Y transition consisted of those involved in protein synthesis and regulation of gene expression, both by transcription factors and chromatin remodeling. The phase transition is a strictly regulated process resulting in cell shape change, alterations in cell wall composition, and several adjustments in protein expression (Felipe et al., 2005; Nunes et al., 2005; Klein and Tebbets, 2007). Although this dimorphic transition has been extensively studied, little is known of the cellular mechanisms orchestrating these events. Our findings suggest that HPs might play important roles in this process.

Domains related to the oxidative stress response were found in HPs expressed under conditions mimicking infection. Recent studies have shown that *Paracoccidioides* species respond strongly to the high concentrations of reactive oxygen species produced by cells of the host immune system (de Oliveira et al., 2013, Pigosso et al., 2013).

Interestingly, domains involved in the regulation of gene expression were identified in all the conditions analyzed. This suggests that HPs are involved in reprogramming expression in response to a change in temperature, an event triggering differentiation, and the presence of host-derived factors.

Functional classification

GO is a functional classification scheme that defines, when possible, three categories for each gene product: biological process, molecular function, and cellular component (Ashburner et al., 2000). Of the 5636 HPs, 3044 were classified into at least one GO category (Figure 3 and **Table S2**). The HPs were implicated in a wide range of processes, including metabolism, establishment of localization, signaling, stimulus and stress responses, and cell wall organization and biogenesis. These results indicate that HPs take part in several cellular activities and may also participate in other, as yet undescribed, processes.
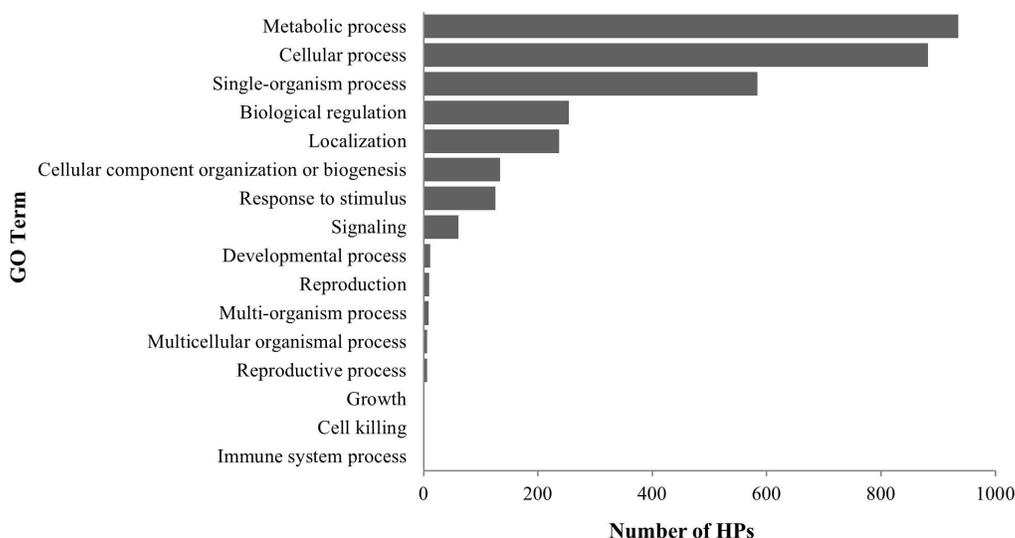
Figure 3. Functional classification of *Paracoccidioides lutzii* hypothetical proteins (HPs). GO = Gene Ontology.

HP gene expression analysis using qPCR

In order to verify their transcription profiles, qPCR assays (Figure 4) were performed for six randomly selected HP genes. All the genes tested were successfully amplified, demonstrating that they were indeed expressed. The qPCR results confirmed those obtained using the cDNA libraries, and in addition, revealed that some HPs were expressed under specific conditions but remained undetected in the corresponding library. For instance, the gene identified as PAAG_02090 was expressed in the M to Y transition, and that known as PAAG_08614 was expressed in M and Y states, as well as during transition, but neither were detected in any of the expression libraries.
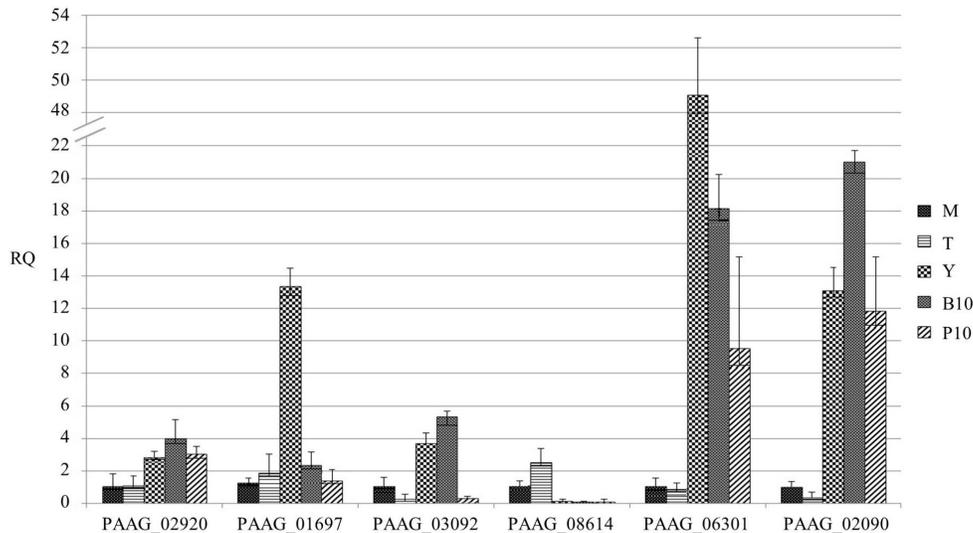
Figure 4. Transcriptional profiling of selected hypothetical proteins under different growth conditions. RNA samples were obtained from mycelial (M) and yeast forms (Y), cells undergoing transition (T; 48 h of M to Y transition), and Y cells incubated with either human blood (B10) or plasma (P10). RQ = relative quantity. The experiments were conducted in triplicate and the error bars represent standard deviations.

CONCLUSIONS

The large number of HPs encoded in the *Paracoccidioides* genome and expressed in its transcriptomes reveals a substantial gap in our knowledge regarding the functions of the proteins produced by these fungi. The characterization of such proteins constitutes a great challenge to our understanding of cellular metabolism and regulation of gene expression in microorganisms. This work highlights the significant contribution made by a large-scale computational analysis of *P. lutzii* HPs to the assignment of protein functions, and its ability to facilitate understanding of genomic and transcriptomic data.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

[**Supplementary material**](#)

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.

Ashburner M, Ball CA, Blake JA, Botstein D, et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: 25-29.

Bailão AM, Schrank A, Borges CL, Dutra V, et al. (2006). Differential gene expression by *Paracoccidioides brasiliensis* in host interaction conditions: representational difference analysis identifies candidate genes associated with fungal pathogenesis. *Microbes Infect.* 8: 2686-2697.

Bailão AM, Shrank A, Borges CL, Parente JA, et al. (2007). The transcriptional profile of *Paracoccidioides brasiliensis* yeast cells is influenced by human plasma. *FEMS Immunol. Med. Microbiol.* 51: 43-57.

Bastos KP, Bailão AM, Borges CL, Faria FP, et al. (2007). The transcriptome analysis of early morphogenesis in *Paracoccidioides brasiliensis* mycelium reveals novel and induced genes potentially associated to the dimorphic process. *BMC Microbiol.* 7: 29.

Borges-Walmsley MI, Chen D, Shu X and Walmsley AR (2002). The pathobiology of *Paracoccidioides brasiliensis*. *Trends Microbiol.* 10: 80-87.

Conesa A, Götz S, García-Gómez JM, Terol J, et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

Costa M, Borges CL, Bailão AM, Meirelles GV, et al. (2007). Transcriptome profiling of *Paracoccidioides brasiliensis* yeast-phase cells recovered from infected mice brings new insights into fungal response upon host interaction. *Microbiology* 153: 4194-4207.

de Oliveira MV, Oliveira AC, Shida CS, de Oliveira RC, et al. (2013). Gene expression modulation by paraquat-induced oxidative stress conditions in *Paracoccidioides brasiliensis*. *Fungal Genet. Biol.* 60: 101-109.

Desjardins CA, Champion MD, Holder JW, Muszewska A, et al. (2011). Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. *PLoS Genet.* 7: e1002345.

Desler C, Suravajhala P, Sanderhoff M, Rasmussen M, et al. (2009). *In Silico* screening for functional candidates amongst hypothetical proteins. *BMC Bioinformatics* 10: 289.

Doerks T, van Noort V, Minguez P and Bork P (2012). Annotation of the *M. tuberculosis* hypothetical orfeome: adding functional information to more than half of the uncharacterized proteins. *PLoS One* 7: e34302.

Emanuelsson O, Brunak S, von Heijne G and Nielsen H (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2: 953-971.

Felipe MS, Andrade RV, Arraes FB, Nicola AM, et al. (2005). Transcriptional profiles of the human pathogenic fungus *Paracoccidioides brasiliensis* in mycelium and yeast cells. *J. Biol. Chem.* 280: 24706-24714.

Finn RD, Bateman A, Clements J, Coggill P, et al. (2014). The Pfam protein families database. *Nucleic Acids Res.* 42: 222-230

Franco M, Montenegro MR, Mendes RP, Marques SA, et al. (1987). Paracoccidioidomycosis: a recently proposed classification of its clinical forms. *Rev. Soc. Bras. Med. Trop.* 20: 129-132.

Goldman GH, dos Reis Marques E, Duarte Ribeiro DC, de Souza Bernardes LA, et al. (2003). Analysis of the human pathogen *Paracoccidioides brasiliensis* yeast phase: identification of putative homologues of *Candida albicans* virulence and pathogenicity genes. *Eukaryot. Cell* 2: 34-48.

Ijaq J, Chandrasekharan M, Poddar R, Bethi N, et al. (2015). Annotation and curation of uncharacterized proteins- challenges. *Front. Genet.* 6: 119.

Klein BS and Tebbets B (2007). Dimorphism and virulence in fungi. *Curr. Opin. Microbiol.* 10: 314-319.

Livak KJ and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 25: 402-408.

Loewenstein Y, Raimondo D, Redfern OC, Watson J, et al. (2009). Protein function annotation by homology-based inference. *Genome Biol.* 10: 207.

Matute DR, McEwen JG, Puccia R, Montes BA, et al. (2006). Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. *Mol. Biol. Evol.* 23: 65-73.

Medoff G, Painter A and Kobayashi GS (1987). Mycelial- to yeast-phase transitions of the dimorphic fungi *Blastomyces dermatitidis* and *Paracoccidioides brasiliensis*. *J. Bacteriol.* 169: 4055-4060.

Nunes LR, Costa de Oliveira R, Leite DB, da Silva VS, et al. (2005). Transcriptome analysis of *Paracoccidioides brasiliensis* cells undergoing mycelium-to-yeast transition. *Eukaryot. Cell* 4: 2115-2128.

Pigosso LL, Parente AF, Coelho AS, Silva LP, et al. (2013). Comparative proteomics in the genus *Paracoccidioides*. *Fungal Genet. Biol.* 60: 87-100.

Qiao J, Shao M, Chen L and Wang J (2013). Systematic characterization of hypothetical proteins *in Synechocystis* sp. PCC 6803 reveals proteins functionally relevant to stress responses. *Gene* 512: 6-15.

Restrepo A (1985). The ecology of *Paracoccidioides brasiliensis*: a puzzle still unsolved. *Sabouraudia* 23: 323-334.

Shahbaaz M, Hassan MI and Ahmad F (2013). Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* Rd KW20. *PLoS One* 8: e84263.

Teixeira MM, Theodoro RC, de Carvalho MJ, Fernandes L, et al. (2009). Phylogenetic analysis reveals a high level of speciation in the *Paracoccidioides genus*. *Mol. Phylogenet. Evol.* 52: 273-283.