ORIGINAL ARTICLE

25S rDNA genotyping and multi-locus sequence typing of *Candida albicans* of pathogenic and commensal origins in the Klang Valley, Malaysia

Zain ILLYAASEEN, Yun Fong NGEOW*, Sook Fan YAP, Hien Fuh NG

Centre for Research on Communicable Diseases, Faculty of Medicine and Health Sciences, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, 43000 Kajang, Selangor, Malaysia

Abstract

Candida albicans is an important opportunistic fungal pathogen capable of causing fatal systemic infections in humans. Presently in Malaysia, there is little information available on the genetic diversity of this organism and trends in behavioural characteristics. In this project, three genotyping methods: 25S rDNA genotyping, Alternative Lengthening of Telomerase (ALT) sequence typing and Multi-Locus Sequence Typing (MLST) were applied to study the genetic diversity of strains from infected hospital in-patients and asymptomatic individuals in the community. The results showed that, with the 25S rDNA genotyping, as in other parts of the world, the most common genotype was type A which accounted for approximately 70% of the 111 isolates tested. Further typing with the ALT sequence showed type 3 to be the most common in the isolates tested. MLST analysis revealed many possibly novel sequence types, as well as a statistically significant association between pathogenicity and a group of closely related isolates, most of which were from hospital samples. Further work on genotypes associated with enhanced virulence will help to clarify the value of genotyping for clinical and epidemiological investigations.

Keywords: Candida albicans, genotyping, multi-locus sequence typing

INTRODUCTION

Candida albicans is a formidable opportunistic fungal pathogen capable of causing superficial and life-threatening systemic infections in humans. It has been identified as the most common causative species of invasive candidiasis among solid organ transplant recipients and other critically-ill and immunocompromised patients particularly in long-term health facilities.[1,2] It is the major cause of candidaemia[3] and vulvovaginal candidiasis which affects approximately 70-75% of women at least once during their lifetime.[4]

Several genotyping methods have been described for *C. albicans*. A simple technique is the use of PCR amplification of the 25S rDNA to differentiate strains into A, B, C, and D types (Fig. 1). As this method lacks discriminatory power, the genotypes obtained are sometimes further differentiated by a second test based on the PCR amplification of ALT repeats in the repetitive sequences in the *C. albicans* genome

(Fig. 2). The ALT genotypes are designated numbers depending on the presence or absence of bands and their respective intensities as viewed under UV transillumination. Hence, a combination of both 25S rDNA and ALT typing would generate genotypes labelled by a letter A to E followed by a number. With the combined 25S rDNA and ALT typing, a number of studies found the genotype A3 to be the most common in Burkina Faso and Japan where samples were of clinical origin isolated from various sites of infection.[5,6]

A more widely used genotyping method is Multi-locus sequence typing (MLST) that opts for strain differentiation by the sequencing of PCR products of selected house-keeping genes as opposed to agarose gel electrophoretic demonstration of amplicons. The high discriminatory power of MLST as well as the ease of data storage and sharing has made it an attractive option for the strain differentiation

Address for correspondence: Yun Fong Ngeow, Centre for Research on Communicable Diseases, Faculty of Medicine and Health Sciences, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, 43000 Kajang, Selangor, Malaysia. Tel: 03-9086 0288 (ext: 158). Fax: Nil. E-mail: ngeowyf@utar.edu.my
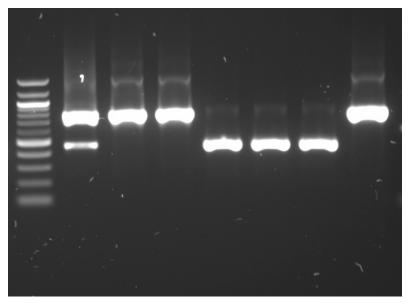
FIG. 1: Agarose gel electrophoresis image of ABC genotyping PCR products. Lane 1: 100bp ladder, lane 2: genotype C, lanes 3, 4 and 8: genotype B, lanes 5, 6 and 7: genotype A.

of many organisms including *C. albicans*.[7] To the best of our knowledge, there is little data on the distribution *C. albicans* genotypes in Malaysia. Tay *et al*.[8] studied 221 Malaysian isolates of clinical origin with the 25S rDNA method and found genotype A to be the most common. This is in contrast to a similar study in Thailand which found genotype B to be the most frequent in patients with oral candidiasis and healthy volunteers.[9] In the MLST online international database (https://pubmlst.org/calbicans/), there are to date, only 7 isolates submitted from Malaysia, dating as far back as 1995.

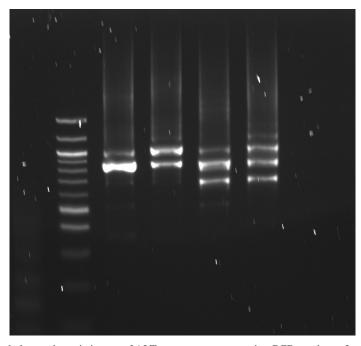In this study, we hope to generate more information on local *C. albicans* genotypes for

FIG. 2: Agarose gel electrophoresis image of ALT repeat sequence typing PCR products. Lane 1: 100bp ladder, lane 2: type 3, lane 3: type 2/3, lane 4: type 3/4, lane 5: type 2/3/4.

comparison of genetic diversity with strains from other parts of the world.

MATERIALS AND METHODS

C. albicans strains

Two groups of *C. albicans* were retrieved from -80°C storage to represent pathogenic and commensal strains. The pathogenic group consisted of 61 isolates from inpatients of a local hospital who were diagnosed to have *C. albicans* infections by positive cultures in a variety of specimen types including blood, urine, sputum, high vaginal and wound swabs. The commensal group consisted of 50 isolates obtained from the buccal swabs of asymptomatic individuals in the community who had no overt signs of oral infection such as white lesions on the inner cheeks or tongue. These swabs were taken during a community survey approved by the Universiti Tunku Abdul Rahman Ethics Committee (Ref.U/SERC/31/2015).

All isolates were revived on CHROMagar Candida (Becton Dickinson) incubated at 37°C for 48 hours. Single green colonies were selected and used for a germ tube test followed by Gram staining and microscopic observation. Germ tube positive isolates were identified by ITS PCR using protocols previously described.[10] PCR products were sequenced and the results were checked in the ISHAM Barcoding Database to confirm species identity.

25S rDNA and ALT repeat sequence typing of *C. albicans*

The DNA of isolates confirmed to be *C. albicans* were used for 25S rDNA-based PCR using the primers CA-INT-L: ATA AGG GAA GTC GGC AAA ATA GAT CCG TAA and CA-INT-R: CCT TGG CTG TGG TTT CGC [5]. PCR products were separated in agarose gel stained with Sybr Safe (Thermo Fisher Scientific) and the results were visualised with a UV transilluminator. The genotype of an isolate was determined based on the size of the DNA band, or bands obtained. Genotype A was observed as a 450 bp band, genotype B a 840 bp band and genotype C 450 bp and 840 bp bands. Subsequently, the PCR for ALT sequence typing was performed in a similar fashion but using the primers ASDcF: TGA ACC TGA ACT TGT GCT ACA AAG and pCSCR: CGC CTC TAT TGG AGC TCG AGT AGT C.[5] PCR products were electrophoresed to determine the ALT type by the presence of

a band or bands and their respective intensities as they appeared under UV transillumination. The results from the 25S rDNA and ALT typing were combined to determine the final genotype of each isolate.

MLST of *C. albicans*

MLST was performed using primers described in other works, consistent with set standards targeting seven housekeeping genes of *C. albicans*.[11,12] PCR products were sequenced and the results were analysed for SNPs. Where applicable, heterogeneous SNPs determined from the chromatogram were marked according to the respective IUPAC nucleotide codes. Analysed sequences were cross-referenced with existing MLST database entries for *C. albicans* to identify matches and determine sequence type designation. In addition, the occurrence of individual SNPs was analysed by extracting the specific loci for all sequences and applying the method described in published literature.[13] Briefly, where heterozygous data was present, the locus was assigned both nucleotide bases constituting the heterozygous IUPAC code. For example, K was replaced with G + T, R with A + G, etc. Homozygous loci were duplicated to ensure proper alignment of sequences for analysis, with A written as AA, T as TT and so on. Such edits enabled the construction of an Unweighted pair group method with an arithmetic mean (UPGMA) phylogenetic tree using MEGA software. This step was necessary given the large number of novel allelic profiles discovered in this study.

Statistical analysis

Results were analysed using Graphpad PRISM and Microsoft Excel. The statistical analysis performed was Fisher's Exact Test. MEGA7[14] was used for UPGMA phylogenetic analysis with bootstrapping (1000 iterations). The Hunter Gaston Discriminatory Index (HGDI)[15] and 95% confidence interval of the respective genotyping methods were determined by using an online software, Comparing Partitions, available at http://www.comparingpartitions. info/?link=Tool.[16] Adjusted Rand coefficient was also calculated using the same software to investigate the congruence of the results between the genotyping methods (with 1 indicating both methods generate completely concordant genotyping results and with 0 indicating both methods generate completely discordant findings).

RESULTS AND DISCUSSION

25S rDNA genotyping and ALT repeat sequence typing

The pathogenic isolates consisted of 75% genotype A (46/61), 20% genotype B (12/61) and 5% genotype C (3/61). The 50 isolates forming the commensal group consisted of 64% genotype A (32/50), 22% genotype B (11/50) and 14% genotype C (7/50). Genotype C was the least abundant genotype observed in this study, as shown in Fig. 3, which is consistent with findings in other regions across the globe.[17–19] The difference in genotype distribution between the pathogenic and commensal groups was not statistically significant. This could be due to the pathogenic strains being isolates from various body sites whereas the commensal strains were all from the buccal cavity. Tantivitayakul *et al*.[9] studied buccal *C. albicans* and reported a statistically significant association between genotype B and the manifestation of oral candidiasis. The difference between their findings and ours seems to suggest that different genotypes have a propensity to express pathogenicity in different ways and in different sites in the host body. Investigations to elucidate these associations could be a focal point for future research.

The grouping of *C. albicans* ABC genotypes and ALT repeat sequence subtypes in this study, displayed in Table 1, is consistent with the findings of other publications, where genotype A3 was found to be the most abundant.[5,6] Details of the specific composition of each genotype are listed in Supplementary Table 1. No statistical significance was found in the distribution of genotypes and subtypes among commensal or pathogenic groups.

MLST

All 111 isolates were subjected to MLST using the globally established standard of 7 housekeeping genes. Of these isolates, 45 had a match to an existing database entry for allelic profiles while the remaining 66 were found to have either a novel allelic profile or a completely new allele number for at least one of seven alleles studied, indicating a novel sequence type. Applying the method described by Tavanti *et al*. (2005)[13], SNP loci were identified and sequence data extracted, edited, aligned and used for phylogenetic analysis. The SNPs of all isolates obtained in this study are shown in Supplementary Table 2. This alternative method was used because typical phylogenetic analysis would make use of allelic profiles and sequence types as assigned by a database query. In this study, the large number of novel sequence types found necessitated an alternative as new allele numbers and sequence type designations could not be obtained due to budgetary limitations. A UPGMA phylogenetic tree was constructed to illustrate the genetic relatedness between the different isolates with a p-distance cut-off of 0.04 to define clades[7], as shown in Fig. 4. The
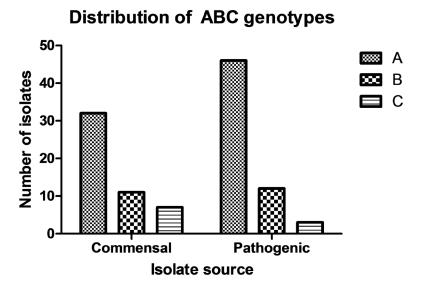

FIG. 3: Distribution of ABC genotypes according to source.

TABLE 1: Numbers of commensal and pathogenic isolates and their MLST clades in each ABC/ALT genotype

ABC/Subgroups	Commensal (clades)	Pathogenic (Clades)
A1/3	0 (0)	2 (3)
A2	0 (0)	2 (1)
A3	20 (1, 5, 12, 14, 17, 19, 21, 30)	16 (1, 8, 17, 19, 25, 29)
A4	0 (0)	1 (19)
A2/3	1 (1)	3 (1, 15)
A3/4	6 (1, 3, 8, 19)	15 (1, 3, 8, 19, 20, 25, 26, 27, 31)
A4/5	0 (0)	1 (1)
A2/3/4	5 (1, 4, 28)	5 (1)
A3/4/5	0 (0)	1 (1)
B3	4 (2, 6, 13, 19)	2 (6, 32)
B4	0 (0)	3 (7, 18, 23)
B2/3	0 (0)	1 (24)
B3/4	7 (6, 9, 10, 14)	5 (7, 9)
B4/5	0 (0)	1 (6)
C3	1 (32)	1 (16)
C4	1 (8)	0 (0)
C3/4	5 (11, 27, 32)	1 (22)
C3/4/5	0 (0)	1 (1)

largest clades in this study numbered arbitrarily, are 1 and 19, with clade 1 containing the largest number of pathogenic isolates (21). A Fisher's Exact test found a statistically significant association between clade 1 and expression of pathogenicity in the isolates (p-value = 0.003) and an odds ratio of 4.725, indicated that *C. albicans* isolates in clade 1 were 4.725 times more likely to be observed in a clinical setting, suggesting that they could have a greater tendency to exhibit pathogenicity. Various assays for the determination of pathogenicity, including gene expression analysis and proteinase activity may further support this data.

The comparison of genotyping methods

As expected, MLST generated a significantly higher score (p-value <0.001) in the HGDI, at 0.985 (95% confidence interval: 0.976-0.995), compared to ABC and ALT sequence typing which scored only 0.837 (95% confidence interval: 0.792-0.883). This indicates that MLST is a very powerful genotyping method capable of capturing the great genetic diversity of *C. albicans* in comparison to ABC and ALT sequence typing.

Ideally, concordant genotyping results are expected to have strains typed as one genotype by a less discriminatory method (ABC+ALT) to be further subdivided into more genotypes by another more discriminatory method (MLST). However, the findings from both genotyping methods appear to be discordant (adjusted Rand coefficient = 0.221). For example, the largest ABC+ALT type, A3 (36 members), can be further divided into 11 MLST clades (clade 1, 5, 8, 12, 14, 17, 19, 21, 25, 29, and 30). In contrast, the largest clade (clade 1 with 26 members) identified by the supposedly more discriminatory MLST method, can be further divided into 8 ABC+ALT types (A2, A2/3, A2/3/4, A3, A3/4, A3/4/5, A4/5, and C3/4/5) (see Supplementary Table 1). This is not entirely unexpected as both methods utilized different genes and interpretations to type the *C. albicans* strains. Differentiation by clades in the UPGMA phylogenetic tree, utilising a p-distance cut-off of 0.04 yielded an HGDI value of 0.895 (95% confidence interval: 0.858-0.932). As such, even using the alternative method of SNP analysis instead of clustering by allelic profile, the MLST is still a better discriminatory tool than ABC and ALT sequence typing combined
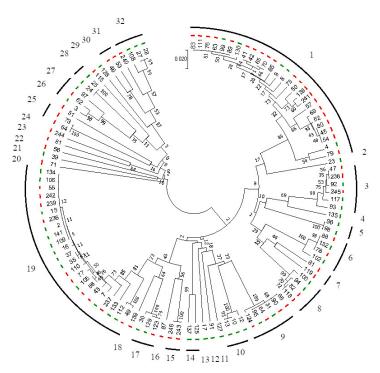
FIG. 4: UPGMA dendrogram constructed using concatenated sequences of seven MLST genes. The bootstrap consensus tree was inferred from 1000 replicates. Red bars indicate pathogenic origin, green bars indicate commensal origin and black bars indicate close genetic relatedness with a p-distance of less than 0.04. Clade numbers appear in the outermost perimeter denoting clades or singletons with a distant genetic relatedness as denoted by a p-distance of greater than 0.04.

(p-value = 0.027). In addition, the unambiguity of using sequencing data for analysis as compared to the visualisation of band intensity via agarose gel electrophoresis is a significant strengthening factor for the use of MLST.

CONCLUSION

In general, the distribution of *C. albicans* genotypes as determined by 25S rDNA analysis and ALT repeat sequences matched the distribution reported elsewhere. MLST data gave great discriminatory power, granting insight into the genetic diversity of isolates in this study. Given that more than half of the isolates studied did not match any of the existing MLST database entries, much more work needs to be done on local isolates. This finding, in addition to the low number of South East Asian submissions to the database (7 from Malaysia, 2 from Vietnam and 16 from Thailand) further strengthens the possibility of a widely undiscovered genetic diversity of *C. albicans* in the region. International cooperation is imperative in this area to confirm the novelty of newly discovered sequence types and

investigate possible reasons for the difference between local and foreign genotypes. In the future, a wider study using MLST to catalogue more *C. albicans* isolates would be beneficial. Further extending research into the cause for the disproportionate representation of isolates expressing pathogenicity in Clade 1 as shown by MLST could lead to a better understanding of *C. albicans* pathogenicity. This may be achieved by the use of virulence gene expression analysis, growth behaviours in systems replicating host conditions as well as protease activity assessments. This study has been limited in terms of database submission due to insufficient funds for isolate submission, a task which should be undertaken in the future.

REFERENCES

1. Andes DR, Safdar N, Baddley JW, *et al*. The epidemiology and outcomes of invasive Candida infections among organ transplant recipients in the United States: results of the Transplant-Associated Infection Surveillance Network (TRANSNET). Transpl Infect Dis. 2016; 18(6): 921-31.

2. Pfaller MA, Castanheira M. Nosocomial candidiasis: Antifungal stewardship and the importance of rapid diagnosis. Med Mycol. 2016; 54(1): 1-22.

3. Haydar A. Epidemiology and Outcomes of Candidaemia among Adult Patients Admitted at Hospital Universiti Sains Malaysia (HUSM): A 5-Year Review. IIUM Med J Malaysia [Internet]. 2018; 17(1). Available from: https://journals.iium. edu.my/kom/index.php/imjm/article/view/300

4. Sobel JD. Vulvovaginal candidosis. Lancet. 2007; 369(9577): 1961-71.

5. Sawadogo PM, Zida A, Soulama I, *et al*. Genotype analysis of clinical candida albicans isolates using pcrs targeting 25s rdna and alt repeat sequences of the rps and antifungal susceptibility in ouagadougou (Burkina Faso). Infect Drug Resist. 2019; 12: 3859–66.

6. Iwata T, Hattori H, Chibana H, *et al*. Genotyping of Candida albicans on the basis of polymorphisms of ALT repeats in the repetitive sequence (RPS). J Dermatol Sci. 2006; 41(1): 43-54.

7. Odds FC, Bougnoux ME, Shaw DJ, *et al*. Molecular phylogenetics of Candida albicans. Eukaryot Cell. 2007; 6(6): 1041-52.

8. Tay ST, Chai HC, Na SL, Ng KP, Soo-Hoo TS. Molecular subtyping of clinical isolates of Candida albicans and identification of Candida dubliniensis in Malaysia. Mycopathologia. 2005; 159(3): 325-9.

9. Tantivitayakul P, Panpradit N, Maudcheingka T, Klaophimai A, Lapirattanakul J. Genotyping of Candida albicans and Candida dubliniensis by 25S rDNA analysis shows association with virulence attributes in oral candidiasis. Arch Oral Biol. 2019; 97: 18-24.

10. Saranpal CS, Chong PP, Ng KP, Lim CSY. Genetic relatedness of Candida albicans bloodstream infection clinical isolates in Malaysia. Malays J Microbiol. 2015; 11(3): 294-9.

11. Bougnoux ME, Morand S, D'Enfert C. Usefulness of multilocus sequence typing for characterization of clinical isolates of Candida albicans. J Clin Microbiol. 2002; 40(4): 1290-7.

12. Tavanti A, Gow NAR, Senesi S, Maiden MCJ, Odds FC. Optimization and Validation of Multilocus Sequence Typing for Candida albicans. J Clin Microbiol. 2003; 41(8): 3765-76.

13. Tavanti A, Davidson AD, Fordyce MJ, Gow NAR, Maiden MCJ, Odds FC. Population Structure and Properties of Candida albicans,as Determined by Multilocus Sequence Typing. J Clin Microbiol. 2005; 43(11): 5601-13.

14. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016; 33(7): 1870-4.

15. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: An application of Simpson's index of diversity. J Clin Microbiol. 1988; 26(11): 2465-6.

16. Carriço JA, Silva-Costa C, Melo-Cristino J, *et al*. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant Streptococcus pyogenes. J Clin Microbiol. 2006; 44(7): 2524-32.

17. Dalvand A, Katiraee F, Joozani RJ, Shokri H. Genotyping of Candida albicans isolated from animals using 25S ribosomal DNA and ALT repeats polymorphism in repetitive sequence. Curr Med Mycol. 2018; 4(4): 12-9.

18. Zida A, Issiaka S, Sanata B, *et al*. 25S rDNA genotype and antifungal susceptibility of clinical Candida albicans in Ouagadougou (Burkina Faso).

List of Supplementary Tables

Table 1: Distribution of *C. albicans* ABC genotypes, ALT sequence repeat subtypes and clades based on a UPGMA phylogenetic tree analysis. Download link: https://drive.google. com/file/d/1_8QFJgrTEhj0T3vHdY8s4trmg-B0dQlB/view?usp=sharing

Table 2: List of SNPs, their respective genes and loci for all isolates obtained in this study. Redundant sequences have been removed. SNP positions are determined based on their positions in the respective reference genes of *C. albicans* SC5314 available in the NCBI database. Download link: https://drive.google. com/file/d/1Sv5xGvmMp5kqJy3jrhaoSyo6M1r qCP50/view?usp=sharing