

Molecular identification of pathogenic fungi

Andrew M. Borman^{1*}, Christopher J. Linton¹, Sarah-Jane Miles² and Elizabeth M. Johnson¹

¹*Mycology Reference Laboratory, Health Protection Agency, Myrtle Road, Kingsdown, Bristol BS2 8EL, UK;*

²*Department of Pathology and Microbiology, University of Bristol, Bristol, UK*

Systemic fungal infections represent a major cause of morbidity and mortality in immunocompromised patients. The ever-increasing number of yeast species associated with human infections that are not covered by conventional identification kits, and the fact that moulds isolated from deep infections are frequently impossible to identify using classical methods due to lack of sporulation, has driven the need for rapid, robust molecular identification techniques. We recently developed a rapid method of preparing fungal genomic DNAs using Whatman FTA filters, which has greatly facilitated molecular identification. Mould isolates cultured from dark grain mycetomas (destructive infections of skin/subcutaneous tissues that progress to involve muscle and bone) invariably fail to produce features by which they can be identified and were taxonomic mysteries. PCR amplification and sequencing of 250 bp of the internal transcribed spacer region 1 (ITS1) allowed us to distinguish between the known agents of mycetoma, to describe three new species associated with this disease and to define phylogenetic relationships. For yeasts, 153 isolates encompassing 47 species that had failed to be identified using classical methods were unambiguously identified by conventional sequencing of 350 bp of the 26S rRNA D1D2 region. These represented 5% of the isolates examined and included common species with atypical biochemical and phenotypic profiles, and rarer species infrequently associated with infection. Our recent studies indicate that FTA extraction coupled with pyrosequencing of 25 bp of ITS2 could potentially identify most common yeast species from pure culture in half a day. Together, these data underscore the importance of molecular techniques for fungal identification.

Keywords: moulds, yeasts, nuclear ribosomal DNA genes, ITS, pyrosequencing

Introduction

Invasive infections by mould and yeast species remain a significant cause of mortality in immunocompromised patients and those undergoing invasive procedures.^{1,2} Rapid identification of yeast isolates from clinical samples is particularly important given their innately variable antifungal susceptibility profiles, but is complicated by the increasing number of emerging pathogenic species that are not included in the repertoires of commercially available conventional identification kits. The identification of mould species associated with deep, invasive infections is equally complicated, due to the astonishing diversity of moulds capable of causing infections in immunocompromised hosts and the fact that isolates cultured from deep sites have often undergone pleomorphism and ceased to produce structures by which they can be identified.³ Thus, the identification, taxonomy and epidemiological analyses of fungal pathogens are increasingly dependent on modern molecular techniques, based on PCR amplification of conserved regions of the genome and sequencing of the resulting PCR products.^{4–6} However, such approaches have long been hindered by the presence of potent PCR inhibitors in

fungal cultures, and difficulties inherent in breaking fungal cell walls.⁷

Although many current fungal DNA extraction procedures eliminate the contaminants that commonly inhibit PCR, the resulting protocols are laborious and costly. These limitations drove us recently to examine the possibility of using Whatman FTA filter matrices as a rapid method for generating and storing PCR-grade fungal genomic DNA.⁸ Whatman's proprietary FTA filter cards have successfully been employed for the rapid preparation and archiving of DNA from a wide variety of sources, including viruses, bacteria, plants, corals, protozoa and mammalian tissue.⁹

Many molecular identification approaches have been evaluated to date for fungi.^{4–6,8} Of these, PCR amplification of genomic DNA followed by sequencing of resulting amplicons has shown the most promise, at least for the more unusual isolates from human infections. Target regions for sequence-based approaches ideally should evolve slowly and show sufficient conservation among fungi to allow cross-species amplification, while being variable enough to allow robust discrimination between closely related species. In studies using limited numbers of different

*Corresponding author: Tel: +44-117-926-8683; Fax: +44-117-922-6611; E-mail: andy.borman@ubht.nhs.uk

isolates, sequencing of portions of the nuclear ribosomal RNA gene cassette, and specifically the internal transcribed spacer (ITS) regions, proved sufficient to discriminate between many species of clinically important yeasts^{10,11} and useful for phylogenetic analyses of a variety of moulds.¹²

Here, we summarize the recent advances in the field of molecular identification of fungal pathogens, highlighting the approaches used for both yeast and mould isolates, and the potential for more rapid identification of yeasts using the innovative and ultra-rapid pyrosequencing of short regions of specific yeast genes.

Rapid extraction of ‘molecular-grade’ fungal genomic DNA using Whatman FTA paper

Prior to our use of Whatman FTA cards, the simplest protocols for fungal DNA preparation were relatively expensive, involved several centrifugation steps,^{13,14} and took over 1 h to complete. Whatman FTA filter matrices are fibrous cards pre-treated with chelators and denaturants that lyse and inactivate most microorganisms on contact. The large nucleic acids released after lysis become physically entangled in the fibres of the FTA matrix, whereas cellular debris is rapidly removed by washing the inoculated card. We demonstrated that application of aqueous suspensions of yeast cells or hyphal fragments and conidia (in the case of moulds) to dry FTA filters, followed by brief microwave treatment, inactivated all 38 yeast and 75 mould species tested. Moreover, 218 of 226 fungal isolates tested (96.5%) liberated amplifiable DNA after application to FTA filters, with a total preparation time of ~15 min (see Borman *et al.*⁸ for the full list of the organisms tested). PCR detection limits evaluated with yeast suspensions were ~10 cfu per punch. In addition, recent experiments with stored inoculated cards indicate that fungal genomic DNA archived on FTA filters is stable at least for several years (data not shown).

Molecular identification and phylogenetic relationships of the agents of dark grain mycetoma

Mycetoma (Madura Foot) is a destructive infection of the skin and subcutaneous tissues, progressing to involve muscle and bone. Infection follows traumatic implantation of fungal spores or hyphal fragments, present in soil or on plant material. Although infections mainly occur in the tropical areas of India, Africa and the Americas, imported cases of fungal mycetoma are not uncommon in the UK. Diagnosis of mycetoma requires culture of the responsible agent from biopsy material. Unfortunately, identification of the causative agent poses particular problems. In effect, the three principal species of fungi known to cause dark grain mycetoma (*Madurella grisea*, *Madurella mycetomatis* and *Pyrenochaeta romeroi*) rarely produce structures in culture by which they can be identified. Moreover, virtually no data exist concerning the genetic relationships between these organisms, and the possibility that other moulds may provoke mycetoma has not been addressed.

To investigate these issues, isolates from cases of mycetoma stored in the National Collection of Pathogenic Fungi (NCPF) housed at the Mycology Reference Laboratory (MRL) were cultured and fungal genomic DNA extracted using Whatman FTA papers. PCR was used to amplify the ITS region 1 and the D1/D2 region of the large ribosomal subunit for each isolate as described previously,¹⁵ and PCR products were sequenced.

FTA filters allowed rapid preparation of PCR-grade fungal genomic DNA from mycetomal agents in <1 h. PCR amplification and sequencing of either the D1/D2 or ITS1 regions discriminated between the various known fungal agents of mycetoma (S.-J. Miles, A. M. Borman, C. J. Linton, C. K. Campbell, P. D. Bridge and E. M. Johnson, unpublished results). Importantly, four distinct genetic groups were identified among agents previously identified as *M. grisea* by classical methods (Figure 1). All share low genetic relatedness with *M. mycetomatis*. Group I fungi are identical to *P. romeroi* and are thus isolates of *P. romeroi* that had never been

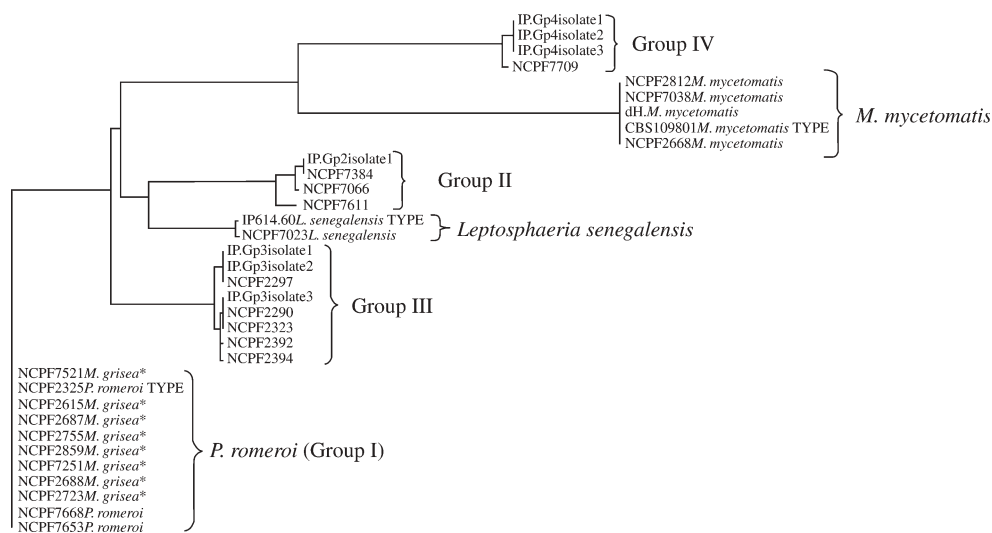


Figure 1. Molecular phylogenetic analysis of the agents of dark grain mycetoma. Sequences corresponding to the ITS1 region were aligned using Clustal W,²⁷ treating gaps as missing data. An unrooted phylogram is shown. Type strains for each species are indicated where available. NCPF, National Collection of Pathogenic Fungi; IP, Institut Pasteur Culture Collection, Paris; CBS, Centraalbureau voor Schimmelcultures, The Netherlands; dH, Sybren de Hoog, CBS. Asterisks indicate isolates of *P. romeroi* misidentified as *M. grisea* on the basis of lack of sporulation.

Molecular identification of pathogenic fungi

induced to sporulate. Groups II–IV are genetically distinct from *P. romeroi* (group I) and from each other and matched no related sequences in the available public databases. On the basis of sequence comparisons with the main fungal orders, group II isolates, which were all from the Indian subcontinent, are best placed in Pleosporales and are possible Leptosphaeraceae. Group III isolates (all from South America) and group IV isolates (from diverse geographical locations) are probably members of the Pleosporales and Sordariales, respectively. Interestingly, the three new species of mycetomal agents identified by this approach were also shown in the same study to have significantly different profiles of antifungal drug susceptibility, underscoring the importance of molecular identification of these agents for optimized patient management.

Large ribosomal subunit sequence-based identification of unusual pathogenic yeasts

To date, more than 700 species of yeast have been described, of which more than 200 belonging to 25 genera have been isolated from human infections. In addition, the increasing role of non-*Candida albicans* species, and the intrinsic or potential resistance of some emerging species to certain antifungal agents, makes rapid species-level identification even more important for informed therapeutic decisions in acutely ill patients. Currently available biochemical systems for yeast identification perform admirably with the 30 or 40 most commonly encountered species, but are inadequate to identify the more unusual potential pathogens.

Over a 2 year period, the MRL received 3033 yeast isolates for identification. Of these, 2880 (95%) were identified using a combination of germ tube testing and the commercial AUXACOLOR2 or API 20C AUX biochemical kits.¹⁶ The remaining 153 isolates were only satisfactorily identified by PCR amplification and sequencing of the D1/D2 region of the large ribosomal subunit gene. These 153 isolates encompassed rare examples of commonly encountered yeasts that had yielded erroneous or incomplete phenotypic or biochemical profiles in conventional identification systems, but also a large number of the less common yeast species (Table 1). In all cases, a combination of BLAST searches of resulting sequences against those in the public databases and phylogenetic analyses with sequences in our own laboratory's database allowed unambiguous identification of the unknown isolate (Figure 2). Moreover, molecular identifications correlated extremely well with the known biochemical and phenotypic properties of the given organisms.

The ultra-rapid identification of yeast isolates using pyrosequencing

Although our past studies have demonstrated that sequence-based molecular identification of yeast isolates is robust, the methodologies involved (including out-of-house automated sequencing) are still time-consuming (turnaround times of 5–7 days) and relatively expensive. Given these drawbacks, we recently evaluated the possibility of identifying yeasts using pyrosequencing (Biotage™) technology, using the commercially available Fungal ID kit marketed by Biotage (Uppsala, Sweden),

Table 1. List of the yeast species identified by molecular methods, in the ascending order of the frequency of referral to the MRL

Organism	Number received	Number requiring molecular ID	% requiring molecular ID
<i>C. albicans</i>	1192	7	0.6
<i>C. glabrata</i>	659	6	0.9
<i>C. parapsilosis</i>	430	3	0.7
<i>C. tropicalis</i>	156	0	0
<i>C. lusitaniae</i>	99	19	19.2
<i>Candida krusei</i>	74	4	5.4
<i>C. neoformans</i>	64	3	4.7
<i>Saccharomyces cerevisiae</i>	60	4	6.7
<i>C. guilliermondii</i>	53	19	35.8
<i>Malassezia pachydermatis</i>	36	2	5.5
<i>C. dubliniensis</i>	33	3	9.1
<i>Trichosporon</i> sp. ^a	23	4	17.4
<i>Rhodotorula</i> sp. ^a	19	0	0
<i>C. inconspicua</i>	17	5	29.4
<i>Candida kefyr</i>	15	4	26.7
<i>Candida famata</i>	12	3	25
<i>C. nivariensis</i>	11	11	100
<i>Geotrichum</i> sp. ^a	10	3	30
<i>C. pelliculosa</i>	8	2	25
<i>C. fabianii</i>	5	5	100
<i>C. lipolytica</i>	5	3	60
<i>Candida blankii</i>	4	4	100
<i>Candida utilis</i>	4	1	25
<i>Cryptococcus albidus</i>	4	0	0
<i>C. rugosa</i>	3	1	33.3
<i>C. norvegensis</i>	3	3	100
<i>Candida pararugosa</i>	3	3	100
<i>Candida catenulata</i>	3	3	100
<i>Kloeckera</i> sp. ^a	3	1	33.3
<i>Saccharomyces elongasporus</i>	2	2	100
<i>C. zeylanoides</i>	2	2	100
<i>Candida eremophila</i>	2	2	100
<i>Candida lambica</i>	2	2	100
<i>Candida ciferii</i>	1	1	100
<i>Candida boidinii</i>	1	1	100
<i>Candida palmiophila</i>	1	1	100
<i>Candida freyschussii</i>	1	1	100
<i>Candida magnoliae</i>	1	1	100
<i>Candida viswanathii</i>	1	1	100
<i>Candida haemulonii</i>	1	1	100
<i>Candida pseudointermedia</i>	1	1	100
<i>Candida pseudoglebosa</i>	1	1	100

^aMolecular identification to species level was not attempted. Adapted from Linton *et al.*¹⁶

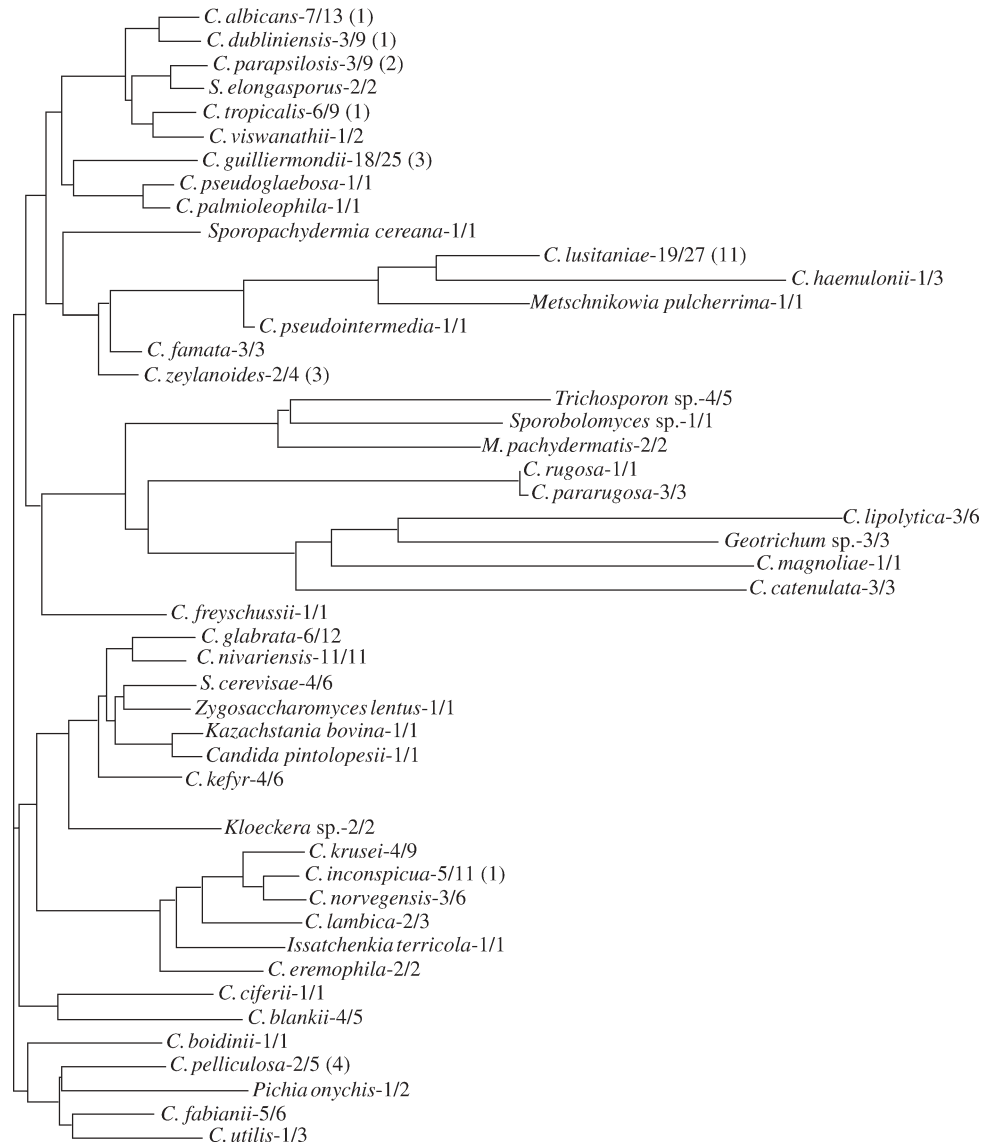


Figure 2. Phylogram of multiply aligned yeast D1/D2 sequences. Alignments were made using ClustalW, treating gaps as missing data. An unrooted phylogram is shown. Numbers after species names indicate the number of sequences generated in the study/total number of clinical isolate sequences analysed by the MRL for the particular organism and (number of additional NCPF reference strain sequences) which show concordant groupings. The NCPF strains used were *C. albicans* NCPF 3179; *Candida dubliniensis* NCPF 3949; *C. parapsilosis* NCPF 8334 and 8766, NCPF 8717; *Candida tropicalis* NCPF 8717; *Candida guilliermondii* NCPF 3334, 3896 and 3099; *Candida lusitaniae* NCPF 3968, 8012, 8031, 8032, 8150, 8235, 8245, 8310, 8352, 8369 and 8377; *C. pelliculosa* NCPF 3115, 3358, 3370 and 8480; *Candida zeylanoides* NCPF 3335, 3871 and 8120; and *C. inconspicua* NCPF 3859.

and based on pyrosequencing of some 25 bp of the ITS2 region. Pyrosequencing correctly identified all test organisms, which included multiple isolates of each of the 11 most common *Candida* species from Table 1 and also isolates of the rarer organisms *Candida norvegensis*, *Candida inconspicua*, *Candida rugosa*, *Candida fabianii*, *Candida lipolytica* and *Candida pelliculosa*. Interestingly, pyrosequencing of a very short region of ITS2 was sufficient to discriminate between *Candida glabrata* and *Candida nivariensis*, a recently described yeast species,¹⁷ which we believe to be an emerging pathogen (R. Petch, C. J. Linton, M. D. Palmer, A. M. Borman and E. M. Johnson, unpublished results). This approach was also able to distinguish *Candida parapsilosis* from *Candida orthopsilosis* and *Candida metapsilosis* (A. M. Borman, C. J. Linton and E. M. Johnson,

unpublished results), which are species that have only recently been distinguished from the *C. parapsilosis* complex on the basis of multilocus sequence typing.¹⁸

Conclusions and perspectives

Our recent studies have demonstrated that Whatman FTA technology represents a cheap, ultra-rapid method of fungal genomic DNA preparation, which has greatly facilitated our own molecular approaches. The experimental results presented here are an attempt to demonstrate the utility of molecular methods for fungal identification. PCR amplification and sequencing of conserved regions of the fungal genome allow unambiguous

Molecular identification of pathogenic fungi

identification of both mould species that are impossible to identify with classical mycological techniques and yeast species that fall outside the range of currently available commercial identification kits and which would require laborious biochemical analyses for correct identification.

We believe that molecular biology will impact more and more heavily on medical mycology. Species identification based on conventional phenotypic methods is often time-consuming and laborious and is hindered by the unstable and subjective nature of phenotypic characteristics, which are readily influenced by culture conditions. Conversely, molecular methods involving gene sequencing are objective, yield results that are uninfluenced by growth conditions, are frequently more rapid than phenotypic approaches and are capable of discriminating between fungi that fail to produce distinctive morphological features. Indeed, recent molecular studies have shown that many 'single species' of fungi pathogenic to humans in fact contain several phylogenetically distinct entities. Such 'cryptic' species have been described for *C. parapsilosis*,¹⁸ *C. albicans*,¹⁹ *Cryptococcus neoformans*,²⁰ *Aspergillus flavus*,²¹ *Histoplasma capsulatum*,²² *Coccidioides immitis*²³ and even *Aspergillus fumigatus*.^{24,25} Moreover, at least in certain cases, the ability to distinguish these cryptic species is of clinical importance as they have been shown to have significantly different antifungal susceptibility profiles or pathogenicity characteristics.^{22,25,26} Many such cryptic species are extremely difficult or even impossible to identify using conventional phenotypic methods.

However, several caveats remain concerning the over-reliance on molecular techniques. First, to date, there is no accepted rule on the number of nucleotide differences in rRNA genes required to define different species or genera. Secondly, the accuracy and especially the true identities of a proportion of sequences submitted to public searchable databases appear dubious. In such cases, spurious identifications can be guarded against by careful examination of the correlation between the results of molecular techniques and available phenotypic and biochemical features. Molecular approaches involving the sequencing of multiple loci for each isolate are likely to help resolve some of these problems.²³

Acknowledgements

We are indebted to the members of the MRL past and present for their contributions to the molecular identification of fungi and to our many national and international collaborators, for sharing unpublished data.

Funding

These studies were supported in part by Whatman International and Biotage.

Transparency declarations

None to declare.

References

1. Ruhnke M. Epidemiology of *Candida albicans* infections and role of non-*Candida albicans* yeasts. *Curr Drug Targets* 2006; **7**: 495–504.
2. Nucci M, Marr KA. Emerging fungal diseases. *Clin Infect Dis* 2005; **41**: 521–6.
3. Leslie CE, Flannigan B, Milne LJR. Morphological studies on clinical isolates of *Aspergillus fumigatus*. *J Med Vet Mycol* 1988; **26**: 335–41.
4. Haynes K, Westerneng T, Fell J *et al*. Rapid detection and identification of pathogenic fungi by polymerase chain reaction amplification of large subunit ribosomal DNA. *J Med Vet Mycol* 1996; **33**: 319–25.
5. Makimura K, Murayama SY, Yamaguchi H. Detection of a wide range of medically important fungi by polymerase chain reaction. *J Med Microbiol* 1994; **40**: 358–64.
6. Sandhu GS, Kline BC, Stockman L *et al*. Molecular probes for diagnosis of fungal infections. *J Clin Microbiol* 1995; **33**: 2913–9.
7. Tang C, Holden D, Aufauvre-Brown A *et al*. The detection of *Aspergillus* spp. by the polymerase chain reaction and its evaluation of bronchoalveolar lavage fluid. *Am Rev Respir Dis* 1993; **148**: 1313–7.
8. Borman AM, Linton CJ, Miles S-J *et al*. Ultra-rapid preparation of total genomic DNA from isolates of yeast and mould using Whatman FTA filter paper technology—a re-usable DNA archiving system. *Med Mycol* 2006; **44**: 389–98.
9. Smith LM, Burgoyne LA. Collecting, archiving and processing DNA from wildlife samples using FTA databasing paper. *BMC Ecol* 2004; **4**: 4.
10. Chen Y-C, Eisner JD, Kattar MM *et al*. Identification of medically important yeasts using PCR-based detection of DNA sequence polymorphisms in the internal transcribed spacer region 2 of the rRNA genes. *J Clin Microbiol* 2000; **38**: 2302–10.
11. Chen Y-C, Eisner JD, Kattar MM *et al*. Polymorphic internal transcribed spacer region 1 DNA sequences identify medically important yeasts. *J Clin Microbiol* 2001; **39**: 4042–51.
12. Iwen PC, Hinrichs SH, Rupp ME. Utilisation of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. *Med Mycol* 2002; **40**: 87–109.
13. Cenis JL. Rapid extraction of fungal DNA for PCR amplification. *Nucleic Acids Res* 1992; **20**: 2380.
14. Liu D, Coloe S, Baird R *et al*. Application of PCR to the identification of dermatophyte fungi. *J Med Microbiol* 2000; **49**: 493–7.
15. Borman AM, Campbell CK, Linton CJ *et al*. *Polycytella hominis* is a mutated form of *Scedosporium apiospermum*. *Med Mycol* 2006; **44**: 33–9.
16. Linton CJ, Borman AM, Cheung G *et al*. Molecular identification of unusual pathogenic yeast isolates by large ribosomal subunit gene sequencing: 2 years experience at the UK Mycology Reference Laboratory. *J Clin Microbiol* 2007; **5**: 1152–8.
17. Alcoba-Florez J, Mendez-Alvarez S, Cano J *et al*. Phenotypic and molecular characterization of *Candida nivariensis* sp. nov., a possible new opportunistic fungus. *J Clin Microbiol* 2005; **43**: 4107–11.
18. Tavanti A, Davidson AD, Gow NAR *et al*. *Candida orthopsilosis* and *Candida metapsilosis* spp. nov. to replace *Candida parapsilosis* groups II and III. *J Clin Microbiol* 2005; **43**: 284–92.
19. Sullivan DJ, Westerneng TJ, Haynes KA *et al*. *Candida dubliniensis* sp. nov.: phenotypic and molecular characterisation of a novel species associated with oral candidosis in HIV-infected individuals. *Microbiology* 1995; **141**: 1507–21.
20. Franzot SP, Salkin IF, Casadevall A. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. *J Clin Microbiol* 1999; **37**: 838–40.

21. Geiser DM, Pitt JI, Taylor JW. Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. *Proc Natl Acad Sci USA* 1998; **95**: 388–93.
22. Kasuga T, White TJ, Koenig G *et al.* Phylogeography of the fungal pathogen *Histoplasma capsulatum*. *Mol Ecol* 2003; **12**: 3383–401.
23. Taylor J, Fisher M. Fungal multilocus sequence typing—it's not just for bacteria. *Curr Opin Microbiol* 2003; **6**: 351–6.
24. Balajee SA, Gribskov JL, Hanley E *et al.* *Aspergillus lentulus* sp. nov., a new sibling species of *A. fumigatus*. *Eukaryot Cell* 2005; **4**: 625–32.
25. Balajee SA, Nickle D, Varga J *et al.* Molecular studies reveal frequent misidentification of *Aspergillus fumigatus* by morphotyping. *Eukaryot Cell* 2006; **5**: 1705–12.
26. Balajee SA, Gribskov J, Brandt M *et al.* Mistaken identity: *Neosartorya pseudofischeri* and its anamorph masquerading as *Aspergillus fumigatus*. *J Clin Microbiol* 2005; **43**: 5996–9.
27. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673–80.