Yeast
Yeast 2003; **20**: 263–271.
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/yea.964

Research Article

Transcriptome characterization of the dimorphic and pathogenic fungus *Paracoccidioides brasiliensis* by EST analysis

M. S. S. Felipe[1]*, R. V. Andrade[1], S. S. Petrofeza[2], A. Q. Maranhão[1], F. A. G. Torres[1], P. Albuquerque[1], F. B. M. Arraes[1], M. Arruda[5], M. O. Azevedo[1], A. J. Baptista[1], L. A. M. Bataus[2], C. L. Borges[2], E. G. Campos[1], M. R. Cruz[1], B. S. Daher[1], A. Dantas[1], M. A. S. V. Ferreira[1], G. V. Ghil[1], R. S. A. Jesuino[2], C. M. Kyaw[1], L. Leitão[1], C. R. Martins[1], L. M. P. Moraes[1], E. O. Neves[1], A. M. Nicola[1], E. S. Alves[1], J. A. Parente[2], M. Pereira[2], M. J. Poças-Fonseca[1], R. Resende[1], B. M. Ribeiro[1], R. R. Saldanha[1], S. C. Santos[1], I. Silva-Pereira[1], M. A. S. Silva[1], E. Silveira[1], I. C. Simões[1], R. B. A. Soares[2], D. P. Souza[1], M. T. De-Souza[1], E. V. Andrade[2], M. A. S. Xavier[1], H. P. Veiga[1], E. J. Venancio[3], M. J. A. Carvalho[1], A. G. Oliveira[1], M. K. Inoue[1], N. F. Almeida[4], M. E. M. T. Walter[1], C. M. A. Soares[2] and M. M. Brígido[1]

[1] *Laboratório de Biologia Molecular, Instituto de Ciências Biológicas and Departamento de Ciência da Computação, Universidade de Brasília, 70910-900 Brasília, DF, Brasil*
[2] *Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, Universidade Federal de Goiás, 74001-970 Goiânia, GO, Brasil*
[3] *Departamento de Ciências Patológicas, Centro de Ciências Biológicas, Universidade Estadual de Londrina, 86051-970 Londrina, PR, Brasil*
[4] *Departamento de Computação e Estatística, Centro de Ciências Exatas e Tecnologia, Universidade Federal do Mato Grosso do Sul, 79070-900 Campo Grande, MS, Brasil*
[5] *Departamento de Ciências Biológicas, Instituto de Ciências Naturais e Tecnológicas, Universidade do Estado de Mato Grosso, 78200-000 Cáceres, MT, Brasil*

**Correspondence to:*
Dr M. S. S. Felipe, Laboratório de Biologia Molecular, Instituto de Ciências Biológicas and Departamento de Ciência da Computação, Universidade de Brasília, 70910-900 Brasília, DF, Brasil.
E-mail: msueli@unb.br

Abstract

Paracoccidioides brasiliensis **is a pathogenic fungus that undergoes a temperature-dependent cell morphology change from mycelium (22 °C) to yeast (36 °C). It is assumed that this morphological transition correlates with the infection of the human host. Our goal was to identify genes expressed in the mycelium (M) and yeast (Y) forms by EST sequencing in order to generate a partial map of the fungus transcriptome. Individual EST sequences were clustered by the CAP3 program and annotated using Blastx similarity analysis and InterPro Scan. Three different databases, GenBank *nr*, COG (clusters of orthologous groups) and GO (gene ontology) were used for annotation. A total of 3938 (Y = 1654 and M = 2274) ESTs were sequenced and clustered into 597 contigs and 1563 singlets, making up a total of 2160 genes, which possibly represent one-quarter of the complete gene repertoire in *P. brasiliensis*. From this total, 1040 were successfully annotated and 894 could be classified in 18 functional COG categories as follows: cellular metabolism (44%); information storage and processing (25%); cellular processes — cell division, posttranslational modifications, among others (19%); and genes of unknown functions (12%). Computer analysis enabled us to identify some genes potentially involved in the dimorphic transition and drug resistance. Furthermore, computer subtraction analysis revealed several genes possibly expressed in stage-specific forms of *P. brasiliensis*. Further analysis of these genes may provide new insights into the pathology and differentiation of *P. brasiliensis*. All EST sequences have been deposited in GenBank under Accession Nos CA580326–CA584263. Copyright © 2003 John Wiley & Sons, Ltd.**

Received: 14 October 2002
Accepted: 28 November 2002

Keywords: **ESTs; transcriptome; functional genomics; human pathogenic fungus; dimorphism; *Paracoccidioides brasiliensis***

Introduction

Large-scale genome analysis has proved to be a valuable tool for unraveling gene function in several fungi, e.g. *Saccharomyces cerevisiae* (Goffeau *et al*., 1996), *Schizosaccharomyces pombe* (Wood *et al*., 2002), *Neurospora crassa* (Zhu *et al*., 2000; Braun *et al*., 2000), *Aspergillus nidulans* (Jeong *et al*., 2000; Prade *et al*., 2001), *Yarrowia lipolytica* (Casaregola *et al*., 2000) and *Candida albicans* (Tzung *et al*., 2001). The comparative study of the *S. cerevisiae* genome with the genomes of several human pathogenic fungi is leading to the discovery of important biological processes in these organisms, such as sexual cycle, meiosis, genetic recombination and mating. Strong evidence for the conservation of sexual functions in *C. albicans*, a dimorphic opportunistic human pathogenic fungus, was obtained by comparative genomic analysis with *S. cerevisiae* (Tzung *et al*., 2001). Furthermore, genomic analysis can reveal genes involved in pathogenicity and virulence, thus leading to new strategies for disease treatment.

An important and efficient approach for the comparative study of various genomes is the sequencing of expressed sequence tags (ESTs). ESTs represent the 5′ or 3′ ends of cDNAs and can generate the expression profile of different tissues, cell types or developmental stages (Adams *et al*., 1991; Verdun *et al*., 1998; Prade *et al*., 2001). EST sequences from over 250 different organisms have been deposited in large databases, known as dbEST (**http://www.ncbi.nlm.nih.gov/dbEST/**), which are used for data mining of homologous genes, detection of polymorphism and alternative splicing of mRNAs; identification of vaccine candidate molecules, new drug targets, gene prediction and expression studies (Pandey and Lewitter, 1999; Ohlrogge and Benning, 2000).

Paracoccidioides brasiliensis is a dimorphic human pathogenic fungus that is the aetiological agent of paracoccidioidomycosis (PCM), a systemic mycosis with restricted geographical distribution in Latin America which affects mainly rural workers and patients with immunological deficiency. Over 10 million people are estimated to be infected with *P. brasiliensis* but only up to 2% develop the disease (McEwen *et al*., 1995). The fungus grows as a multicellular mycelium at 26 °C or as multinucleated yeast cells at 36 °C. The phase transition can be triggered *in vitro* by a temperature shift from 26 °C to 36 °C or vice versa. It is generally assumed that the establishment of the human infection depends on the mycelium-to-yeast transition, which may reflect adaptive changes required to ensure survival of the pathogen in the warm-blooded human host (Franco, 1987). Although the molecular basis of the dimorphism in *P. brasiliensis* is largely unknown, several genes that may be involved in dimorphism and/or pathogenicity have recently been described in this fungus (Silva *et al*., 1999; Nino-Vega *et al*., 1998; Pereira *et al*., 2000; Izacc *et al*., 2001; Jesuino *et al*., 2002; Costa *et al*., 2002; Venancio *et al*., 2002b). Recently, Venancio *et al*. (2002a) identified three cDNA fragments (M73, M51 and M32) specific to the mycelial form and another two (MY-1 and MY-2) probably upregulated during the M to Y transition. Although these findings have greatly contributed to the understanding of the biology of *P. brasiliensis*, it is necessary to improve the identification and characterization of a larger number of expressed genes potentially involved in the dimorphism, virulence and pathogenicity, to ultimately unravel the mechanisms involved in the human host infection.

A network of laboratories from the central region of Brasil has been established to develop an EST genome project entitled 'Functional and Differential Genome of the Fungus *P. brasiliensis*' (**http://www.biomol.unb.br/Pb**), with the main goal of generating information about the mycelium and yeast transcriptomes of this fungus. In this study we describe the results of a partial transcriptome analysis comprising a total of 3938 ESTs and 1040 genes successfully annotated in *P. brasiliensis.*

Materials and methods

Microorganisms and growth conditions

The *P. brasiliensis* isolate *Pb*01 (ATCC-MYA-826) was grown as a mycelium at 22 °C or as a yeast at 36 °C on semi-solid Fava-Neto's medium, as previously described (Silva *et al*., 1994). *Escherichia coli* XL1-Blue MRF′ was used as a recipient for the phage cDNA libraries.

ESTs generation and sequencing

Non-normalized yeast and mycelium cDNA libraries were constructed in λZAPII (Stratagene, USA)

according to the manufacturer's instructions, using oligo dT-primed poly(A)$^+$ RNA. A total of 500 pfu of each library were used to infect *E. coli* XL1-Blue MRF′ cells and individual plaques were randomly picked and ordered into 96-well microtitre plates. Each well was previously filled with 100 µl SM buffer (100 mM NaCl, 8 mM $MgSO_4.7H_2O$, 50 mM Tris–HCl, 0.01% gelatin) and 5 µl chloroform. Phage elution was performed at 4 °C overnight. Inserts were amplified by PCR using eluted phages as templates. Briefly, PCRs were performed in a 50 µl reaction consisting of: 5 µl eluted phage ($\sim 5 \times 10^6$ phages); 0.04 mM each T3 and T7 universal primers; 0.2 mM dNTPs; 2.5 U *Taq* DNA polymerase (Cenbiot, Brasil); 5 µl 10 × *Taq* DNA polymerase buffer; 1.5 mM $MgCl_2$ and 0.1 mg/ml BSA. Thermal cycling was as follows: 93 °C/2 min; 35 cycles of 93 °C/1 min; 50 °C/1 min; 72 °C/1.5 min and a final extension at 72 °C/5 min. After the PCR amplification, 10 µl of the reactants were analysed in a 0.8% agarose gel. Sequencing reactions were performed with 1 µl PCR product (60–100 ng) following the dideoxide method, using the MegaBACE Dye Terminator procedure and analysed in a MegaBACE 1000 automatic sequencer (Amersham Biosciences, USA) at the Molecular Biology laboratories in the University of Brasília and Federal University of Goiás, Brasil.

EST assembly and computer analysis

Base calling and quality control of individual EST sequences were performed by PHRED (Ewing *et al*., 1998a,b). Vector sequences were screened out using Cross_Match (Green, 1996) before inclusion in the *P. brasiliensis* EST database. ESTs were assembled into clusters of overlapping sequences using the CAP3 program (Huang and Madan, 1999). All sequences with at least 200 nucleotides having a PHRED index greater than 20 were considered for clustering. The resulting contigs and singlets, hereafter denoted as groups, were preannotated using a pipeline that included automatic processing using Blastx and InterPro Scan. Three databases were used as input to Blastx (Altschul *et al*., 1997): GenBank *nr* (all non-redundant GenBank CDS; **http://www.ncbi.nlm.nih.gov**), COG (**http://www.ncbi.nlm.nih.gov/COG**) and GO (**http://ww.geneontology.org/**). Blastx analysis was performed using default parameters values. All files and databases (SMART, ProDom, PROSITE, PRINTS, TIGRFAMs, Pfam) used by InterPro Scan were obtained from the EBI website (**http://www.ebi.ac.uk/interpro/**). Functional annotation was based on this pipeline and confirmed by a human referee.

Additional analysis were performed using Blast programs (Altschul *et al*., 1997) for transcriptome comparisons with *S. cerevisiae* and *C. albicans*. Briefly, Blastx was used to compare each *P. brasiliensis* group against annotated ORFs of these two organisms. Conversely, TBlastn was used to compare *S. cerevisiae* and *C. albicans* ORFs with *P. brasiliensis* groups. A *P. brasiliensis* group and an ORF (or two ORFs) were considered orthologous if one of them was the other's best hit at an e-value lower than e-5, thus constituting a true 'bidirectional best hit' (BBH). On the other hand, a group (or an ORF) was considered specific for *P. brasiliensis*, with respect to the other organism, if it found no hits with e-value greater than e-3.

All computer analyses were performed on Intel-based computers (P4 and Xeon) using the Red Hat Linux 7.3 operation system. The scripts and programs were developed using PERL language and the web pages using Javascript and html.

Results and discussion

Generation and clustering of ESTs

For the production of ESTs we have used non-normalized cDNA libraries from the mycelium and yeast cells of *P. brasiliensis*. DNA fragments amplified from individual and randomly picked cDNA phage plaques showed size distribution in the range 0.5–3.5 kb. Typically, the yield of single-product PCR amplification from a 96-well microtitre plate was 90.5%; double-products represented 2.5% and unsuccessful amplification was 7%. Single-pass sequencing was performed from the 5′-end of the cDNA clones using the T3 sequencing primer.

We have initially sequenced 4992 ESTs and most of them (3938) were of high quality. We systematically accepted high-quality sequences having more than 200 nucleotides and a PHRED $\geq$ 20. In order to improve gene discovery we used CAP3 to cluster the EST data set. From a total of 3938 high-quality EST sequences, 2375 were clustered into 597 contigs with an average size of 897 nucleotides. The

total number of singlets (1563) plus 597 contigs resulted in 2160 groups, representing the number of genes expressed in the pathogenic fungus *P. brasiliensis* identified so far. The frequency of EST distribution after CAP3 analysis is shown in Figure 1. Most clusters (92.7%) contained 1–3 ESTs and the largest cluster contained 137 ESTs. The redundancy of the cDNA libraries, as estimated by the number of reads that were assembled into clusters, was approximately 60%, which is an acceptable value considering that a non-normalized cDNA library was used in this study.

We used an assembler program to reduce redundancy and to improve the quality of the sequences. We decided to use CAP3 because of its capacity for quickly clustering and assembling gene fragments. It is expected that CAP3 will create a number of clusters that represent multiple species of RNA, mainly for large families of closely related paralogues and for alternatively spliced genes (Telles and Silva, 2001). Therefore, for consistency, we decided to call the assembled cluster 'groups of related transcripts'. This is probably over-cautious for most fungal genes, since large families of paralogous genes represent only a small fraction of fungal genomes (Llorente *et al.*, 2000). Moreover, we have observed that CAP3 was able to discriminate related genes that were apparently polymorphic or alternatively spliced (data not shown).

The most abundant ESTs are shown in Table 1. Among these are the genes that have been previously shown to be highly expressed in mycelium cells, e.g. M51 and M32 (Venancio *et al.*, 2002a), and yeast cells, e.g. Y20 [Cunha *et al.*, 1999; Daher *et al.*, 2002. Expression and characterization of *Paracoccidioides brasiliensis* yeast-specific y20 gene (submitted)]. Despite having performed only single-pass sequencing of the 5′-end of the cDNAs, we identified 75 full-length ORFs after CAP3 assembly of groups of four or more reads with an average size of 626 nucleotides, e.g. formamidase, hsp90 and ubi4, among others (**https://www.biomol.unb.br/Pb**).

Table 1. Summary of the most abundant expressed genes in *P. brasiliensis* (only the 18 most frequent ESTs are shown)

No. ESTs/ frequency (%)	Gene annotation
137 (3.42)	*P. brasiliensis* M51
56 (1.40)	Formamidase
41 (1.02)	*P. brasiliensis* small subunit rRNA
33 (0.82)	Mono-oxygenase
30 (0.75)	Acid ribosomal protein P2
30 (0.75)	*P. brasiliensis* M32
25 (0.62)	*P. brasiliensis* Y20
24 (0.60)	Proline-rich protein
22 (0.55)	Nucleoside diphosphate kinase
20 (0.50)	Unknown
19 (0.47)	Low molecular weight heat shock protein (HSP30)
19 (0.47)	Lactoylglutathione lyase
18 (0.45)	Unknown
16 (0.40)	Cytochrome *c* oxidase related peptide
16 (0.40)	L24 ribosomal protein
16 (0.40)	Thioredoxin
15 (0.37)	Ty-like retro-element
15 (0.37)	Zinc-containing alcohol dehydrogenase

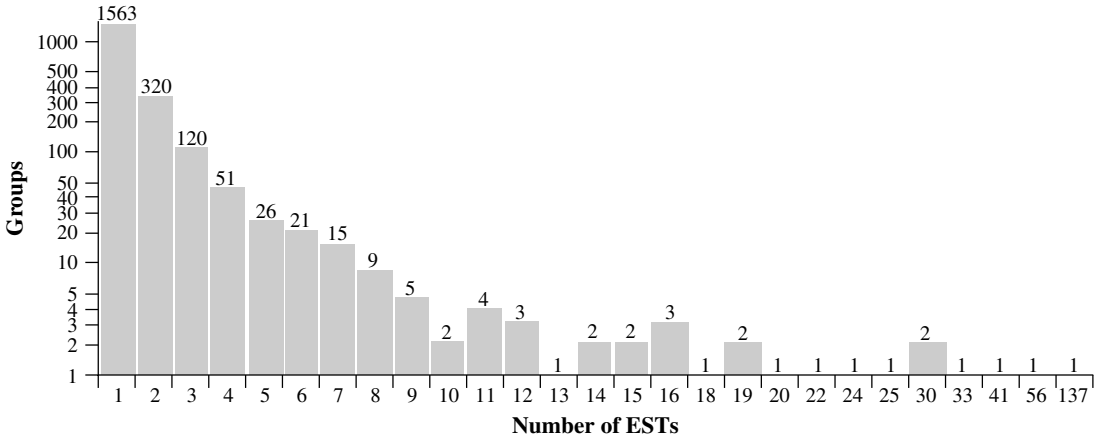

Figure 1. Distribution and number of clustered sequences. The frequency of EST distribution was obtained after CAP3 analysis. All sequences with at least 200 nucleotides having a PHRED $\geq$ 20 were considered for clustering. Most clusters (92.7%) contained 1–3 ESTs and the largest cluster was formed by 137 reads

Biological survey of identified genes in *P. brasiliensis*

As previously mentioned, we obtained a total of 2160 clusters: 1040 successfully annotated, 289 non-conclusively annotated and 831 not identified (Table 2). Group annotation was considered non-conclusive when either the e-value of blast hits was not significant, or when databases comparisons showed inconsistent results. Also, groups that matched only hypothetical genes were also denoted as non-conclusive. Therefore non-conclusively annotated groups were a set of genes clearly distinct from the non-identified groups, which were not annotated. COG annotation was also automatically generated and human revised. We classified 894 groups in 18 functional COG categories as follows: 44% corresponded to genes involved in cellular metabolism; 25% in information storage and processing; 19% involved in cellular processes (cell division, post-translational modifications, etc.) and 12% corresponded to genes with unknown functions (Figure 2 and Table 2). The most frequent gene categories are related to energy production and conversion (18%) and protein synthesis (15%). The COG categories are based mainly on bacterial genomic information and are not expected to cover all genes found in a eukaryotic organism.

Computer analysis identified several genes potentially involved in the dimorphic transition and drug resistance. Genes involved in signal transduction mechanisms are also represented and corresponded to 3% of the total number of annotated genes. MAP kinases, Ca^{2+}/calmodulin and cAMP/PKA signal transduction pathways have been described as having important roles in the cellular differentiation process of dimorphic and pathogenic fungi (Carvalho *et al*., 2002). Functional and genetic characterization of calmodulin from the dimorphic

Table 2. Annotation of clusters identified in *P. brasiliensis*. Annotated contigs and singlets. The clusters are shown as singlets and contigs

	No. annotated clusters[a]	No. non-conclusive clusters[b]	No. non-identified clusters[c]	Total no. clusters[d]	No. categorized clusters
Contigs	354	94	149	597	306
Singlets	686	195	682	1563	588
Total	1040	289	831	2160	894

[a] Confidently annotated and human curated.
[b] Non-conclusive annotation.
[c] Unable to be annotated.
[d] Sum of a,b and c.

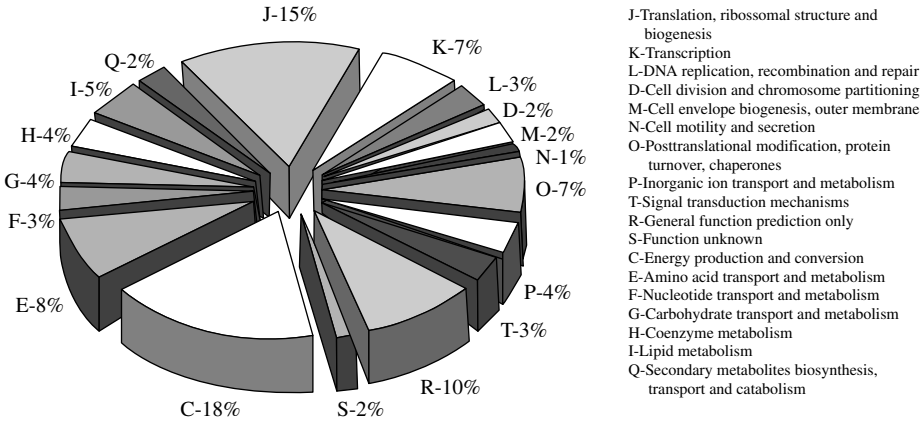

Figure 2. Functional classification of *P. brasiliensis* groups. This shows the percentage of predicted genes from *P. brasiliensis* according to their putative biological functions. A total of 894 non-redundant groups were classified into 18 COG categories. Pre-annotation was performed using automatic pipeline and functional annotation was confirmed by a human referee

and pathogenic fungus *Paracoccidioides brasiliensis* (submitted) have recently reported the inhibition of the mycelium-to-yeast transition using drugs that block the Ca^{2+} /calmodulin-dependent kinases. An increase in cAMP concentration was observed in *P. brasiliensis* during the mycelium-to-yeast transition, suggesting that the cAMP–PKA pathway is involved in this process (Paris and Duran, 1985). The identification of an EST encoding a PKA regulatory subunit in *P. brasiliensis* is interesting, but its functional significance needs to be established experimentally in this pathogen.

Multi-drug resistance proteins (MDR) are prime targets for the design of new drugs for the control of fungal infections. Among several protein families related to the multi-drug resistance phenotype are the ATP-binding cassettes ABC-type transporters (Young and Holland, 1999) involved in drug efflux. The human pathogen *C. albicans* has been the focus of drug therapy because this fungus is considered resistant to multidrug treatment. Five genes encoding for ABC-type transporters have been described in *C. albicans* and these findings should lead to the development of new drug therapy approaches. We have identified an EST similar (82% identity) to a MDR gene from *Emericella nidulans* in the mycelium form of *P. brasiliensis*. This EST probably represents an ABC-type transporter that could be a target for new drugs aiming at the infection control of this fungus.

EST representing members of the heat shock protein (HSP) gene families were identified (mycelium, HSP82, HSP10, HSP30 and HSP 88; yeast, HSP70, HSP60 and HSP104; in both: GroEL). These proteins could be involved in the thermoregulated dimorphic process because a HSP70 has been described as a differentially expressed gene during the M to Y transition of *P. brasiliensis* (Silva *et al*., 1999). In addition, HSP60 and HSP70 have been described as elicitors of the humoral and cellular immune responses in *Histoplasma capsulatum* infection (Gómez *et al*., 1992; Allendoerfer *et al*., 1996). HSP60 and HSP70 proteins from *P. brasiliensis* react with serum from patients with PCM (Izacc *et al*., 2001; Cunha *et al*., 2002). The ESTs identified in this work represent novel members of the HSP70 and HSP60 families because they were different to previously described genes from *P. brasiliensis* (Silva *et al*., 1999; Izacc *et al*., 2001). As mentioned above for the clustering procedure, it is not possible to determine whether a group represents a single species of heat shock protein (or its cognate) or even a mixture of transcripts from paralogous genes.

Polyubiquitin consists of an in tandem array of ubiquitin monomers (Ozkaynak *et al*., 1987) that is found in all eukaryotic organisms and participates in the intracellular protein turnover (Hershko and Ciechanover, 1998). In *Aspergillus nidulans* and *C. albicans*, the expression of the *ubi* genes is controlled during heat shock and other stress conditions (Noventa-Jordão *et al*., 2000; Roig *et al*., 2000). In plant pathogenic fungi, such as *Tuber borchii* and *Magnaporthe grisea*, the *ubi* genes are differentially expressed during dimorphism and infection, respectively (Zeppa *et al*., 2001; McCafferty and Talbot, 1998). We have identified a mycelium polyubiquitin EST but the role of this protein in *P. brasiliensis* has not yet been described, although we can postulate that it could be involved in the temperature-dependent dimorphic transition required for the establishment of the human infection.

We have also identified many other genes, including two groups encoding for prohibition homologues which may exhibit anti-proliferative activity controlling senescence and may also be involved in the maintenance and regulation of mitochondrial morphology. An EST related to a protein putatively involved in meiotic sister-chromatid recombination (Msc1p) was detected in the mycelium form, suggesting that *P. brasiliensis* is a diploid, reinforcing previous data (Cano *et al*., 1998). The detection of this EST is the first genetic evidence supporting the existence of a sexual cycle in *P. brasiliensis*. Eight different groups of retrotransposons were identified in the transcriptome of *P. brasiliensis*, including Ty-like elements, one of the most abundant ESTs described in this work (Table 1). Recombinational events involving dispersive Ty sequences can lead to inversions, deletions and translocations, which may account for the chromosomal polymorphism observed in different strains of *P. brasiliensis* (Oliveira *et al*., 2002). Functional analysis will result in important information about expression, cellular differentiation, pathogenicity and/or virulence of *P. brasiliensis*, but these issues can only be further addressed when genetic approaches such as gene disruption and RNA interference (siRNA) become available for this pathogen. A protocol for genetic transformation of *P. brasiliensis* has recently been developed

[Soares *et al*., 2002. Hygromycin B acquired phenotype in *Paracoccidioides brasiliensis* via plasmid DNA (submitted)], that will certainly help to perform genetic analysis in this system.

Genomic comparison

The partial transcriptome of *P. brasiliensis* was compared to both *S. cerevisiae* and *C. albicans* predicted proteomes. We have used the BBH to probe for the closest homologue in every compared pair of transcripts. Because BBH will detect the best hit in both directions, and because it tends to eliminate the interference from paralogous genes, we consider this pair of genes to be the orthologue between two genomic sets. From the 2160 *P. brasiliensis* groups compared, we have found 795 (47.7%) orthologues in *S. cerevisiae* and 868 (50.1%) in *C. albicans*. Once we are analysing a partial transcriptome from *P. brasiliensis*, the numbers obtained may be underestimated but, so far, our data suggest that *P. brasiliensis* has more genes in common with *C. albicans* than with *S. cerevisiae*. Because *C. albicans* and *P. brasiliensis* are human pathogens, shared genes between these species, and its absence in a non-pathogenic species such as *S. cerevisiae*, is of great interest when searching for disease-associated genes.

Identification of differentially expressed genes by computer subtraction

The *P. brasiliensis* EST database contains sequences from either yeast or mycelium origin. All sequences were submitted to CAP3 clustering. Morphotype specific groups were identified by computing the frequency of yeast or mycelium ESTs in each contig. Thus, to probe for differentially expressed mRNAs in the *P. brasiliensis* morphotypes, we considered the yeast : mycelium EST ratio in each contig as an indication of the expression level of the respective mRNA in each morphotype. A significant number of groups are shown to be specific by such criteria. A histogram showing the distribution of contigs is shown in Figure 3. In this Figure, we show the contigs composed of four or more ESTs in order to improve the significance of the analysis. We have found 14 yeast-specific groups, 12 of which are composed only of yeast-derived ESTs. On the other hand, 27 groups are considered to be mycelium-specific, 22 of which contain only mycelium-derived ESTs. Furthermore, preliminary experiments involving macroarray analysis, using cDNA probes generated from mRNA from mycelium and yeast cells (data not shown), showed that some *P. brasiliensis* genes (hydrophobin, M51 and M32) were highly expressed in mycelium cells, as indicated by EST computer subtraction. These data confirm previous results reported by Venancio *et al*. (2002a).

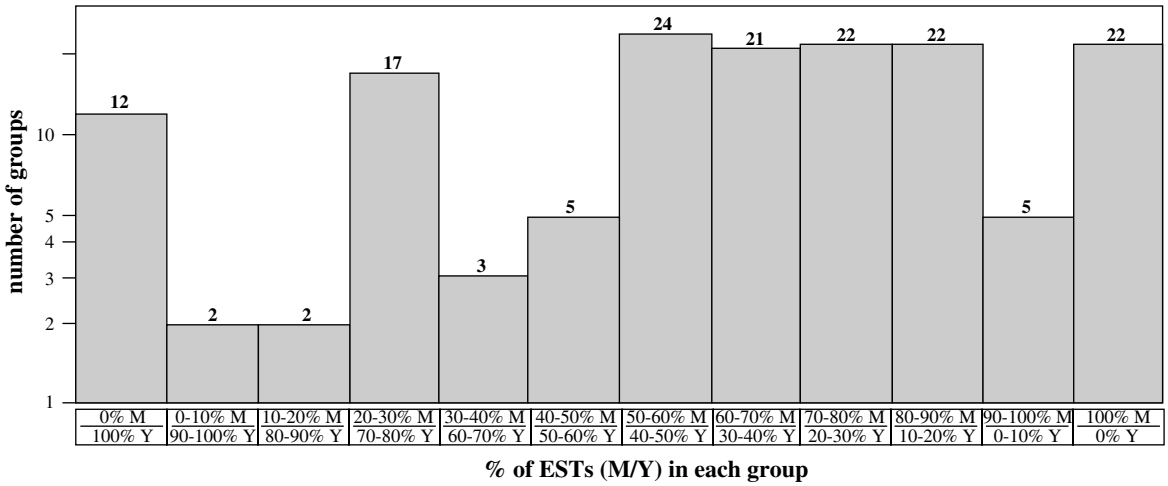

Figure 3. Morphotype stage-specific groups from mycelium and yeast cells of *P. brasiliensis*. The mycelium : yeast ratio in each of the groups suggests the expression level of the respective mRNA in each morphotype. The histogram shows the distribution of the specific groups in mycelium and yeast cells of *P. brasiliensis*. The groups are composed of four or more ESTs to improve the significance of the analysis

Together, our results suggest that, in *P. brasiliensis*, some genes are upregulated in mycelium cells; however, in view of the physiology, morphogenesis and life cycle of this fungus, the meaning of this observation remains to be investigated.

Acknowledgements

We thank Dr David John Bertioli for English review and M. F. L. Cesário and J. I. S. de Santana for laboratory assistance. This work was supported by MCT, CNPq and University of Brasília, DF, Brasil.

References

Adams MD, Kelley JM, Gocayne JD, *et al*. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**: 1651–1656.

Allendoerfer R, Maresca B, Deepe GS Jr. 1996. Cellular immune response to recombinant heat shock protein 70 from *Histoplasma capsulatum*. *Infect Immun* **64**: 4123–4128.

Altschul SF, Madden TL, Schaffer AA, *et al*. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Braun EL, Halpern AL, Nelson MA, Natvig DO. 2000. Large-scale comparison of fungal sequence information: mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*. *Genome Res* **10**: 416–430.

Cano MIN, Cisalpino PS, Galindo I, *et al*. 1998. Electrophoretic karyotypes and genome sizing of the pathogenic fungus *Paracoccidioides brasiliensis*. *J Clin Microbiol* **36**: 742–747.

Casaregola S, Neuvéglise C, Lépingle A, *et al*. 2000. Genomic exploration of the hemiascomycetous yeast: *Yarrowia lipolytica*. *FEBS Lett* **487**: 95–100.

Costa AA, Gómez FJ, Pereira M, *et al*. 2002. Characterization of a gene which encodes a mannosyltransferase homologue of *Paracoccidioides brasiliensis*. *Microbes Infect* **4**: 1027–1034.

Cunha AF, Sousa MV, Silva SP, *et al*. 1999. Identification, N-terminal region sequencing and similarity analysis of differentially expressed proteins in *Paracoccidioides brasiliensis*. *Med Mycol* **37**: 115–121.

Cunha DA, Zancopé-Oliveira RM, Felipe MSS, *et al*. 2002. Heterologous expression, purification and immunological reactivity of a recombinant HSP60 from *Paracoccidioides brasiliensis*. *Clin Diagn Lab Immuno* **l9**: 374–377.

Ewing B, Hillier L, Wendl MC, Green P. 1998a. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175–185.

Ewing B, Green P. 1998b. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**: 186–194.

Franco M. 1987. Host–parasite relationships in paracoccidioidomycosis. *J Med Vet Mycol* **25**: 5–18.

Goffeau A, Barrell BG, Bussey H, *et al*. 1996. Life with 6000 genes. *Science* **274**: 546, 563–567.

Gómez FJ, Gómez AM, Deepe GS Jr. 1992. An 80 kDa antigen from *Histoplasma capsulatum* that has homology to heat shock protein 70 induces cell-mediated responses and protection in mice. *Infect Immunol* **60**: 2565–2571.

Green P. 1996. PHRAP documentation. University of Washington, Seattle, USA. **http://bozeman.mbt.washington.edu/phrap.docs/phrap.html**.

Hershko A, Ciechanover A. 1998. The ubiquitin system. *Ann Rev Biochem* **67**: 425–479.

Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res* **9**: 868–877.

Izacc SMS, Gómez FJ, Jesuino RSA, *et al*. 2001. Molecular cloning, characterization and expression of the heat shock protein 60 gene from the human pathogenic fungus *Paracoccidioides brasiliensis*. *Med Mycol* **39**: 445–455.

Jeong H-Y, Han DM, Jahng K-Y, Chae K-S. 2000. The rpl16a gene for ribosomal protein L16A identified from expressed sequence tags is differentially expressed during sexual development of *Aspergillus nidulans*. *Fungal Genet Biol* **31**: 69–78.

Jesuino RS, Azevedo MO, Felipe MSS, Pereira M, Soares CMA. 2002. Characterization of a chaperone ClpB homologue of *Paracoccidioides brasiliensis*. *Yeast* **19**: 963–972.

Llorente B, Durrens P, Malpertuy A, *et al*. 2000. Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. *FEBS Lett* **487**: 122–133.

McCafferty HR, Talbot NJ. 1998. Identification of three ubiquitin genes of rice blast fungus *Magnaporthe grisea*, on of which is highly expressed during initial stages of plant colonization. *Curr Genet* **33**: 352–361.

McEwen JG, Garcia AM, Ortiz BL, Botero S, Restrepo A. 1995. In search of the natural habitat of *Paracoccidioides brasiliensis*. *Arch Med Res* **26**: 305–306.

Nino-Vega GA, Burman ET, Gooday GW, San-Blas G, Gow NA. 1998. Molecular cloning and sequencing of chitin syntase gene (CHS2) of *Paracoccidioides brasiliensis*. *Yeast* **14**: 181–187.

Noventa-Jordão MA, do Nascimento AM, Goldman MH, Terenzi HF, Goldman GH. 2000. Molecular characterization of ubiquitin genes from *Aspergillus nidulans*: mRNA expression of different stress and growth conditions. *Biochim Biophys Acta* **1490**: 237–244.

Ohlrogge J, Benning C. 2000. Unraveling plant metabolism by EST analysis. *Curr Opin Plant Biol* **3**: 224–228.

Oliveira AA, Alves VS, Moreira ESA. 2002. Genetic polymorphism of *Paracoccidioides brasiliensis* by PFGE and RAPD. *Ann Rev Biomed Sci.* (special issue) 39.

Ozkaynak E, Finley D, Solomon MJ, Varshavsky A. 1987. The yeast ubiquitin genes: a family of natural gene fusions. *EMBO J* **6**: 1429–1439.

Pandey A, Lewitter F. 1999. Nucleotide sequence databases: a gold mine for biologists. *Trends Biochem Sci* **24**: 276–280.

Paris S, Duran S. 1985. Cyclic adenosine 3′,5′ monophosphate (cAMP) and dimorphism in the pathogenic fungus *Paracoccidioides brasiliensis*. *Mycopathologia* **92**: 115–120.

Pereira M, Felipe MSS, Brígido MM, Soares CMA, Azevedo MO. 2000. Molecular cloning and characterization of a glucan synthase gene from the human pathogenic fungus *Paracoccidioides brasiliensis*. *Yeast* **16**: 451–462.

Prade RA, Ayoubi P, Krishnan S, Macwana S, Russell H. 2001. Accumulation of stress and inducer-dependent plant-cell-wall-degrading enzymes during asexual development in *Aspergillus nidulans*. *Genetics* **157**: 957–967.

Roig P, Martinez JP, Gil ML, Gozalbo D. 2000. Molecular cloning and characterization of the *Candida albicans* UBI3 gene coding for a ubiquitin–hybrid protein. *Yeast* **16**: 1413–1419.

Silva SP, Felipe MSS, Pereira M, Azevedo MO, Soares CMA. 1994. Phase transition and stage-specific protein synthesis in the dimorphic fungus *Paracoccidioides brasiliensis*. *Exp Mycol* **18**: 294–299.

Silva SP, Borges-Walmsley MI, Pereira IS, *et al*. 1999. Differential expression of a hsp70 gene during transition from the mycelial to the infective yeast form of the human pathogenic fungus *Paracoccidioides brasiliensis*. *Mol Microbiol* **31**: 1039–1050.

Telles GP, Silva FR. 2001. Trimming and clustering sugarcane ESTs. *Genet Mol Biol* **24**: 17–23.

Tzung KW, Williams RM, Scherer S, *et al*. 2001. Genomic evidence for a complete sexual cycle in *Candida albicans*. *Proc Natl Acad Sci USA* **98**: 3249–3253.

Venancio EJ, Kyaw CM, Mello CV, *et al*. 2002a. Identification of differentially expressed transcripts in the human pathogenic fungus *Paracoccidioides brasiliensis* by differential display. *Med Mycol* **40**: 45–51.

Venancio EJ, Daher BS, Andrade RV, *et al*. 2002b. The *kex2* gene from the dimorphic and human pathogenic fungus *Paracoccidioides brasiliensis*. *Yeast* **19**: 1221–1231.

Verdun RE, Di Paolo N, Urmenyi TP, *et al*. 1998. Gene discovery through expressed sequence tag sequencing in *Trypanosoma cruzi*. *Infect Immun* **66**: 5393–5398.

Wood V, Gwilliam R, Rajandream MA, *et al*. 2002. The genome sequence *of Schizosaccharomyces pombe*. *Nature* **415**: 871–880.

Young J, Holland B. 1999. ABC transporters: bacterial exporters revisited 5 years on. *Biochim Biophys Acta* **1461**: 177–200.

Zeppa S, Potenza L, Polidori E, *et al*. 2001. Cloning and characterization of a polyubiquitin gene from the ectomycorrhizal fungus *Tuber borchii* Vittad. *Curr Genet* **40**: 49–53.

Zhu H, Nowrousian M, Kupfer D, *et al*. 2000. Analysis of expressed tags from two starvation, time-of-day-specific libraries of *Neurospora crassa* reveals novel clock-controlled genes. *Genetics* **157**: 1057–1065.

