JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 2003, p. 5265–5266 Vol. 41, No. 11
0095-1137/03/$08.00+0 DOI: 10.1128/JCM.41.11.5265–5266.2003

Collaborative Consensus for Optimized Multilocus Sequence Typing of *Candida albicans*

M.-E. Bougnoux,[1,2] A. Tavanti,[3] C. Bouchier,[4] N. A. R. Gow,[3] A. Magnier,[4] A. D. Davidson,[3] M. C. J. Maiden,[5] C. d'Enfert,[1] and F. C. Odds[3]*

Unité Biologie et Pathogénicité Fongiques, CNRS URA 2172,[1] and Génopole,[4] Institut Pasteur, Paris, and Laboratoire de Parasitologie-Mycologie, Service de Microbiologie, Hôpital Ambroise-Paré, Université Versailles Saint-Quentin en Yvelines, Boulogne-Billancourt,[2] France, and Department of Molecular & Cell Biology, Institute of Medical Sciences, University of Aberdeen, Aberdeen,[3] and The Peter Medawar Building for Pathogen Research and Department of Zoology, University of Oxford, Oxford,[5] United Kingdom

Received 23 July 2003/Returned for modification 5 August 2003/Accepted 19 August 2003

A panel of 86 different *Candida albicans* isolates was subjected to multilocus sequence typing (MLST) in two laboratories to obtain sequence data for 10 published housekeeping gene fragments. Analysis of data for all possible combinations of five, six, seven, eight, and nine of the fragments showed that a set comprising the fragments *AAT1a*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b* was the smallest that yielded 86 unique diploid sequence types for the 86 isolates. This set is recommended for future MLST with *C. albicans*.

Multilocus sequence typing (MLST) is becoming a widely used approach to microbial isolate differentiation for epidemiological purposes (6). For *Candida albicans*, the species most often involved in deep-organ fungal diseases, MLST was introduced in 2002 (1), and a central internet database has been set up for deposition and analysis of *C. albicans* MLST data from any global source (http://calbicans.mlst.net). This MLST system is based on fragments of six *C. albicans* genes (in alphabetical order): *ACC1*, *ADP1*, *GLN4*, *RPN2*, *SYA1*, and *VPS13*.

A second study of *C. albicans* MLST involved four of these genes plus four other gene fragments, *AAT1a*, *AAT1b*, *MPIb*, and *ZWF1b* (5). Both sets of gene fragments provided highly discriminatory typing systems that gave stable, reproducible results and could distinguish even closely related strains. While in principle the use of as many gene sequences as possible should enhance the discriminatory power of an MLST scheme, in practical and technical terms a compromise is required to provide the maximum level of isolate differentiation with the minimum set of fragments. Our two groups therefore agreed to collaborate by exchanging *C. albicans* isolates that had already been described in the published MLST papers and compiling a data set of sequences based on all 10 fragments that have been used for MLST. Analysis of these data allows us to propose an optimized gene set for routine use in *C. albicans* MLST research.

A total of 92 *C. albicans* isolates (1, 5) were shared between the laboratories. These included duplicate cultures of isolate SC5314 and one culture of CAF2, derived from SC5314. As expected, identical MLST results were obtained for these three isolates, so two of the three data sets were excluded from analysis. Among the remaining 90 unique isolates, incomplete sequence data were obtained for 4, so the results for 86 isolates were analyzed to determine the optimal set of gene fragments for MLST.

The method used for MLST was as previously described (1, 5). Both DNA strands in this diploid fungus were sequenced for each of 10 fragments (Table 1), and the sequences were recorded by the one-letter code for nucleotides from the International Union of Pure and Applied Chemistry nomenclature. For each fragment, each different genotype was assigned a unique number. Diploid sequence types (DSTs) are the numbers assigned to each unique combination of genotypes.

Table 1 summarizes the characteristics of the 10 DNA fragments used for MLST with the 86 *C. albicans* isolates. The sizes of the fragments were similar, ranging from 306 to 491 bases. An internet database, http://cbr-rbc.nrc-cnrc.gc.ca/biovis/candida/, compiled by Whiteway and colleagues with the input of unpublished data from the Stanford *Candida* genome project (http://alces.med.umn.edu/candida/), allows tentative assign-

TABLE 1. Properties of the 10 DNA fragments used for *C. albicans* MLST

Gene fragment	*C. albicans* chromosome[a]	No. of bases analyzed	No. of variable bases		dN/dS[c]	No. of Genotypes	
			This study	Previous study (4)[b]		Found	Per variable base
AAT1a	2	373	10	7	0.17	23	2.3
AAT1b	2	339	6	6	0.08	15	2.5
ACC1	3	407	7	6	0.20	15	2.1
ADP1	1	443	16	15	0.41	23	1.4
GLN4	3	404	11	11	1.00	20	1.8
MPIb	2	375	11	11	0.33	20	1.8
RPN2	1	306	13	11	0.11	20	1.5
SYA1	6	391	13	13	0.34	26	2.0
VPS13	4	403	17	16	0.70	38	2.2
ZWF1b	1	491	9	8	0.12	36	4.0

[a] Tentative assignment of the chromosome on which the fragment is located, from http://cbr-rbc.nrc-cnrc.gc.ca/biovis/candida/.
[b] The number of variable bases has increased because of the larger panel of isolates sequenced.
[c] Calculated according to Nei and Gojobori (3).

* Corresponding author. Mailing address: Department of Molecular and Cell Biology, Institute of Medical Sciences, Foresterhill, University of Aberdeen, Aberdeen AB25 2ZD, United Kingdom. Phone: 44 1224555828. Fax: 44 0224555828. E-mail: f.odds@abdn.ac.uk.

TABLE 2. Discrimination of 86 *C. albicans* isolates by MLST with all possible combinations of five, six, seven, eight, and nine DNA fragments

No. of:		No. (%) of DSTs determined for indicated no. of *C. albicans* isolates (n = 86)								
Fragments	Combinations	<60	60–70	70–74	75–79	80–82	83	84	85	86
5	252	2 (0.8)	60 (23.8)	90 (35.7)	85 (33.7)	13 (5.2)	1 (0.4)	1 (0.4)	0 (0.0)	0 (0.0)
6	210	0 (0.0)	5 (2.4)	42 (20.0)	108 (51.4)	39 (18.6)	8 (3.8)	6 (2.9)	2 (1.0)	0 (0.0)
7	120	0 (0.0)	0 (0.0)	2 (1.7)	46 (38.3)	43 (35.8)	8 (6.7)	12 (10.0)	8 (6.7)	1 (0.8)
8	45	0 (0.0)	0 (0.0)	0 (0.0)	4 (8.9)	21 (46.7)	0 (0.0)	7 (15.6)	10 (22.2)	3 (6.7)
9	10	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (30.0)	0 (0.0)	0 (0.0)	4 (40.0)	3 (30.0)

ment of *C. albicans* genes to individual chromosomes (3). From this database our 10 DNA fragments are probably distributed over five of the eight *C. albicans* chromosomes. The lowest proportion of bases that varied between isolates for any fragment was 3.7%, for *ACC1*, and the highest was 9.4%, for *VPS13*. The ratio of nonsynonymous to synonymous base changes (dN/dS), calculated by the method of Nei and Gojobori (4), was <1 for all fragments except *GLN4* (Table 1), indicating neutral effects of selective pressure for most of the fragments.

The number of different genotypes determined for 86 *C. albicans* isolates from the 10 DNA fragments varied from 15 to 38, with *ZWF1b* giving the highest number of genotypes per variable base (Table 1). On the basis of this ratio, fragments *ADP1*, *RPN2*, *GLN4*, and *MPIb* gave the poorest level of isolate discrimination.

When the genotypes for the 86 *C. albicans* isolates were analyzed across all 10 DNA fragments, 86 unique DSTs were found, indicating perfect discrimination for MLST based on all the fragments. To deduce the minimum set of fragments that could differentiate all 86 isolates by MLST, the number of DSTs for the isolates was determined for every possible combination of five, six, seven, eight, and nine gene fragments. Results are shown in Table 2. With combinations of five fragments, 94% of results showed fewer than 80 DSTs for the isolates, and no combination resulted in 86 unique DSTs for the 86 isolates. The most discriminatory set of five fragments was *AAT1a* + *ACC1* + *ADP1* + *VPS13* + *ZWF1b*, yielding 84 different DSTs (Table 2). The two most discriminatory sets of six fragments added either *MPIb* or *SYA1* to the best five-fragment set, in both cases increasing the number of DSTs obtained to 85. The minimum fragment set of seven—the smallest set that gave 86 unique DSTs for the 86 isolates—was *AAT1a* + *ACC1* + *ADP1* + *MPIb* + *SYA1* + *VPS13* + *ZWF1b*.

A number of considerations influence the choice of the set of gene fragments used for MLST. In principle, the fragments used for typing should be from housekeeping genes that are under stabilizing selective pressure for conservation of function. This property is indicated by the ratio of nonsynonymous to synonymous amino acid changes that result from base variation, which should ideally be less than 1.0 (2). Among our 10 fragments, *GLN4*, with a dN/dS of 1.0 (Table 1), was least suitable for MLST on the basis of this criterion. Four of our 10 fragments yielded fewer than two genotypes per base variation: *ADP1*, *GLN4*, *MPIb*, and *RPN2* (Table 1). In both our laboratories, the greatest number of technical problems with reliable PCR amplification were encountered with *GLN4*. Based on these three considerations, fragment *GLN4* is the least suitable of the 10 to be used for MLST; its absence from the minimum fragment set capable of discriminating all 86 *C. albicans* isolates confirms this.

The two other fragments not included in the minimum seven-fragment, fully discriminatory set were *AAT1b* and *RPN2*. *AAT1b* yielded only six variable bases in its 339-base sequence. Although the differentiating power of these variable bases was the second highest, at 2.5 genotypes per variation (Table 1), its contribution to the DST was poor; it was represented in only one of the three eight-fragment sets that discriminated all 86 isolates. *RPN2* sequence variability represented a low ratio of nonsynonymous to synonymous amino acid changes but also one of the lowest numbers of genotypes per variable base (Table 1). Like *AAT1b*, *RPN2* contributed to a full 86-DST discrimination between the 86 test isolates only as part of sets of eight or more fragments.

On the basis of this study we propose following gene set as an international standard for *C. albicans* MLST: *AAT1a* + *ACC1* + *ADP1* + *MPIb* + *SYA1* + *VPS13* + *ZWF1b*.

We are grateful to the Ministčre de la Recherche et de la Technologie (Programme de Recherche Fondamentale en Microbiologie, Maladies Infectieuses et Parasitaires-Réseau Infections Fongiques) and the Wellcome Trust for grants supporting our fungal MLST research. M.C.J.M. is a Wellcome Trust Senior Research Fellow.

We also acknowledge the British Society for Antimicrobial Chemotherapy and the BBSRC for laboratory support. We acknowledge the several colleagues who have supplied clinical isolates of *C. albicans* to our collection over the last 30 years.

The first two authors contributed equally to this study.

REFERENCES

1. **Bougnoux, M.-E., S. Morand, and C. d'Enfert.** 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans*. J. Clin. Microbiol. **40:**1290–1297.
2. **Jones, N., J. F. Bohnsack, S. Takahashi, K. A. Oliver, M.-S. Chan, F. Kunst, P. Glaser, C. Rusniok, D. W. M. Crook, R. M. Harding, N. Bisharat, and B. G. Spratt.** 2003. Multilocus sequence typing system for group B *Streptococcus*. J. Clin. Microbiol. **41:**2530–2546.
3. **Magee, P. T., and H. Chibana.** 2002. The genomes of *Candida albicans* and other *Candida* species. *In* R. A. Calderone (ed.), *Candida* and candidiasis. ASM Press, Washington, D.C.
4. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
5. **Tavanti, A., N. A. R. Gow, S. Senesi, M. C. J. Maiden, and F. C. Odds.** 2003. Optimization and validation of multilocus sequence typing for *Candida albicans*. J. Clin. Microbiol. **41:**3765–3776.
6. **Urwin, R., and M. C. J. Maiden.** 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. **11:**479–487.