Arch Microbiol (2003) 179 : 339–353
DOI 10.1007/s00203-003-0533-8

ORIGINAL PAPER

**Kwok-yung Yuen · Géraldine Pascal · Samson S. Y. Wong
Philippe Glaser · Patrick C. Y. Woo · Frank Kunst
James J. Cai · Elim Y. L. Cheung · Claudine Médigue
Antoine Danchin**

Exploring the *Penicillium marneffei* genome

Received: 2 September 2002 / Revised: 17 February 2003 / Accepted: 17 February 2003 / Published online: 15 March 2003

Abstract *Penicillium marneffei* is a dimorphic fungus that intracellularly infects the reticuloendothelial system of humans and bamboo rats. Endemic in Southeast Asia, it infects 10% of AIDS patients in this region. The absence of a sexual stage and the highly infectious nature of the mould-phase conidia have impaired studies on thermal dimorphic switching and host-microbe interactions. Genomic analysis, therefore, could provide crucial information. Pulsed-field gel electrophoresis of genomic DNA of *P. marneffei* revealed three or more chromosomes (5.0, 4.0, and 2.2 Mb). Telomeric fingerprinting revealed 6–12 bands, suggesting that there were chromosomes of similar sizes. The genome size of *P. marneffei* was hence about 17.8–26.2 Mb. G+C content of the genome is 48.8 mol%. Random exploration of the genome of *P. marneffei* yielded 2303 random sequence tags (RSTs), corresponding to 9% of the genome, with 11.7, 6.3, and 17.4% of the RSTs having sequence similarity to yeast-specific sequences, non-yeast fungus sequences, and both (common sequences), respectively. Analysis of the RSTs revealed genes for information transfer (ribosomal protein genes, tRNA synthetase subunits, translation initiation, and elongation factors), metabolism, and compartmentalization, including several multi-drug-resistance protein genes and homologues of fluconazole-resistance gene. Furthermore, the presence of genes encoding pheromone homologues and ankyrin repeat-containing proteins of other fungi and algae strongly suggests the presence of a sexual stage that presumably exists in the environment.

Keywords *Penicillium marneffei* · Genome · Genomic analysis

K. Yuen (✉) · A. Danchin
HKU-Pasteur Research Centre, 8 Sassoon Road, Hong Kong
Tel.: +852-2816-8403, Fax: +852-2872-2782,
e-mail: kyyuen@hkucc.hku.hk

K. Yuen · S. S. Y. Wong · P. C. Y. Woo · J. J. Cai
E. Y. L. Cheung
Department of Microbiology, The University of Hong Kong,
University Pathology Building, Queen Mary Hospital,
Hong Kong

G. Pascal · A. Danchin
Unité Génétique des Génomes Bactériens, Institut Pasteur,
28 rue du Docteur Roux, 75724 Paris Cedex 15, France

P. Glaser · F. Kunst
Laboratory of Pathogenic Microbial Genomes, Institut Pasteur,
25 rue du Docteur Roux, 75724 Paris Cedex 15, France

C. Médigue
Genoscope and CNRS UMR-8030,
2 Rue Gaston Cremieux, CP 5706, 91057 Evry Cedex, France

Introduction

Penicillium marneffei is the most important thermal dimorphic fungus causing respiratory, skin and systemic mycosis in Southeast Asia (Yuen et al. 1994; Lo et al 1995; Kwan et al. 1997; Chim et al. 1998; Wong et al. 1999; Wong et al. 2001). Discovered in 1956 in hepatic abscesses of the Chinese bamboo rat *Rhizomys sinensis*, only 18 cases of human diseases were reported (in HIV-negative patients) until 1985 (Deng and Connor 1985). The appearance of the HIV pandemic, especially in Southeast Asian countries, saw the emergence of the infection as an important opportunistic mycosis in immunocompromised patients. About 10% of AIDS patients in Hong Kong are infected with *P. marneffei* (Wong and Lee 1998). In northern Thailand, penicilliosis is the third most common indicator disease of AIDS following tuberculosis and cryptococcosis (Supparatpinyo et al. 1994). Clinically, penicilliosis manifests as a systemic febrile illness, which results from intracellular infection of the reticuloendothelial cells by the yeast phase of the fungus and the associated inflammatory response of the host.

Despite its medical importance and its unusual thermal dimorphic capability, a large part of the ecology and epidemiology of *P. marneffei* remains unknown. The natural habitat of the fungus and its exact route of transmission have not been described. Molecular studies of this fungus at the molecular level have been limited. Only one cell-wall mannoprotein gene has been characterized and suc-

cessfully used in serodiagnosis and prevention of this infection (Cao et al. 1998a, b, 1999; Wong et al. 2001; Wong et al. 2002). Based on the mitochondrial and spacer rRNA, which allowed investigators to suggest a strong phylogenetic connection with *Talaromyces* species (LoBuglio and Taylor 1995), a PCR/hybridization assay was designed for molecular identification of this fungus in positive cultures (Vanittanakom et al. 1998).

P. marneffei is a model organism for understanding the molecular basis of thermal dimorphism. Given its propensity to cause disease in AIDS patients, the genome of *P. marneffei* may also provide insights into its pathogenic mechanisms and its possible interactions with the immune system. We describe in this report a random analysis of the genome of *P. marneffei*, which will facilitate further molecular research and lay the foundation for the complete genomic sequencing project of this fungus. A comprehensive knowledge of the genome will enable researchers to understand the basic mechanisms of thermal dimorphism, disease pathogenesis, virulence, and immune defense.

Materials and methods

Strains and DNA preparation

P. marneffei strain PM1 was isolated from an HIV-negative patient suffering from culture-documented penicilliosis in Hong Kong. In addition, ten additional strains of *P. marneffei*, isolated from the tissue and blood cultures of ten (7 HIV-positive and 3 HIV-negative) patients in Hong Kong, were used for electrokaryotyping. The arthroconidia ("yeast form") of PM1 was used throughout the DNA sequencing experiments. Genomic DNA was prepared from the arthroconidia grown at 37 °C. A single colony of the fungus grown on Sabouraud dextrose agar at 37 °C was inoculated into yeast peptone broth and incubated in a shaker at 30 °C for 3 days. Cells were cooled in ice for 10 min, harvested by centrifugation at 2,000×*g* for 10 min, washed twice and resuspended in ice-cold 50 mmol EDTA/l buffer (pH 7.5). Subsequently, 20 mg novazym/ml was added and incubated at 37 °C for 1 h followed by digestion in a mixture of 1 mg proteinase K/ml, 1% *N*-lauroylsarcosine, and 0.5 mol EDTA/l pH 9.5 at 50 °C for 2 h. Genomic DNA was then extracted by phenol, phenol-chloroform, and finally precipitated and washed in ethanol. After digestion with RNase A, a second ethanol precipitation was washed with 70% ethanol, air-dried and dissolved in 500 l of TE (pH 8.0).

Electrokaryotyping

The numbers and sizes of *P. marneffei* chromosomes in the 11 *P. marneffei* isolates were determined using pulsed-field gel electrophoresis. *P. marneffei* arthroconidia were grown for 3 days at 37 °C in yeast peptone broth as above. The harvested arthroconidia were washed and then inoculated into lyticase buffer containing 20 U of lyticase (Sigma). The protoplasts were then embedded in low melting point agarose plugs (2% prepared in isotonic solution and warmed to 50 °C) which were incubated in lyticase buffer at 37 °C for 1 h before treatment with lysis solution containing 1% *N*-laurylsarcosine and 1 mg proteinase K/ml. The mixture was incubated at 50 °C for 48 h.

Gels were cast using chromosomal-grade agarose (0.8%) in 0.5×TBE buffer and pulsed-field gel electrophoresis was carried out for 120 h at 12 °C and 2 V per cm, using a contour-clamped homogenous electric field and a CHEF Mapper XA System (Bio-Rad Laboratories, Hercules, Calif., USA). In the first 96 h, an included angle of 120° and pulse times of 3–60 min were used, and in the next 48 h, an included angle of 106° and pulse times of 3–20 min were used. *Saccharomyces cerevisiae* and *Hansenula wingei* molecular mass standards were used. Gels were stained with ethidium bromide and photographed with the Eagle System (Strategene, La Jolla, Calif., USA).

Telomeric fingerprinting

Genomic DNA of strain PM1 was digested to completion with *Eco*RI, *Eco*RV, *Hae*III, or *Sal*I and separated on a 0.8% (w/v) agarose gel. For southern hybridization, Hybond N^+ membranes (Amersham) were used according to the manufacturer's instructions. After hybridization at 50 °C for 5 h with the DIG-labeled (DIG Oligonucleotide 3' End Labeling Kit, Roche) probe $(TTAGGG)_6$, the nylon membrane was washed twice with 2×SSC/0.1% SDS (1×SSC is 0.15 M NaCl with 0.015 M sodium citrate) at room temperature for 10 min, followed by washing twice with 0.1× SSC/0.1% SDS at 50 °C for 15 min. Signal was detected according to the manufacturer's instructions.

Library construction

In order to make a library of *P. marneffei* DNA, 10 g of genomic DNA was partially digested by *Sau*3A. Fragments of 1–2 kb were gel purified and ligated onto the pBK-CMV vector at the *Bam*HI site. The ligation mix was used to transform *Escherichia coli* DH5α cells by electroporation. Bacteria were plated on LB medium containing 100 mg ampicillin/l; 5,0000 clones were obtained. One hundred clones were randomly picked and checking by miniprep, which confirmed that 98% of the clones has inserts of 800–2,300 bp.

DNA sequencing

DNA was sequenced using the chain-termination reaction according to published protocols (Frangeul et al. 1999) with slight modifications. To ensure high and even sequence quality for the entire set of RSTs, each sequencing profile was inspected on a Sparc II workstation using the Alfsplit and Ted programmes of the Staden package; sequences containing any base-calling ambiguity, or <100 nucleotides were eliminated. Suspected frameshifts-issued form errors in our single-read sequences detected by BLASTX comparisons or by using DNA-Strider dot plot matrices were corrected according to the sequence alignments. The average error rate of the RSTs was determined to be 0.5% for nucleotide substitution and 0.3% for base insertion or deletion, by repeated sequencing of the pCMV vector from empty clones.

Sequence analysis

All sequences were submitted as a batch file (FASTA format) to the LASSAP software (Glemet and Codani 1997) for a general BLAST (Altschul et al. 1990) study against the aggregated SwissProt+GenePept+PIR Banks. For the identification of the best hits, the non-redundant library SwissProt+TREMBL+TREMBL_New was used. The output was analyzed by a software constructed ad hoc, filtering data of interest (all sequences similar to known sequences were retained for further studies).

For the reconstitution of the rDNA genes, a specialized library of rDNA sequences was constructed using the World-Wide DNA Data Library (GenBank/EBI-EMBL/DDBJ). The rDNA sequences were collected and assembled for further study. The phylogenetic relationships of *P. marneffei* to other related species were determined using PileUp method with GrowTree (Genetics Computer Group). A total of 1,726 nucleotide positions of the 18S rRNA genes were included in the analysis.

Nucleotide sequence accession numbers

Nucleotide sequences reported in this article are deposited in the GenBank database under accession numbers AL683898 to AL686199.

Results

Physical characteristics of the genome and telomeric fingerprinting

The genomes of all 11 *P. marneffei* strains consist of three or more chromosomes (Fig. 1). The sizes of the three bands are about 5.0, 4.0, and 2.2 Mb. The results of telomeric fingerprinting of *P. marneffei* are shown in Fig. 2. Six to 12 bands were detected after digestion with *Eco*RI, *Eco*RV, *Hae*III, or *Sal*I, suggesting that chromosomes of similar sizes co-migrated (Wu et al. 1996). Assuming 12 telomeric fragments representing six chromosomes, the genome size of *P. marneffei* was hence about 17.8 (5.0+4.0+2.2+2.2+2.2+2.2) to 26.2 (5.0+5.0+5.0+5.0+4.0+2.2) Mb.

Random sequences

The overall G+C content of the genome sequences identified was 48.8 mol%. The 2,303 sequence tags that were generated had an average length of 773±103 bp. These

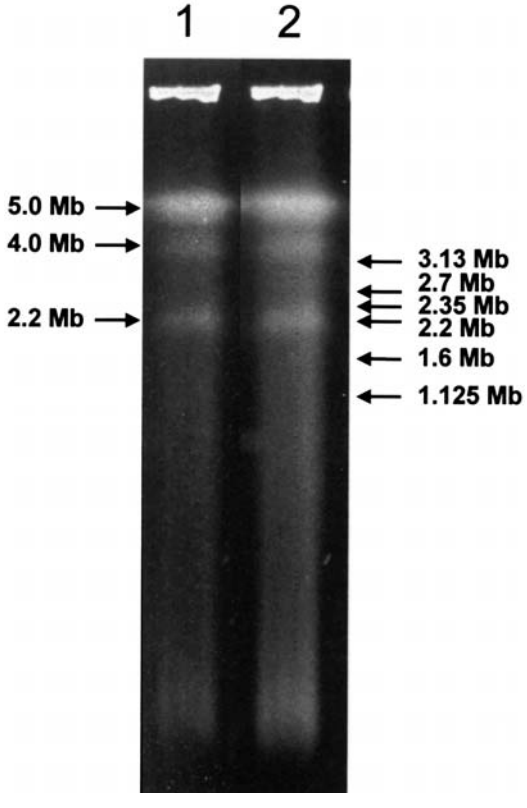

Fig. 1 Pulsed field gel electrophoresis of genomic DNA of *Penicillium marneffei*. *Lanes 1, 2 P. marneffei* strains PM1 and PM2, respectively. Sizes of markers (*Saccharomyces cerevisiae* and *Hansenula wingei* standards) are indicated on the *right*

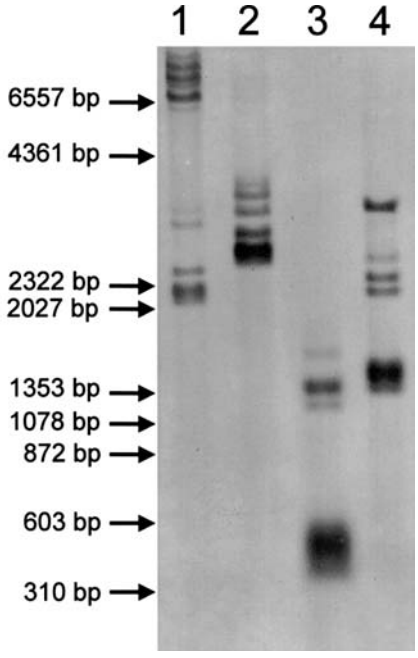

Fig. 2 Telomeric fingerprinting of *P. marneffei*. Six to 12 bands were detected after digestion with *Eco*RI (*lane 1*), *Eco*RV (*lane 2*), *Hae*III (*lane 3*), or *Sal*I (*lane 4*). Sizes of markers (λ *Hin*dIII digest and ΦX174 *Hae*III digest) are indicated on the *left*

2,303 sequence tags represented about 9% of the whole genome, if the genome size is 20 Mb. Table 1 lists the sequence tags of *P. marneffei* that show significant similarities to gene sequences in public databases. All major processes characterizing life are represented (Danchin 1989): metabolism (33%), information transfer (59%), and compartmentalization (8%) (Fig. 3).

One third of the genes are classified as primary metabolic genes, including genes coding for membrane-bound enzymes. They consist of genes for metabolism. Examples of genes involved in energy metabolism include cytochrome P450 enzymes, glucokinase, and enzymes of the glycolytic pathway. This is in line with the phenotypic observations showing that *P. marneffei* uses β-glucosides as carbon sources (Wong et al. 2001). Amino acid metabolism is represented by various aminotransferases and amino acid synthases; a number of enzymes involved in fatty-acid and phospholipids synthesis are also present, including a homologue of the lovastatin nonaketide synthase from *Aspergillus terreus*. Genes coding for an endoglucanase (which may be a cellulose) and a chitinase precursor are noted; these are likely to be involved in cell wall synthesis of the fungus. An interesting feature of the sequence tags identified here is that 3.4% code for secondary metabolism genes for non-ribosomal peptide synthesis and polyketide synthesis (this is likely to be an underestimate since polyketide synthases are usually highly repeated, and may have been removed from our sample as putative duplicates).

In the category of information transfer, six ribosomal protein genes as well as three tRNA synthetase subunits and five translation initiation and elongation factors (two

Table 1 Random sequence tags of *Penicillium marneffei* matched to known sequences in public databases. *sp* SWISS-PROT, *gb* GenBank

Functions	Organism	Description	GenBank accession no.	Accession no. of closest hit	*Penicillium marneffei* RST	E value
Cell cycle	*Aspergillus nidulans*	G1/S regulator	AL684013	sp/O93843	PM10E8.B	7.57 e^{-25}
	Aspergillus nidulans	Chromosome segregation SepB protein	AL684194	sp/Q00210	PM11E2.G	1.13 e^{-57}
	Saccharomyces cerevisiae	Ubiquitin-like protein DSK2	AL683940	sp/P48510	PM10B7.G	8.32 e^{-10}
	Saccharomyces cerevisiae	SIT4-associating protein SAP185	AL684015	sp/P40856	PM10E9.B	3.59 e^{-29}
	Saccharomyces cerevisiae	SIT4-associating protein SAP190	AL684016	sp/P36123	PM10E9.G	3.63 e^{-16}
	Schizosaccharomyces pombe	Cyclosome subunits	AL683936	sp/O42839	PM10B5.G	2.94 e^{-5}
	Schizosaccharomyces pombe	Pelota protein	AL684046	sp/Q9USL5	PM10G11.G	2.92 e^{-22}
Cell envelope	*Arabidopsis thaliana*	Mycolic acid methyl-transferase-like protein	AL684059	sp/Q9LUH5	PM10G7.B	4.55 e^{-10}
	Arabidopsis thaliana	Mycolic acid methyl-transferase-like protein	AL684060	sp/Q9LUH5	PM10G7.G	1.11 e^{-6}
	Arabidopsis thaliana	Putative pectin esterase	AL684185	sp/Q9SIJ9	PM11E1.B	9.44 e^{-11}
	Arabidopsis thaliana	Multi-spanning membrane protein	AL684292	sp/Q9LIC2	PM12A3.G	2.13 e^{-42}
	Beta vulgaris	Chitinase precursor	AL684111	sp/Q42421	PM11A9.B	3.89 e^{-14}
	Beta vulgaris	Chitinase precursor	AL684288	sp/Q42421	PM12A12.G	3.30 e^{-41}
	Beta vulgaris	Chitinase precursor	AL684297	sp/Q42421	PM12A6.B	2.05 e^{-38}
	Beta vulgaris	Chitinase precursor	AL684305	sp/Q42421	PM12B1.B	2.43 e^{-28}
	Beta vulgaris	Chitinase precursor	AL684310	sp/Q42421	PM12B11.G	2.03 e^{-28}
	Beta vulgaris	Chitinase precursor	AL684311	sp/Q42421	PM12B12.B	5.39 e^{-31}
	Beta vulgaris	Chitinase precursor	AL684319	sp/Q42421	PM12B5.B	1.63 e^{-34}
	Beta vulgaris	Chitinase precursor	AL684326	sp/Q42421	PM12B8.G	5.20 e^{-25}
	Beta vulgaris	Chitinase precursor	AL684351	sp/Q42421	PM12C9.B	2.58 e^{-26}
	Beta vulgaris	Chitinase precursor	AL684364	sp/Q42421	PM12D3.G	2.75 e^{-10}
	Candida glabrata	Cell wall synthesis protein KNH1 (cell wall β-1,6-glucan synthesis)	AL683994	sp/O74684	PM10E1.G	1.27 e^{-7}
	Entamoeba histolytica	Surface antigen ariel1	AL684259	sp/O96609	PM11H10.B	9.43 e^{-4}
	Homo sapiens	Mucin 2 precursor	AL684069	sp/Q02817	PM10H11.B	1.55 e^{-14}
	Volvox carteri	Sulfated surface glycoprotein SSG185	AL684056	sp/P21997	PM10G5.G	8.34 e^{-4}
Cellular metabolism						
Biosynthesis of cofactors	*Clostridium thermocellum*	Acetate kinase	AL683898	sp/O52594	PM10A1.B	2.96 e^{-24}
Carbohydrate metabolism	*Aspergillus nidulans*	Trehalose-6-phosphate synthase subunit 1	AL684282	sp/O59921	PM12A1.G	1.08 e^{-36}
Amino acid metabolism	*Aeropyrum pernix*	Dihydroxy-acid dehydratase	AL684193	sp/Q9YG88	PM11E2.B	5.98 e^{-12}
	Arabidopsis thaliana	NADH-dependent glutamate synthase	AL684347	sp/Q9LV03	PM12C7.B	1.51 e^{-55}
	Aspergillus nidulans	Homogentisate dioxygenase	AL683907	sp/Q00667	PM10A2.G	2.97 e^{-37}
	Aspergillus nidulans	Pentafunctional arom polypeptide (polyaromatic amino acid biosynthesis)	AL684053	sp/P07547	PM10G4.B	4.32 e^{-99}
	Aspergillus nidulans	Pentafunctional arom polypeptide (polyaromatic amino acid biosynthesis)	AL684054	sp/P07547	PM10G4.G	3.15 e^{-68}

Table 1 (continued)

Functions	Organism	Description	GenBank accession no.	Accession no. of closest hit	*Penicillium marneffei* RST	E value
	Cochliobolus carbonum	Branched chain amino acid aminotransferase	AL684166	sp/Q9Y885	PM11D11.G	8.64 e^{-16}
	Cochliobolus carbonum	Putative branched chain amino acid aminotransferase	AL684201	sp/Q9Y885	PM11E6.B	6.11 e^{-16}
	Cochliobolus heterostrophus	Polyketide synthase	AL684273	sp/Q92217	PM11H6.B	1.09 e^{-55}
	Escherichia coli	Arginase family	AL684071	sp/P16936	PM10H12.B	6.17 e^{-25}
	Lactococcus lactis	Aminotransferase	AL684114	sp/Q9CE18	PM11B1.G	1.72 e^{-15}
	Legionella pneumophila	Homogentisate dioxygenase	AL683911	sp/Q9S4T0	PM10A4.G	3.74 e^{-18}
	Saccharomyces cerevisiae	Acetolactate synthase small subunit	AL683948	sp/P25605	PM10C10.G	9.11 e^{-11}
	Saccharomyces cerevisiae	Glutamate synthase [NADPH] precursor	AL684348	sp/Q12680	PM12C7.G	2.64 e^{-47}
	Vibrio cholerae	2-Hydroxyacid dehydrogenase family	AL684117	sp/Q9KP72	PM11B11.B	5.73 e^{-27}
Energy metabolism	*Aspergillus nidulans*	Cytochrome P450	AL684154	sp/Q9Y7G5	PM11C6.G	1.39 e^{-13}
	Bacillus halodurans	N-acetylglucosamine-6-phosphate deacetylase	AL683980	sp/Q9KFQ7	PM10D3.G	5.26 e^{-17}
	Bacillus halodurans	Acetamidase	AL684270	sp/Q9KGN3	PM11H4.G	5.25 e^{-28}
	Boophilus microplus	Cytochrome P450	AL683972	sp/Q9Y1T8	PM10D10.G	4.64 e^{-23}
	Burkholderia cepacia	2,4-D dioxygenase	AL684226	sp/P96312	PM11F6.G	6.73 e^{-7}
	Coprinus cinereus	Cytochrome P450	AL684373	sp/O74643	PM12D8.B	6.93 e^{-7}
	Globodera pallida	NADH-ubiquinone oxidoreductase subunit 4	AL684327	sp/Q9T6M3	PM12B9.B	7.35 e^{-8}
	Glycine max	Cytochrome P450	AL684153	sp/O48928	PM11C6.B	2.29 e^{-7}
	Neurospora crassa	Mitochondrial NADH dehydrogenase	AL684228	sp/Q9Y7G7	PM11F7.G	2.15 e^{-56}
	Neurospora crassa	64 kDa mitochondrial NADH dehydrogenase	AL684296	sp/Q9Y7G7	PM12A5.G	9.74 e^{-59}
	Saccharomyces cerevisiae	6-Phosphogluconate dehydrogenase	AL684361	sp/P38720	PM12D2.B	4.63 e^{-25}
	Vibrio cholerae	Glucokinase regulatory protein	AL684022	sp/Q9KVE0	PM10F11.G	5.99 e^{-6}
Fatty acid and phospholipid metabolism	*Aspergillus fumigatus*	Inositol phosphoryl-ceramide synthase	AL683939	sp/Q9Y745	PM10B7.B	1.13 e^{-68}
	Aspergillus terreus	Lovastatin nonaketide synthase	AL684225	sp/Q9Y8A5	PM11F6.B	1.36 e^{-17}
	Aspergillus terreus	Lovastatin nonaketide synthase	AL684235	sp/Q9Y8A5	PM11G10.B	1.05 e^{-28}
	Aspergillus terreus	Lovastatin nonaketide synthase	AL684236	sp/Q9Y8A5	PM11G10.G	4.62 e^{-63}
	Avena sativa	UDP-glucose:sterol glucosyltransferase	AL684217	sp/O22678	PM11F2.B	2.22 e^{-35}
	Caenorhabditis elegans	Acyl-CoA dehydrogenase	AL684224	sp/Q19057	PM11F5.G	2.44 e^{-21}
	Neurospora crassa	Long-chain fatty acid CoA ligase	AL684087	sp/Q9P3D2	PM10H9.B	9.87 e^{-34}
	Neurospora crassa	Long-chain fatty acid CoA ligase	AL684088	sp/Q9P3D2	PM10H9.G	1.11 e^{-41}
Purines or pyrimidines	*Schizosaccharomyces pombe*	Ribonucleoside-diphosphate reductase small chain	AL684332	sp/P36603	PM12C10.G	1.39 e^{-56}

Table 1 (continued)

Functions	Organism	Description	GenBank accession no.	Accession no. of closest hit	*Penicillium marneffei* RST	E value
Others	*Acetobacter pasteurianus*	Carboxylesterase	AL684119	sp/O66374	PM11B12.B	1.80 e^{-5}
	Arthrobacter globiformis	Choline oxidase	AL683979	sp/Q59117	PM10D3.B	1.62 e^{-5}
	Aspergillus fumigatus	Catalase	AL683912	sp/P78574	PM10A5.B	1.95 e^{-78}
	Aspergillus nidulans	Acetyl-CoA carboxylase	AL684151	sp/O60033	PM11C5.B	3.41 e^{-59}
	Aspergillus nidulans	Alcohol dehydrogenase I	AL684354	sp/P08843	PM12D1.G	6.02 e^{-13}
	Bacillus stearothermophilus	Alcohol dehydrogenase	AL683998	sp/P42328	PM10E11.G	6.89 e^{-7}
	Bacillus subtilis	YVRD protein (short chain dehydrogenase/reductase)	AL683976	sp/O34782	PM10D12.G	4.83 e^{-5}
	Cladosporium herbarum	Aldehyde dehydrogenase	AL684360	sp/P40108	PM12D12.G	3.07 e^{-24}
	Homo sapiens	Short-chain dehydrogenases/reductase	AL684281	sp/Q9NRW0	PM12A1.B	4.16 e^{-6}
	Nectria haematococca mpVI	Aromatic-ring hydroxylase	AL684135	sp/Q01446	PM11B9.B	2.02 e^{-23}
	Neurospora crassa	Folylpolyglutamate synthetase	AL683975	sp/O13492	PM10D12.B	4.07 e^{-7}
	Ophiostoma novo-ulmi	Polygalacturonase	AL684159	sp/O59934	PM11C9.B	1.85 e^{-11}
	Pyrococcus abyssi	Chlorohydrolase	AL683982	sp/Q9V0Y5	PM10D4.G	3.12 e^{-4}
	Schizosaccharomyces pombe	Probable acid phosphatase	AL684049	sp/Q9USS6	PM10G2.B	1.63 e^{-16}
	Shewanella frigidimarina	Fumarate reductase flavoprotein subunit precursor	AL684051	sp/Q9Z4P0	PM10G3.B	9.31 e^{-17}
	Thermotoga maritima	Folylpolyglutamate synthetase	AL684143	sp/Q9WY13	PM11C12.B	1.21 e^{-10}
Cell signalling						
Receptor and their associated proteins	*Saccharomyces cerevisiae*	Pheromone receptor	AL683922	sp/P06842	PM10B1.G	4.75^{-14}
Protein kinase and phosphatase	*Arabidopsis thaliana*	Dual-specificity protein phosphatase	AL683901	sp/Q9M8K7	PM10A10.G	2.38 e^{-5}
	Arabidopsis thaliana	Histidine kinase 1 osmosensor	AL684195	sp/Q9SXL4	PM11E3.B	3.69 e^{-5}
	Candida albicans	Serine/threonine-protein kinase	AL683946	sp/Q92212	PM10C1.G	5.46 e^{-6}
	Saccharomyces cerevisiae	Probable serine/threonine-protein kinase	AL684172	sp/Q12399	PM11D3.G	8.90 e^{-5}
	Schizosaccharomyces pombe	Protein-tyrosine phosphatase (dual specificity protein phosphatase)	AL684242	sp/O13819	PM11G2.G	2.75 e^{-35}
Others	*Aspergillus nidulans*	Nuclear migration protein NUDF	AL684050	sp/Q00664	PM10G2.G	5.70 e^{-6}
	Neurospora crassa	GTP-binding protein YPT1	AL684157	sp/P33723	PM11C8.B	2.22 e^{-56}
	Neurospora crassa	G-protein β WD-40 repeats	AL684170	sp/Q9P6V7	PM11D2.G	2.44 e^{-4}
	Saccharomyces cerevisiae	Ankyrin repeat-containing protein Akr1p	AL683959	sp/P39010	PM10C5.B	1.06 e^{-7}
	Saccharomyces cerevisiae	Ankyrin repeat-containing protein Akr1p	AL683960	sp/P39010	PM10C5.G	5.18 e^{-7}
	Volvox carteri	Pherophorin-S precursor	AL684293	sp/P93797	PM12A4.B	4.75 e^{-8}
	Volvox carteri	Pherophorin-S precursor	AL684333	sp/P93797	PM12C11.B	1.98 e^{-32}
	Volvox carteri	Pherophorin-S precursor	AL684337	sp/P93797	PM12C2.B	2.85 e^{-25}
	Volvox carteri	Pherophorin-S precursor	AL684349	sp/P93797	PM12C8.B	6.45 e^{-18}
	Volvox carteri	Pherophorin-S precursor	AL684353	sp/P93797	PM12D1.B	2.95 e^{-18}
	Volvox carteri	Pherophorin-S precursor	AL684362	sp/P93797	PM12D2.G	2.75 e^{-6}
	Volvox carteri	Pherophorin-S precursor	AL684363	sp/P93797	PM12D3.B	1.90 e^{-13}

Table 1 (continued)

Functions	Organism	Description	GenBank accession no.	Accession no. of closest hit	*Penicillium marneffei* RST	E value
DNA replication and metabolism						
DNA replication, modification	*Aspergillus flavus*	O-methyltransferase	AL684183	sp/Q9P900	PM11D9.B	1.44 e^{-15}
	Saccharomyces cerevisiae	Mitochondrial membrane GTPase	AL683977	sp/P38297	PM10D2.B	5.45 e^{-40}
	Saccharomyces cerevisiae	Mitochondrial membrane GTPase	AL683978	sp/P38297	PM10D2.G	1.94 e^{-4}
	Schizosaccharomyces pombe	Replication factor-A protein 2	AL684123	sp/Q92373	PM11B3.B	4.09 e^{-17}
DNA repair	*Arabidopsis thaliana*	Ubiquitin-protein ligase 2	AL684025	sp/P42745	PM10F2.B	7.48 e^{-10}
	Pyrococcus kodakaraensis	Methylated DNA protein cysteine methyltransferase	AL683969	sp/O74023	PM10D1.B	2.25 e^{-9}
	Saccharomyces cerevisiae	DNA repair protein RAD5	AL684331	sp/P32849	PM12C10.B	5.53 e^{-9}
DNA binding	*Gallus gallus*	RING zinc-finger protein	AL684278	sp/Q90972	PM11H8.G	3.62 e^{-10}
	Trypanosoma cruzi	Kinetoplast-associated protein	AL684304	sp/Q26938	PM12A9.G	4.45 e^{-18}
Chromosomal structure	*Ensis minor*	Nuclear protein (linker histone H1)	AL684302	sp/Q24898	PM12A8.G	1.29 e^{-16}
	Ensis minor	Nuclear protein (linker histone H1)	AL684308	sp/Q24898	PM12B10.G	8.21 e^{-15}
	Ensis minor	Nuclear protein (linker histone H1)	AL684368	sp/Q24898	PM12D5.G	4.24 e^{-11}
	Schizosaccharomyces pombe	Chromosome region maintenance protein 1	AL683937	sp/P14068	PM10B6.B	2.79 e^{-54}
	Schizosaccharomyces pombe	Chromosome region maintenance protein 1	AL683938	sp/P14068	PM10B6.G	2.15 e^{-43}
	Schizosaccharomyces pombe	Histone H4	AL684175	sp/P09322	PM11D5.B	3.24 e^{-4}
RNA-directed DNA polymerase	*Aedes aegypti*	Pol-like protein (similar to RNA-directed DNA polymerase)	AL684213	sp/Q9U4W1	PM11F11.B	6.60 e^{-8}
	Neurospora crassa	Similar to RNA-directed DNA polymerase	AL684140	sp/Q01375	PM11C10.G	9.92 e^{-6}
	Neurospora crassa	Similar to RNA-directed DNA polymerase	AL684334	Q01375	PM12C11.G	1.42 e^{-5}
Intracellular trafficking	*Homo sapiens*	Transmembrane protein Tmp21 precursor	AL684256	sp/P49755	PM11G9.G	4.91 e^{-12}
	Neurospora crassa	Probable mitochondrial membrane dicarboxylate carrier protein	AL684082	sp/Q9P5U2	PM10H6.G	9.79 e^{-6}
	Saccharomyces cerevisiae	Protein transport protein Sec23	AL683984	sp/P15303	PM10D5.G	2.7 e^{-69}
	Schizosaccharomyces pombe	Importin subunit	AL683935	sp/O13864	PM10B5.B	4.02 e^{-64}
	Schizosaccharomyces pombe	Putative vacuolar biogenesis protein	AL684131	sp/Q9P6N4	PM11B7.B	2.02 e^{-6}
	Schizosaccharomyces pombe	Putative vacuolar biogenesis protein	AL684132	sp/Q9P6N4	PM11B7.G	4.91 e^{-9}
	Schizosaccharomyces pombe	Putative involvement in vesicular transport	AL684221	sp/Q9P6K0	PM11F4.B	3.36 e^{-8}
	Xenopus laevis	Nucleolar phosphoprotein	AL684222	sp/Q91803	PM11F4.G	1.51 e^{-6}
Membrane transport	*Amanita muscaria*	Sugar transporter	AL684177	sp/O13411	PM11D6.B	1.34 e^{-24}
	Arabidopsis thaliana	Sugar transporter	AL684261	sp/O23213	PM11H11.B	1.28 e^{-8}
	Botrytis cinerea	ABC transporter	AL684028	sp/O60034	PM10F3.G	3.29 e^{-91}
	Cochliobolus heterostrophus	Fatty-acid transporter protein	AL684113	sp/O42633	PM11B1.B	1.28 e^{-20}

Table 1 (continued)

Functions	Organism	Description	GenBank accession no.	Accession no. of closest hit	*Penicillium marneffei* RST	E value
	Gibberella pulicaris	MFS-multidrug resistance transporter	AL684204	sp/Q9P8F5	PM11E7.G	4.13 e^{-26}
	Homo sapiens	Calcium channel β2a subunit	AL684181	sp/Q9Y341	PM11D8.B	2.42 e^{-4}
	Mus musculus	Peroxisomal protein ALDR (ABC transporter)	AL684298	sp/Q61285	PM12A6.G	6.85 e^{-23}
	Mycobacterium avium	Molybdate uptake secreted protein	AL684161	sp/Q48919	PM11D1.B	3.47 e^{-4}
	Neurospora crassa	Amino acid permease 2	AL684102	sp/O59942	PM11A4.G	2.70 e^{-12}
	Saccharomyces cerevisiae	Purine-cytosine permease	AL683910	sp/P17064	PM10A4.B	1.18 e^{-23}
	Saccharomyces cerevisiae	Possible small-molecule transporter	AL683990	sp/P53134	PM10D8.G	2.93 e^{-37}
	Saccharomyces cerevisiae	Purine-cytosine permease	AL684099	sp/P17064	PM11A3.B	1.43 e^{-30}
	Saccharomyces cerevisiae	Fluconazole resistance protein 1	AL684103	sp/P38124	PM11A5.B	6.74 e^{-13}
	Schizosaccharomyces pombe	ABC transporter	AL683915	sp/P36619	PM10A6.G	6.43 e^{-22}
	Schizosaccharomyces pombe	Membrane transporter	AL684104	sp/O59700	PM11A5.G	1.74 e^{-39}
	Schizosaccharomyces pombe	Amino acid permease	AL684147	sp/Q9US40	PM11C3.B	1.39 e^{-60}
	Schizosaccharomyces pombe	Putative transporter of the allantoate permease family	AL684187	sp/Q10097	PM11E10.B	4.28 e^{-13}
	Schizosaccharomyces pombe	Probable membrane transporter	AL684188	sp/Q9US44	PM11E10.G	6.07 e^{-21}
	Schizosaccharomyces pombe	HMSF membrane transporter	AL684203	sp/O43081	PM11E7.B	4.95 e^{-33}
	Streptomyces fradiae	Transporter	AL684042	sp/Q9RP97	PM10G1.G	6.91 e^{-11}
Chaperone system	*Saccharomyces cerevisiae*	T-complex protein 1, α-subunit	AL683995	sp/P12612	PM10E10.B	2.62 e^{-23}
	Thermotoga maritima	Chaperone protein Dnaj	AL683961	sp/Q9WZV3	PM10C6.B	4.44 e^{-6}
Protein synthesis and degradation						
Ribosomal proteins	*Homo sapiens*	60S acidic ribosomal protein PO	AL684033	sp/Q9UKD2	PM10F6.B	4.12 e^{-11}
	Neurospora crassa	60S ribosomal protein L28	AL684045	sp/P08978	PM10G11.B	1.97 e^{-15}
	Schizosaccharomyces pombe	50S ribosomal protein	AL683930	sp/O94345	PM10B2.G	1.95 e^{-12}
RNA polymerase	*Caenorhabditis elegans*	DNA-directed RNA polymerase II largest subunit	AL684065	sp/P16356	PM10H1.B	2.13 e^{-5}
	Drosophila melanogaster	DNA-directed RNA polymerase II largest subunit	AL684211	sp/P04052	PM11F10.B	1.73 e^{-7}
RNA-binding proteins	*Mycobacterium tuberculosis*	Translation initiation factor 2	AL684306	sp/P71613	PM12B1.G	4.23 e^{-8}
	Nicotiana glutinosa	RNA-binding protein	AL683904	sp/O24106	PM10A12.B	2.66 e^{-6}
Aminoacyl-tRNA synthetase, tRNAs	*Homo sapiens*	Bifunctional aminoacyl-tRNA synthetase	AL683996	sp/P07814	PM10E10.G	2.48 e^{-83}
	Neurospora crassa	Valyl-tRNA synthetase	AL684355	sp/P28350	PM12D10.B	5.13 e^{-33}
	Neurospora crassa	Valyl-tRNA synthetase	AL684356	sp/P28350	PM12D10.G	5.34 e^{-33}
	Saccharomyces cerevisiae	tRNA synthetase	AL683919	sp/P38707	PM10A8.G	1.45 e^{-47}
Protein synthesis	*Tolypocladium inflatum*	Cyclosporin synthetase	AL683903	sp/Q09164	PM10A11.G	5.1 e^{-20}

Table 1 (continued)

Functions	Organism	Description	GenBank accession no.	Accession no. of closest hit	*Penicillium marneffei* RST	E value
Protein modification and translation factors	*Acanthamoeba castellanii*	Myosin I heavy-chain kinase	AL684328	sp/Q93107	PM12B9.G	1.88 e^{-18}
	Homo sapiens	Geranylgeranyl transferase type I β-subunit	AL684199	sp/P53609	PM11E5.B	1.35 e^{-5}
	Metarhizium anisopliae	Peptide synthetase	AL683902	sp/Q01135	PM10A11.B	1.75 e^{-60}
	Schizosaccharomyces pombe	Eukaryotic translation initiation factor	AL683957	sp/Q10425	PM10C4.B	2.74 e^{-37}
Degradation of proteins	*Aspergillus oryzae*	Alanyl dipeptidyl peptidase	AL684260	sp/Q9Y8E3	PM11H10.G	3.24 e^{-30}
	Caenorhabditis elegans	Ubiquitin fusion degradation protein 1 homologue	AL683963	sp/Q19584	PM10C7.B	9.33 e^{-4}
	Mus musculus	Peptide:N-glycanase	AL684043	sp/Q9JI78	PM10G10.B	3.13 e^{-35}
	Saccharomyces cerevisiae	Ubiquitin fusion degradation protein 1	AL684064	sp/P53044	PM10G9.G	7.48 e^{-6}
	Schizosaccharomyces pombe	26S Proteasome regulatory subunit	AL684128	sp/O74762	PM11B5.G	1.03 e^{-11}
	Schizosaccharomyces pombe	26S Proteasome regulatory subunit 12	AL684238	sp/O74440	PM11G11.G	8.14 e^{-12}
	Schizosaccharomyces pombe	26S Proteasome regulatory subunit MTS3	AL684344	sp/P50524	PM12C5.G	1.78 e^{-9}
	Sclerotinia sclerotiorum	Acid protease	AL684037	sp/Q9P8R1	PM10F8.B	6.0 e^{-4}
Transcription and mRNA regulation	*Aspergillus niger*	Transcription factor pacC	AL684253	sp/Q00203	PM11G8.B	1.21 e^{-31}
	Aspergillus niger	Transcription factor pacC	AL684254	sp/Q00203	PM11G8.G	2.62 e^{-37}
	Mus musculus	GA binding protein β-1 chain	AL683986	sp/Q00420	PM10D6.G	2.50 e^{-4}
	Saccharomyces cerevisiae	Transcriptional adaptor	AL683966	gb/NP010736	PM10C8.G	1.47 e^{-6}
	Schizosaccharomyces pombe	Transcriptional adaptor	AL683965	sp/Q9P7J7	PM10C8.B	6.83 e^{-24}
	Schizosaccharomyces pombe	Transcription factor ATF1 required for sexual development	AL684078	sp/P52890	PM10H4.G	2.84 e^{-5}
	Schizosaccharomyces pombe	BTB domain and ankyrin repeat-containing protein	AL684127	sp/O74881	PM11B5.B	2.52 e^{-23}
DEAD box proteins	*Saccharomyces cerevisiae*	ATP-dependent RNA helicase DOB1	AL684084	sp/P47047	PM10H7.G	2.27 e^{-26}
	Leishmania amazonensis	ATP-dependent RNA helicase	AL684124	sp/P90549	PM11B3.G	1.25 e^{-62}
Cytoskeleton	*Acanthamoeba castellanii*	Myosin IC heavy chain	AL684318	sp/P10569	PM12B4.G	2.30 e^{-9}
	Acanthamoeba castellanii	Myosin IA heavy chain	AL684320	sp/O77202	PM12B5.G	2.20 e^{-15}
	Acanthamoeba castellanii	Myosin IA	AL684372	sp/O77202	PM12D7.G	1.62 e^{-20}
	Aspergillus nidulans	Myosin I heavy chain	AL684174	sp/Q00647	PM11D4.G	3.74 e^{-35}
	Bos taurus (Bovine)	N-WASP	AL684375	sp/Q95107	PM12D9.B	5.56 e^{-19}
	Homo sapiens	WASP interacting protein	AL684365	sp/O43516	PM12D4.B	6.99 e^{-15}
	Homo sapiens	Diaphanous protein homologue 1	AL684367	sp/O60610	PM12D5.B	1.09 e^{-23}
	Mus musculus	Lymphocyte-specific formin-related protein	AL683968	sp/Q9Z2V7	PM10C9.G	9.83 e^{-6}
	Nectria haematococca mpVI	Kinesin	AL684125	sp/P78718	PM11B4.B	9.42 e^{-84}

Table 1 (continued)

Functions	Organism	Description	GenBank accession no.	Accession no. of closest hit	*Penicillium marneffei* RST	E value
Hypothetical proteins	*Arabidopsis thaliana*	46.1-kDa protein	AL683905	sp/Q9LF56	PM10A12.G	1.09 e^{-5}
	Arabidopsis thaliana	61.8-kDa TRP-ASP repeats-containing protein	AL684058	sp/O22212	PM10G6.G	5.55 e^{-23}
	Arabidopsis thaliana	gb|AAF34307.1	AL684289	sp/Q9LIR9	PM12A2.B	4.22 e^{-15}
	Caenorhabditis elegans	Coded for by *C. elegans* cDNA YK165E3.3	AL683934	sp/O02173	PM10B4.G	4.94 e^{-5}
	Caenorhabditis elegans	52.8-kDa protein	AL684257	sp/P30640	PM11H1.B	7.00 e^{-4}
	Escherichia coli	47.3-kDa protein	AL684137	sp/P75791	PM11C1.B	4.69 e^{-10}
	Mus musculus	Octapeptide-repeat protein T2	AL684321	sp/Q06666	PM12B6.B	1.10 e^{-4}
	Neurospora crassa	93.3-kDa protein	AL683947	sp/Q9P6B2	PM10C10.B	1.77 e^{-39}
	Saccharomyces cerevisiae	77.7-kDa protein	AL683992	sp/P47077	PM10D9.G	6.21 e^{-14}
	Saccharomyces cerevisiae	Possible role in ribosome biogenesis	AL684068	sp/Q04660	PM10H10.G	1.16 e^{-7}
	Saccharomyces cerevisiae	19.7-kDa protein	AL684077	sp/P36088	PM10H4.B	1.68 e^{-22}
	Saccharomyces cerevisiae	110.9-kDa protein	AL684081	sp/P53920	PM10H6.B	6.01 e^{-14}
	Saccharomyces cerevisiae	Integral membrane protein	AL684098	sp/P40468	PM11A2.G	6.48 e^{-28}
	Saccharomyces cerevisiae	Possible isomerase	AL684109	sp/Q12177	PM11A8.B	2.36 e^{-30}
	Saccharomyces cerevisiae	81.5-kDa integral membrane protein	AL684291	sp/P40071	PM12A3.B	4.09 e^{-28}
	Schizosaccharomyces pombe	52.3-kDa integral membrane protein	AL683916	sp/Q10254	PM10A7.B	1.85 e^{-9}
	Schizosaccharomyces pombe	21.3-kDa protein	AL683917	sp/O94329	PM10A7.G	7.39 e^{-10}
	Schizosaccharomyces pombe	170.7-kDa integral membrane protein	AL683943	sp/Q10250	PM10B9.B	4.43 e^{-76}
	Schizosaccharomyces pombe	68.1-kDa protein	AL683985	sp/O94419	PM10D6.B	1.42 e^{-6}
	Schizosaccharomyces pombe	170.7-kDa transmembrane protein	AL683987	sp/Q10250	PM10D7.B	2.53 e^{-76}
	Schizosaccharomyces pombe	WD-repeat protein	AL684011	sp/Q9USZ0	PM10E7.B	1.15 e^{-10}
	Schizosaccharomyces pombe	181.5-kDa protein	AL684023	sp/Q09853	PM10F12.B	2.92 e^{-24}
	Schizosaccharomyces pombe	16.6-kDa protein	AL684034	sp/Q9UUC8	PM10F6.G	1.43 e^{-13}
	Schizosaccharomyces pombe	Integral membrane protein	AL684063	sp/O94348	PM10G9.B	3.38 e^{-15}
	Schizosaccharomyces pombe	27.4-kDa protein	AL684133	sp/O14359	PM11B8.B	1.22 e^{-18}
	Schizosaccharomyces pombe	63.7-kDa protein	AL684146	sp/O94667	PM11C2.G	2.27 e^{-5}
	Schizosaccharomyces pombe	55.8-kDa protein	AL684271	sp/Q9URX1	PM11H5.B	2.77 e^{-29}

for the mitochondrial translation apparatus) were found. A significant number of regulatory protein genes involved in transcription, control of the cell cycle or in differentiation, in particular serine/threonine and tyrosine kinases and phosphatases as well as, interestingly, a class III adenylyl cyclase (Barzu and Danchin 1994), were also observed. The former part of this category, to which the significant number of sequences extracted from rDNA regions can be added, allowed a rough evaluation of the genome length (this cannot be done with the category of

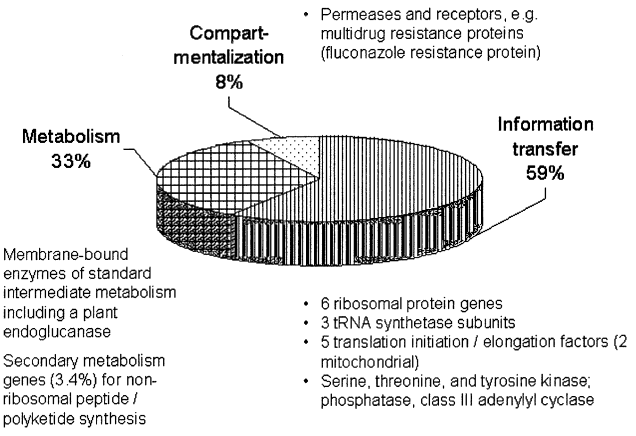

Fig. 3 Distribution of *P. marneffei* genes with regard to cellular functions

regulatory genes, which can be extremely variable from one organism to another).

Similar to other organisms, 8% of the sequences with identified function annotations presumably code for proteins involved in cell compartmentalization, in particular permeases and receptors. A large number of membrane transporters and permeases were found that are involved in the transport of carbohydrates, proteins, and fatty acids. Among them were several multi-drug-resistance protein genes, including a gene similar to that of a fluconazole resistance gene of *S. cerevisiae*.

One most intriguing observation from this sequence tags collection is the indications for the presence of a mating type and of mating pheromones in the genome. Similarities to the *S. cerevisiae* pheromone receptor gene *STE2* were found in some sequence tags, as well as similarities to pherophorins that bear sequence similarities to algal (*Volvox carteri*) sexual pheromone. Further evidence comes from the similarities to the *S. cerevisiae* ankyrin repeat-containing protein Akr1p which is involved in the yeast's pheromone response pathway and contributes to the control of cell shape and signal transduction (Kao et al. 1996; Pryciak and Hartwell 1996), although the expected values of 1.06 e^{-7} and 5.18 e^{-7} were not as high.

The genome appears to contain a large proportion of unknown and repeated sequences. However, there were not many transposase-related genes except for a counterpart of a transposase found in *Talaromyces stipitatus* (WWDDL accession number: CAA09449). Repeated sequences were often similar to segments of genes coding for proteins of the cell wall, or proteins involved in the cytoskeleton, as in the case of other yeasts and fungi. This is also a characteristic feature of pathogenic microorganisms. Among the repeated sequences were those that may code for proteins similar to the PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis* (Espitia et al. 1999). At this stage, however, it is not certain that these GC-rich regions indeed code for proteins (translation termination codons are statistically rare in such regions).

After removing duplicates, the random sequences were submitted to the LASSAP software against two specific databanks built for the purpose of this work—using the SRS tool (Stoesser et al. 2001)—written in the appropriate format: one containing all the known yeast sequences (YeastDB), and one containing the available fungus-non-yeast sequences (FungiDB). Taking the best hits, 269 matched only with yeast-specific sequences, 144 only with fungus-non-yeast sequences, and 400 matched with both (common sequences). The remaining sequence tags (1,183) matched with other sequences in the aggregated databanks.

Ribosomal DNA

The various fragments collected from rDNA loci can be assembled into contigs, allowing identification of most the sequence of the 18S and 28S RNAs of *P. marneffei*. Based on this knowledge, *P. marneffei* is likely to be an anamorph of a *Talaromyces* species. This substantiates the observation that the spacer regions of the rDNA loci are highly similar to that found in *Talaromyces* species (Kappe et al. 1996; Verweij et al. 1995). Indeed the sequence is almost identical with that of *T. flavus* (Fig. 4). It is also very similar to that of *Chromocleista cinnabarina*, a soil fungus that produces a red pigment, as does *P. marneffei* (Udagawa et al. 1973).

Discussion

Little is known about the ecology, transmission, or pathogenesis of *P. marneffei* infection. Although several species of bamboo rats are known to be carriers of the fungus, and, on rare occasions, the fungus has been isolated from bamboo rat burrows, it appears that they do not serve as significant reservoirs for human infection but are coincidentally infected from a common environmental source (Chariyalertsak et al. 1996, 1997). It is generally believed that *P. marneffei* exists in the soil of endemic areas and susceptible hosts—humans and bamboo rats—acquired the infection by inhaling the infectious conidia of the fungus, as in the case of other thermal dimorphic fungi.

The route of transmission of *P. marneffei*, however, has not been convincingly established. The main obstacle lies in the fact that the fungus has never been consistently isolated from any environmental samples, even in highly endemic areas. Conventional mycological culture relies on the identification of the fungus using the criteria of: (1) microscopic morphology of the reproductive structures, (2) elaboration of a diffusible red pigment, and (3) thermal dimorphism. However, this rests on the premise that the fungus exists in the familiar mould form in nature. Morphological identification of *P. marneffei* in environmental samples would fail if the saprophytic form is indeed a teleomorph, which may have a very different morphology. A teleomorph for *P. marneffei* has never been described, although our analysis of the genome does suggest this possibility. The presence of homologues for fun-

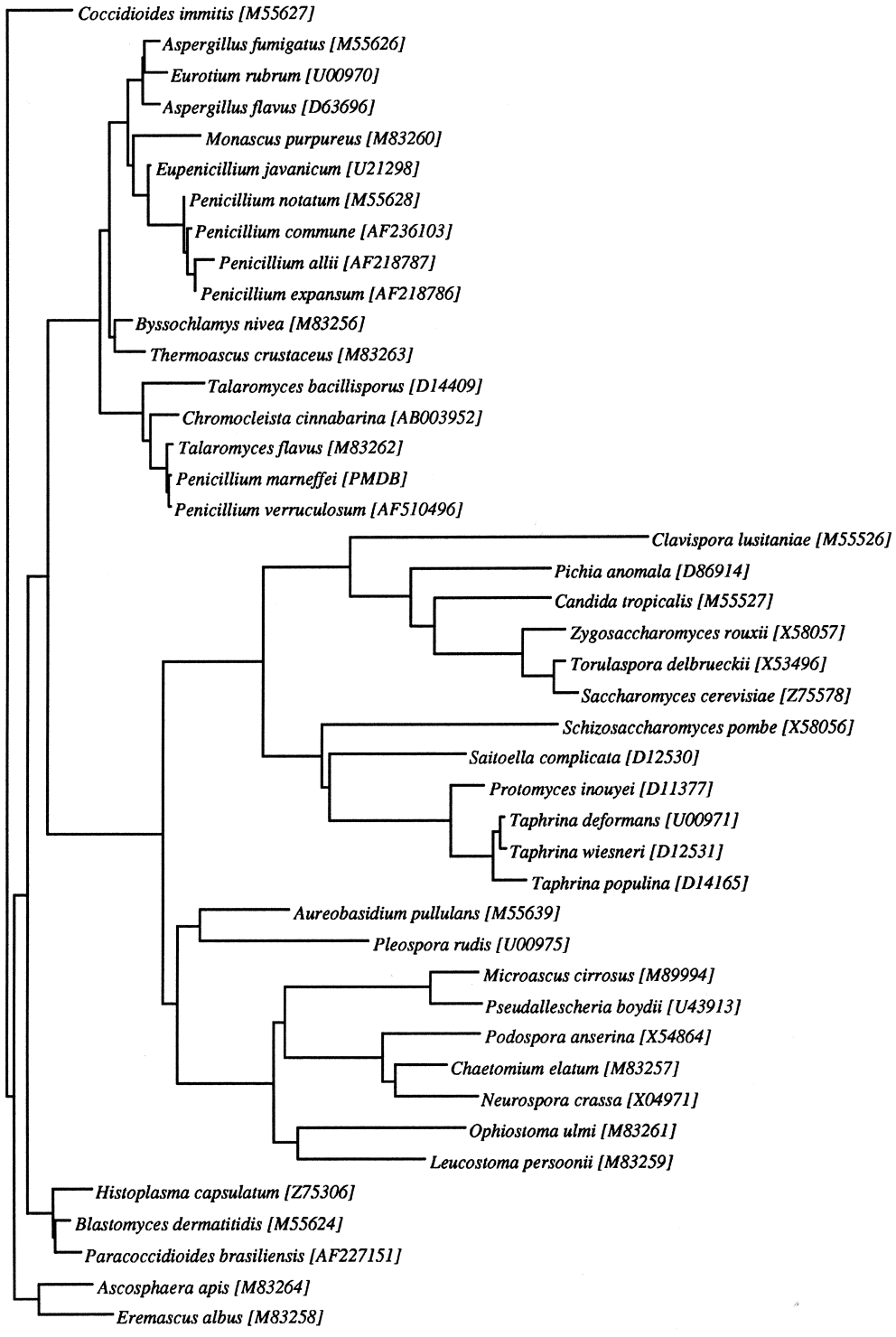

Fig. 4 Phylogenetic tree showing the relationships of *P. marneffei*. to other *Penicillium* and *Talaromyces* species. The tree was inferred from 18S rRNA data by the neighbor-joining method. *Scale bar* Estimated number of substitutions per 100 bases using the Jukes-Cantor correction. Names and accession numbers are given as cited in the GenBank database

gal and algal pheromone as well as molecules involved in the *S. cerevisiae* pheromone response pathway is highly suggestive of the presence of a sexual stage in the life cycle of *P. marneffei*. A further line of evidence comes from the finding that an *STE12* homologue of *P. marneffei* restores the defect in sexual development of *Aspergillus nidulans steA* mutant (Borneman et al. 2001). It is generally accepted that organisms without some sort of sexuality are rare, because sex-associated gene recombination or reassortment processes are the only way to escape the fate of Muller's ratchet, leading to degeneracy and ultimately to disappearance (Kondrashov 1994). It is therefore to be expected that some form of *P. marneffei* should involve mating. Indeed, teleomorphs of several other *Penicillium* species have been described and classified under *Eupenicillium* and *Talaromyces*, an example being *Penicillium*

Table 2 Karyotypes and genome sizes of *P. chrysogenum*, *P. notatum*, *P. nalgiovense*, *P. janthinellum*, *P. paxilli*, *P. purpurogenum*, and *P. marneffei*

Penicillium species	Number of chromosomes	Genome sizes	References
P. chrysogenum	4	34.1	Fierro et al. 1993
P. notatum	4	32.1	Fierro et al. 1993
P. nalgiovense	4	26.5	Farber and Geisen 2000
P. janthinellum	8	39.0–49.0	Kayser and Schulz 1991
P. paxilli	8	–	Young et al. 1998
P. purpurogenum	5	21.2	Chavez et al. 2001
P. marneffei	3–6	17.8–26.2	Present study

emersonii which is the anamorph of *Talaromyces emersonii*, a thermophilic fungus usually isolated from soil (Cimon et al. 1999). The phylogenetic position of *P. marneffei* was recently studied using nuclear and mitochondrial ribosomal DNA sequences. Results of the study placed *P. marneffei* closely related to *Talaromyces*. It is also known that in Ascomycetes (e.g. *Talaromyces*) the haploid forms may differ widely, with the diploid form being transient. As a consequence it is possible that the pathogenic stage of the fungus is just the haploid form of a thermotolerant saprophyte with a short-lived sexual reproduction.

The genome size of *P. marneffei* is relatively small compared to other *Penicillium* species (Chavez et al. 2001). The karyotypes and genome sizes of six *Penicillium* species (*P. chrysogenum*, *P. notatum*, *P. nalgiovense*, *P. janthinellum*, *P. paxilli*, and *P. purpurogenum*) have been published (Table 2). The chromosome number varies from four to eight while the estimated genome size varies from 21.2 to 49.0 Mb. The genome size of *P. marneffei* is small compared to these six *Penicillium* species. The significance of this is unknown. One possible explanation could be that *P. marneffei*, at some stage of its life cycle, is an obligate parasite in susceptible hosts; in fact, it is the only species of *Penicillium* that consistently causes disease in humans. Compared to purely saprophytic species of *Penicillium*, some of the genes may have been lost during evolution. Other obligate parasites such as *Mycoplasma* (Fraser et al. 1995, 1998; Himmelreich et al. 1996), *Treponema pallidum* (Fraser et al. 1998), and *Mycobacterium leprae* (Cole et al. 2001) also characteristically possess relatively small genome sizes compared to their free-living counterparts or species with more elaborate life cycles. Whether this is the case for *P. marneffei* needs to be confirmed by comparative genomic studies with other related *Penicillium* and *Talaromyces* species.

The unique virulence and pathogenicity of *P. marneffei* amongst the otherwise saprophytic *Pencillium* genus are unexplained. The crux of this is the propensity to cause disease in patients with impaired cellular immunity, with AIDS patients being the largest at-risk population. The thermal tolerance of *P. marneffei* is definitely essential for its virulence, but detailed pathogenic mechanisms have not been described, apart from reports on its interactions with leukocytes and adhesion of the conidia to laminin (Hamilton et al. 1998, 1999). The presence of a large number of thioester-mediated non-ribosomal protein synthesis or reduced carbon-chain carboxylate intermediates (polyketides or related molecules) suggests a very rich secondary metabolism (Stachelhaus et al. 1995), as found in *Streptomyces* and other saprophytic organisms where these metabolites are presumably used in complex regulatory pathways. While these molecules may have their role in signaling pathways with concomitant adaptation in their hosts, they may play a more novel role in modulating the immune responses of the host and hence have a crucial role in pathogenesis. Important examples include homologues to lovastatin nonaketide synthase of *Aspergillus terreus* and, more interestingly, cyclosporin synthetase of *Tolypocladium niveum*. Polyketide and the non-ribosomal peptides are two large families of compounds that include many clinically important antimicrobials (e.g. erythromycin, oleandomycin, vancomycin) and immunosuppressants (e.g. cyclosporin, tacrolimus, sirolimus), and cytotoxic agents (e.g. doxorubicin, bleomycin, epothilones). Most of these compounds, including the macrolides, possess immunomodulating (predominantly immunosuppressive) or cytotoxic activities. Cyclosporin A, for example, is noted for its T-cell immunosuppressive activities and hence is clinically important for various anti-rejection treatments in transplant recipients. It is well known that many intracellular pathogens (e.g. *Leishmania* species) actively modulate host cytokine production and/or Th1/Th2 cellular immune responses to enhance their survival (Alexander et al. 1999). It would, therefore, not be surprising if *P. marneffei* also utilizes similar strategies to facilitate its persistence inside susceptible hosts.

Finally, there are a large number of membrane transport proteins in *P. marneffei*. Among these is fluconazole resistance protein 1, encoded by the *FLU1* gene. Fluconazole resistance protein is a member of the major facilitator superfamily of multidrug efflux transporter. This is in accord with the observed fluconazole resistance of *P. marneffei* (Imwidthaya et al. 2001).

Acknowledgements This work was supported by the collaboration between The University of Hong Kong and the Institute Pasteur and was partly funded by the University Development Fund, Research Grants Council, and AIDS Trust Fund, Hong Kong.

References

Alexander J, Satoskar AR, Russell DG (1999) Leishmania species: models of intracellular parasitism. J Cell Sci 112:2993–3002

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Barzu, O, Danchin A (1994) Adenylyl cyclases: a heterogeneous class of ATP-utilizing enzymes. Prog Nucleic Acid Res Mol Biol 49:241–283

Borneman AR, Hynes MJ, Andrianopoulos A (2001) An STE12 homolog from the asexual, dimorphic fungus *Penicillium marneffei* complements the defect in sexual development of an *Aspergillus nidulans* steA mutant. Genetics 157:1003–1014

Cao L, Chan CM, Lee C, Wong SS, Yuen KY (1998) *MP1* encodes an abundant and highly antigenic cell wall mannoprotein in the pathogenic fungus *Penicillium marneffei*. Infect. Immun 66:966–973

Cao L, Chen DL, Lee C, Chan CM, Chan KM, Vanittanakom N, Tsang DN, Yuen KY (1998) Detection of specific antibodies to an antigenic mannoprotein for diagnosis of *Penicillium marneffei* penicilliosis. J Clin Microbiol 36:3028–3031

Cao L, Chan KM, Chen D, Vanittanakom N, Lee C, Chan CM, Sirisanthana T, Tsang DN, Yuen KY (1999) Detection of cell wall mannoprotein Mp1p in culture supernatants of *Penicillium marneffei* and in sera of penicilliosis patients. J Clin Microbiol 37:981–986

Chariyalertsak S, Vanittanakom P, Nelson KE, Sirisanthana T, Vanittanakom N (1996) *Rhizomys sumatrensis* and *Cannomys badius*, new natural animal hosts of *Penicillium marneffei*. J Med Vet Mycol 34:105–110

Chariyalertsak S, Sirisanthana T, Supparatpinyo K, Praparattanapan J, Nelson KE (1997) Case-control study of risk factors for *Penicillium marneffei* infection in human immunodeficiency virus-infected patients in northern Thailand. Clin Infect Dis 24:1080–1086

Chavez R, Fierro F, Gordillo F, Francisco Martin J, Eyzaguirre J (2001) Electrophoretic karyotype of the filamentous fungus *Penicillium purpurogenum* and chromosomal location of several xylanolytic genes. FEMS Microbiol Lett 205:379–383

Chim CS, Fong CY, Ma SK, Wong SS, Yuen KY (1998) Reactive hemophagocytic syndrome associated with *Penicillium marneffei* infection. Am J Med 104:196–197

Cimon, B, Carrere J, Chazalette JP, Vinatier JF, Chabasse D, Bouchara JP (1999) Chronic airway colonization by *Penicillium emersonii* in a patient with cystic fibrosis. Med Mycol 37: 291–293

Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall D, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, and Barrell BG (2001) Massive gene decay in the leprosy bacillus. Nature 409:1007–1011

Danchin A (1989) Homeotopic transformation and the origin of translation. Prog Biophys Mol Biol 54:81–86

Deng ZL, Connor DH (1985) Progressive disseminated penicilliosis caused by *Penicillium marneffei*. Report of eight cases and differentiation of the causative organism from *Histoplasma capsulatum*. Am J Clin Pathol 84:323–327

Espitia C, Laclette JP, Mondragon-Palomino M, Amador A, Campuzano J, Martens A, Singh M, Cicero R, Zhang Y, Moreno C (1999) The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? Microbiology 145:3487–3495

Farber P, Geisen R (2000) Karyotype of *Penicillium nalgiovense* and assignment of the penicillin biosynthetic genes to chromosome IV. Int J Food Microbiol 58:59–63

Fierro F, Gutierrez S, Diez B, Martin JF (1993) Resolution of four large chromosomes in penicillin-producin filamentous fungi: the penicillin gene cluster is located on chromosome II (9.6 Mbp) in *Penicillium notatum* and chromosome I (10.4 Mbp) in *Penicillium chrysogenum*. Mol Gen Genet 241:573–578

Frangeul L, Nelson KE, Buchrieser C, Danchin A, Glaser P, Kunst F (1999) Cloning and assembly strategies in microbial genome projects. Microbiology 145:2625–2634

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270:397–403

Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, Gwinn M, Hickey EK, Clayton R, Ketchum KA, Sodergren E, Hardham JM, McLeod MP, Salzberg S, Peterson J, Khalak H, Richardson D, Howell JK, Chidambaram M, Utterback T, McDonald L, Artiach P, Bowman C, Cotton MD, Venter JC (1998) Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science 281:375–388

Glemet E, Codani JJ (1997) LASSAP, a LArge Scale Sequence compArison Package. Comput Appl Biosci 13:137–143

Hamilton AJ, Jeavons L, Youngchim S, Vanittanakom N, Hay RJ (1998) Sialic acid-dependent recognition of laminin by *Penicillium marneffei* conidia. Infect Immun 66:6024–6026

Hamilton AJ, Jeavons L, Youngchim S, Vanittanakom N (1999) Recognition of fibronectin by *Penicillium marneffei* conidia via a sialic acid-dependent process and its relationship to the interaction between conidia and laminin. Infect Immun 67:5200–5205

Himmelreich R., Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucleic Acids Res 24:4420–4449

Imwidthaya P, Thipsuvan K, Chaiprasert A, Danchaivijitra S, Sutthent R, Jearanaisilavong J (2001) *Penicillium marneffei*: types and drug susceptibility. Mycopathologia 149:109–115

Kao LR, Peterson J, Ji R, Bender L, Bender A (1996) Interactions between the ankyrin repeat-containing protein Akr1p and the pheromone response pathway in *Saccharomyces cerevisiae*. Mol Cell Biol 16:168–178

Kappe R., Fauser C, Okeke CN, Maiwald M (1996) Universal fungus-specific primer systems and group-specific hybridization oligonucleotides for 18S rDNA. Mycoses 39:25–30

Kayser T, Schulz G (1991) Electrophoretic karyotype of cellulolytic *Penicillium janthinellum* strains. Curr Genet 20:289–291

Kondrashov AS (1994) Muller' ratchet under epistatic selection. Genetics 136:1469–1473

Kwan EY, Lau YL, Yuen KY, Jones BM, Low LC (1997) Penicillium marneffei infection in a non-HIV infected child. J Paediatr Child Health 33:267–271

Lo CY, Chan DT, Yuen KY, Li FK, Cheng KP (1995) Penicillium marneffei infection in a patient with SLE. Lupus 4:229–231

LoBuglio KF, Taylor JW (1995) Phylogeny and PCR identification of the human pathogenic fungus *Penicillium marneffei*. J Clin Microbiol 33:85–89

Pryciak PM, Hartwell LH (1996) *AKR1* encodes a candidate effector of the G βγ complex in the *Saccharomyces cerevisiae* pheromone response pathway and contributes to control of both cell shape and signal transduction. Mol Cell Biol 16: 2614–2626

Stachelhaus T, Marahiel MA (1995) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. J Biol Chem 270:6163–6169

Stoesser G, Baker W, Van den Broek A, Camon E, Garcia-Pastor M, Kanz C, Kulikova T, Lombard V, Lopez R, Parkinson H, Redaschi N, Sterk P, Stoehr P, Tuli MA (2001) The EMBL nucleotide sequence database. Nucleic Acids Res 29:17–21

Supparatpinyo K, Khamwan C, Baosoung V, Nelson KE, Sirisanthana T (1994) Disseminated *Penicillium marneffei* infection in southeast Asia. Lancet 344:110–113

Udagawa S, Furuya K, Horie Y (1973) Mycological reports from New Guinea and the Solomon Islands (compiled by Y. Kobayasi). 19. Notes on some ascomycetous microfungi from soil. Bull. Natl Sci Mus Tokyo 16:503–520

Vanittanakom N, Merz WG, Sittisombut N, Khamwan C, Nelson KE, Sirisanthana T (1998) Specific identification of *Penicillium marneffei* by a polymerase chain reaction/hybridization technique. Med Mycol 36:169–175

Verweij PE, Meis JF, Van den Hurk P, Zoll J, Samson RA, Melchers WJ (1995) Phylogenetic relationships of five species of *Aspergillus* and related taxa as deduced by comparison of sequences of small subunit ribosomal RNA. J Med Vet Mycol 33:185–190

Wong KH, Lee SS (1998) Comparing the first and second hundred AIDS cases in Hong Kong. Singapore Med J 39:236–240

Wong LP, Woo PC, Wu, AY, Yuen KY (2002) DNA immunization using a secreted cell wall antigen Mp1p is protective against Penicillium marneffei infection. Vaccine 20:2878–2886

Wong SS, Siau H, Yuen KY (1999) Penicilliosis marneffei—West meets East. J Med Microbiol 48:973–975

Wong SS, Ho TY, Ngan AH, Woo PC, Que TL, Yuen KY (2001a) Biotyping of *Penicillium marneffei* reveals concentration-dependent growth inhibition by galactose. J Clin Microbiol 39: 1416–1421

Wong SS, Wong KH, Hui WT, Lee SS, Lo JY, Cao L, Yuen KY (2001b) Differences in clinical and laboratory diagnostic characteristics of penicilliosis marneffei in human immunodeficiency virus (HIV)- and non-HIV-infected patients. J Clin Microbiol 39:4535–4540

Wong SS, Woo PC, Yuen KY (2001c) *Candida tropicalis* and *Penicillium marneffei* mixed fungaemia in a patient with Waldenstrom's macroglobulinaemia. Eur J Clin Microbiol Infect Dis 20:132–135

Wu S, Guo N, Yin Z, Chai J. (1996) Characterization of pathogenic fungi genomes using pulsed field gel electrophoresis. Chin Med Sci J 11:188–190

Young C, Itoh Y, Johnson R, Garthwaie I, Miles CO, Munday-Finch SC, Scott B (1998) Paxilline-negative mutants of *Penicillium paxilli* generated by heterologous and homologous plasmid integration. Curr Genet 33:368–377

Yuen KY, Wong SS, Tsang DN, Chau PY (1994) Serodiagnosis of *Penicillium marneffei* infection. Lancet 344:444–445