Multilocus Microsatellite Typing System for *Penicillium marneffei*
Reveals Spatially Structured Populations

Matthew C. Fisher, 1* David Aanensen, 1 Sybren de Hoog, 2 and Nongnuch Vanittanakom 3

Department of Infectious Disease Epidemiology, Imperial College Faculty of Medicine, London, United Kingdom 1; Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands 2; and Department of Microbiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand 3

Received 26 April 2004/Returned for modification 18 May 2004/Accepted 23 June 2004

For eukaryotic pathogens that have low levels of genetic variation, multilocus microsatellite typing (MLMT) offers an accurate and reproducible method of characterizing genetic diversity. Here, we describe the application of an MLMT system to the emerging pathogenic fungus *Penicillium marneffei*. Isolates used for this study were those held in the culture collections of the Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands, and the Chiang Mai University Department of Microbiology, Chiang Mai, Thailand. High genetic diversity and extensive spatial structure were observed among clinical isolates, with the geographical area of origin for each isolate strongly correlating with the occurrence of two deeply divided clades. Within each clade, multilocus linkage associations were highly significant and could be explained by genetically differentiated populations or by an exclusively clonal reproductive mode, or both. Our results show that southeast Asian penicilliosis is caused by a fungus with a complex population genetic structure. Furthermore, this MLMT system generates digital data that can be easily queried against a centrally held database via the internet (http://pmarneffei.multilocus.net/); this provides a powerful epidemiological tool for analyzing the underlying parameters that are responsible for the emergence of *P. marneffei* in human immunodeficiency virus-positive populations.

*Penicillium marneffei* is a mitosporic pathogenic fungus of the family *Trichocomaceae* that has emerged since 1990 as a significant human mycosis. Emergence of *P. marneffei* has occurred in concert with the explosive epidemic of AIDS in southeast Asia, where it is classified as an AIDS-related indicator disease (17). In northern Thailand, *P. marneffei* is the third most frequent opportunistic infectious disease organism, after tuberculosis and cryptococcosis (22).

So far, the ecology and epidemiology of *P. marneffei* have remained enigmatic. The range of endemicity of the pathogen is confined to southeast Asia, where autochthonous isolations from northeast India, Thailand, the Guangxi region of China, Hong Kong, Taiwan, Vietnam, and Indonesia are known (1). In these regions, *P. marneffei* is a naturally occurring sylvatic infection in a high proportion of bamboo rat species (3–5). However, it is not known whether bamboo rats are (i) an obligate stage in *P. marneffei*'s life cycle and (ii) a zoonotic focus for human infection. Furthermore, it is not known whether all lineages of *P. marneffei* are equally infectious to bamboo rats and humans or rather represent a subset of a wider, more genetically diverse population. In order to address these questions, there is an urgent need for well-characterized neutral markers with which to analyze the population genetic structure of *P. marneffei*.

Studies of *P. marneffei* using molecular epidemiological techniques are still at an early stage of development, with previous differentiation of *P. marneffei* isolates relying on PCR amplification of anonymous loci (14) or enzyme-based restriction of nuclear DNA (14, 25). While multilocus sequence typing (MLST) (19) is becoming the established method for discriminating individuals in many infectious diseases, insufficient genetic variation in eukaryotic species can pose a problem (23). In such cases, loci with increased variability are required. Studies of the ascomycete fungi *Coccidioides immitis* and *Coccidioides posadasi* have shown that microsatellite-based typing systems work well to discriminate individuals, populations, and species (7, 10), and microsatellites have been shown to be polymorphic in *P. marneffei* (16). Here, we describe the application of a multilocus microsatellite typing (MLMT) system to human *P. marneffei* and illustrate its use in addressing broad-scale patterns of genetic variation.

**MATERIALS AND METHODS**

Isolates. A panel of 24 isolates comprising 17 clinical isolates from the collection held at the Centraalbureau voor Schimmelcultures (CBS) and the type specimen isolated from a bamboo rat from Vietnam in 1959 (3) was assembled for analysis. In addition, clinical isolates were chosen at random from the collection of N.V. at the Maharaj Hospital, University of Chiang Mai, Chiang Mai, Thailand (Table 1). The reproducibility of the typing system was assessed by sampling multiple isolates from the same patients; assuming that mixed infections are rare cases for disseminated penicilliosis, isolates from the same patient should have identical multilocus genotypes. *P. marneffei* isolates were grown on Sabouraud’s agar at room temperature. Subsequently, each isolate was assigned an extraction number and DNA was extracted from 7-day-old cultures as described previously (25). All isolates were typed “blind” to control for user bias.

**MLMT loci and data analysis.** Twenty-three *P. marneffei*-specific microsatellite loci were developed, and isolates were genotyped as described previously (8). Briefly, primers for each of the chosen microsatellite loci were aliquoted into four multiplex pools to a concentration of 2 µM, and 1 µl of a 1/10 dilution of the DNA from each isolate was amplified by using a multiplex PCR kit (QIA-GEN) with a working primer concentration of 0.2 µM. Cycling conditions were...
is the number of loci in the MLMT system. For $D_j$, pairwise distance matrices were calculated between isolates by using MicroSatellite Analyzer software (6), and confidence intervals were calculated by bootstrapping across loci. The occurrence of multilocus linkage associations within the data set were assessed by calculating the index of association, $I_A$ (2, 21), and statistical significance was assessed by comparisons against randomized data sets. The $I_A$ was calculated for data sets containing all genotypes (but excluding the second duplicated isolate from the Chiang Mai patients) and from clone-corrected data sets from which naturally occurring identical genotypes had been removed.

**RESULTS**

MLMT. Of the 23 typeable loci, 21 were polymorphic, with between 2 and 14 alleles within our sample, corresponding to a discriminatory probability that ranged from 0.472 (locus PM23) to 0.929 (locus PM2). The most polymorphic locus, PM2, had 14 alleles that varied by ±76 nucleotides. The large number of alleles spanning a wide size range at this locus suggests an anomalous and hypervariable mode of mutation due to flanking sequence indels. For this reason, all subsequent analyses were performed without locus PM2, resulting in a final data set of 20 loci and 24 isolates. When the blinded had been removed from our panel of DNA, it was found that isolates taken from the same patient on different occasions (isolates NV16 and NV19, isolates NV12 and NV30, and isolates NV56 and NV59) had identical multilocus genotypes, demonstrating high reproducibility of the MLMT system at all 23 loci. For all loci, typeability was high, with 550 out of 552 alleles being successfully characterized (99.6%).

Once the duplicated Thai genotypes were removed from further analyses, the data set contained 19 unique MTs out of the 21 independent isolates (Table 1). This corresponds to a genotypic diversity, $X$, in this data set of 0.995. Two pairs of isolates, isolates CBS122.89 (Indonesia) and CBS451.91 (Thailand) (MT9) and isolates CBS107.89 (Thailand) and CBS120.89 (Thailand) (MT16) were of identical genotype. Calculating the probability of sampling these particular genotypes more than once in this data set using binomial probabilities (11) demonstrated that both genotypes were statistically unlikely to be observed by chance alone (for MT9, $P < 0.001$; for MT16, $P < 0.001$) and could therefore be considered genetically identical clones.

**Genetic analyses.** The clustering of isolates based on pairwise distances using $D_j$ showed that isolates occur within one of two clades separated by a large genetic distance (Fig. 1). The upper clade contained 10 *P. marneffei* strains, six of which were isolated from patients living in areas of mainland China (3), Hong Kong (2), and Indonesia (1), where *P. marneffei* is endemic. This clade also contained the type isolate from Vietnam, CBS388.87. Due to the dominance of isolates in this clade that were recovered from patients living in the eastern part of *P. marneffei*’s range, this clade was subsequently referred to as “Eastern.” Conversely, the lower clade contained isolates that were exclusively recovered from the western sector of *P. marneffei*’s range, Thailand (11) and India (1), and this clade was subsequently referred to as “Western.” The spatial distribution of genetic variation was assessed by using the statistic $\theta$, estimated with the analysis of variance framework of Weir (26). Here, $\theta = 0.262$ ($P < 0.001$), a value which corresponds to 26% of the total genetic diversity attributable to the spatial structure between the Eastern and Western isolates.
Of the 20 loci, 19 contained alleles that were private to either the Eastern or Western clade, comprising 62% of the observed alleles. One locus (PM1) had a distribution of alleles that did not overlap between the two clades and is a candidate diagnostic marker for the phylogenetic structure seen here (Fig. 2). We observed significant differences in the allelic diversity within each clade; the Eastern clade was more polymorphic, with an average of 3.75 alleles per locus, compared with 2.95 alleles per locus within the Western clade (analysis of variance; $F_{1,38} = 5.20, P = 0.028$).

Multilocus linkage associations were highly significant when the data set was considered as a whole ($I_A$ of all isolates, 3.414; $P < 0.001$) or clades were analyzed independently ($I_A$ of Eastern isolates, 2.502 [$P < 0.001$]; $I_A$ of Western isolates, 4.118 [$P < 0.001$]) (Fig. 1). Clone correction of the data is performed by removing isolates with identical genotypes (CBS122.89 and CBS107.89); this has no effect on the statistical significance of the $I_A$ values.

**Population assignments for isolates using BATS.** The occurrence of strongly spatially structured genotypes in this data set suggests that isolates of unknown ancestry can be assigned to a

**FIG. 1.** Unweighted pair group method with arithmetic averages tree produced by using pairwise microsatellite distance $D_2$. Bootstrap distances of >50% are shown as numbers above branches, and the tree is midpoint rooted. The $I_A$ values for the Eastern and Western clades are shown compared against histograms of $I_A$ values for 1,000 permuted data sets.

**FIG. 2.** Allele distribution showing the sizes of PCR products for the microsatellite locus PM1. White bars show alleles present in the Eastern clade, and black bars show alleles present in the Western clade.
source population by using Bayesian assignment tests (BATs) (12). These statistical tests assign individuals to populations by calculating posterior probability densities for multilocus genotypes within all populations based on the observed allele frequencies, with assignment being made to the population with the highest posterior probability. All isolates were sequentially removed from the data set, and the program BayesAss (12) was then used to assign the isolate to a source population. This process was repeated for all autochthonous isolates and the five isolates of unknown origin (Table 1). For each of the 21 isolates, all were assigned to the same clade within which they had been originally clustered by $D_1$. Posterior probabilities were unambiguous ($P = 0.99$) for all isolates except the Vietnamese type isolate CBS388.87, which was assigned to the Eastern clade with a nonsignificant probability of $P = 0.86$.

**DISCUSSION**

The emergence of *P. marneffei* as a significant pathogen in southeast Asia can be explained by (i) the increase in numbers of immunocompromised individuals within this region and (ii) the thermally dimorphic nature of *P. marneffei* that allows it to infect such individuals. However, until now there has been no experimental framework with which to analyze the underlying epidemiological parameters that are responsible for driving the emergence of *P. marneffei* in association with human immunodeficiency virus-AIDS. Informed public health decisions require information on the contribution of sylvatic cycles of infection to human disease, the dispersal distances from infectious foci over which human infections can occur, and whether all isolates or populations are equally infectious. Here, we have developed an MLMT framework within which such questions can be systematically addressed.

We developed our MLMT system by systematically searching the GenBank database for microsatellites that were contained within the *P. marneffei* genomic sequence (8); the majority of the online genomic sequence comprises genome sequence tags from the efforts of the Hong Kong University Pasteur Research Centre (27). Our work resulted in the successful amplification of 23 microsatellite loci which were subsequently tested on a panel of 23 clinical isolates and the type isolate, CBS388.87. Of the amplifiable loci, 21 were polymorphic within the panel of isolates, and 20 of these loci were used to construct an MLMT database of 19 MTs. The MLMT system was validated with duplicated clinical isolates isolated on separate hospital visits. In each case, both MTs were identical, demonstrating that the reproducibility of the MLMT system was 100%. Moreover, this identity of MTs over time shows that the loci are not mutating at a rate that is high enough to generate intrasolate polymorphisms between separate hospital visits. This result corroborates previous data from MLMT of *C. immitis* and *C. posadasii*, where MTs were shown to be stable over 12 years (7); these data suggest that microsatellite mutation rates for these fungi are in line with those observed for *Saccharomyces cerevisiae* ($2.6 \times 10^{-5}$ per generation) (15).

We used a microsatellite-specific genetic distance, $D_1$, to determine interindividual genetic distances. Due to their stepwise mode of mutation (13), microsatellite loci accumulate historical information not only by generating novel alleles via mutation but also by increasing the population variances in allele sizes. This extra component of variation is utilized by $D_1$ and, under conditions of unconstrained mutation, tends to reconstruct deeper divergences better than simple allele-sharing genetic distances (9). Here, the use of $D_1$ has identified two deep-rooted Eastern and Western clades. These clades explain 26% of the total genetic diversity and contain many private alleles, showing deep genetic divergence and possible cryptic species.

Other MLST and MLMT studies of pathogenic fungi have resulted in the definition of cryptic species based on phylogenetic species concepts (24). However, one must be careful in interpreting our data in such a manner; our analyses are based on a small sample of *P. marneffei* isolates taken from a wide geographical range and therefore may represent single samples from as many populations as there are isolates in this data set. In this case, the observed clustering into two clades may reflect limited longitudinal dispersal of *P. marneffei* rather than speciation. If a stepping-stone model does explain dispersal in *P. marneffei*, further sampling will be key to uncovering the populations that are connective between the Eastern and Western clades described here. Values of $I_1$ are significant within each clade, adding support for the argument that the clades are composed of heterogeneous populations. However, the existence of statistically identical MT types (MT9 and MT16) is preliminary evidence that there are widely disseminated clones of *P. marneffei*.

Phylogenetic analyses by LoBuglio and Taylor have shown that *P. marneffei* is closely related to species of *Penicillium* subgenus *Biverticillium* and sexual *Talaromyces* species with asexual biverticilliate states (18). The fact that the *Talaromyces* sexual stage (teleomorph) seen in other biverticilliate *Penicillium* species has never been observed for *P. marneffei* suggests that there will be a strong asexual component to the population structure of *P. marneffei*. Indeed, it may be argued that the two clades observed here may represent two spatially separated asexual lineages. If this is the case, in order to disentangle the effects of asexual reproduction versus spatial population structure, systematic sampling of *P. marneffei* needs to be undertaken on a variety of spatial scales. Hierarchical analyses of the distribution of genetic diversity will then demonstrate (i) introgressed genotypes that are indicative of recombination or (ii) the maintenance of clonal population structures at all scales. Such sampling is being undertaken by utilizing and developing ongoing mycosis surveillance programs centered in Thailand. Further work will focus on identifying and surveying the mating-type loci in these populations in order to measure the potential for recombination, if it exists.

The existence of spatial structure, based on the genetic distance $D_1$, suggested that isolates could be assigned to one of the two clades by use of assignment tests. As BATs are based on underlying population allele frequencies unlike $D_1$, which also depends on allelic size variation, BATs are an independent test of the phylogenetic structure described by $D_1$. Five *P. marneffei* isolates from our panel were isolated from patients outside the region of endemicity and were probably gained from visits to southeast Asia. All five isolates were assigned to one of the two clades with high probability, suggesting that this was the geographical area where the infection was acquired. However, one isolate, CBS451.91, was recovered from a Dutch patient with human immunodeficiency virus who was reported...
to have returned from Thailand, and this isolate was assigned to the Eastern clade with high probability. If the clinical history of this patient is correct, then it suggests either that Eastern and Western isolates co-occur in Thailand or that this isolate represents a long-distance dispersal event. An MT identical to CBS451.91 was recovered from a clinical case in Indonesia (CBS122.89 [MT]), and this finding is further evidence for widely dispersed clones. However, analyses of larger independent data sets are necessary to corroborate these findings.

Like MLST, MLMT is an exact method by which different laboratories can compare their results to those of other laboratories. The increasing availability of automated genotypers means that MLMT genotypes can be rapidly generated and laboratories can compare their results to those of other laboratories co-occurring in Thailand or that this isolate of this patient is correct, then it suggests either that Eastern clones versus environmentally driven population increase: analysis of an epidemic of the human fungal pathogen Coccidioides immitis. J. Clin. Microbiol. 38:807–813.