

Original Research Article

Phylogeny and Diversity of *Candida albicans* Vaginal Isolates from Three Continents

Karima Soudi^{1,2}, Marie-Elisabeth Bougnoux³, Rhimou El Hamoumi¹, Dorothee Diogo³,
Naima El Mdaghri⁴, Christophe d'Enfert³ and Aziza Razki²

¹Laboratoire d'Ecologie et d'Environnement, Faculté des Sciences Ben M'sik.
Casablanca, Maroc.

²Laboratoire de Pathogénicité Fongique, Institut Pasteur, 1, Place Louis Pasteur,
Casablanca 20100. Maroc.

³ Institut Pasteur, Unité Biologie et Pathogénicité Fongiques, INRA USC2019,
Paris, France.

⁴Directrice de l'Institut Pasteur. Casablanca, Maroc.
*Corresponding author

A B S T R A C T

Keywords

Multilocus
sequence
typing,
*Candida
albicans*,
vaginal
isolates

Multilocus sequence typing data for 94 *Candida albicans* isolates cultured from vaginal samples in Morocco, France and Brazil were analyzed for geographically related variations. No significant differences were found between distributions of diploid sequence types (DSTs) and in the clade distributions of isolates from the three geographical sources; 70DSTs were identified. DSTs 69 and 79 were observed for 9 and 6 isolates, respectively. Almost all of the isolates could be assigned by MLST to one of 17 clades, 95% of the isolates were assigned to one of the clades previously described in *C. albicans*. Generally no obvious difference in the clade distribution of the vaginal isolates from the three different countries could be observed.

Introduction

Fungal infections have become a prominent problem over the last 25 years (Martin et al., 2003). This is mainly due to the worldwide increase in the number of immunocompromised patients, who are highly susceptible to opportunistic infections, including mycoses (Fridkin and Jarvis, 1996). Among fungal pathogens responsible for opportunistic infections, species of the genus *Candida* have a central contribution. These species can infect most patients to some degree and are responsible

for superficial infections such as Oropharyngeal Candidiasis and vulvo vaginal candidiasis. *Candida* species associated with man as human commensals and opportunistic pathogens (Calderone, 2002; Odds, 1988).

Candida albicans is the most frequently encountered among *Candida* species associated with man as human commensals and opportunistic pathogens (Calderone, 2002; Odds, 1988). As for vaginal

candidiasis, it has been reported that 75% of all women will experience at least one episode of *Candida* vaginitis in their life-time (Holmes and Hansfield, 1998). Vaginal candidiasis is caused by *C. albicans* predominantly at least 80% (Holmes and Hansfield, 1998). Non *C. albicans* *Candida* species, predominantly *Candida glabrata*, are responsible for the remainder of cases (Dodgson et al., 2003).

The advent of high throughput DNA sequence analysis has been accompanied by the development of highly resolving, reproducible and portable typing methodologies. Among these diverse methods, Multilocus Sequence Typing (MLST) was introduced in 1998 by Maiden et al., for typing bacterial isolates and has now been extended to a wide variety of species. It's becoming the method of choice for typing bacterial and fungal isolates (Bougnoux et al., 2004; Dodgson et al., 2003; Litvintseva et al., 2006; Tavanti et al., 2005a; Taylor and Fisher, 2003; Urwin and Maiden, 2003) including *C. albicans* (Bougnoux et al., 2004; Bougnoux et al., 2003; Bougnoux et al., 2002), *Candida glabrata* (Dodgson et al., 2003), *Candida krusei* (Jacobsen et al., 2007), and *Candida tropicalis* (Tavanti et al., 2005a).

MLST combines a high discriminatory index with exceptional portability since MLST data can be accessed and updated (linked via <http://www.mlst.net/> or <http://pubmlst.org/>). MLST involves obtaining the sequences of internal fragments of seven house-keeping genes for each strain of a particular species. Sequence data from multiple loci can be used not only to analyze clonality and recombination for strain populations within a species but also to provide evidence for species level differentiation (Taylor and Fisher, 2003). Isolates of *C. albicans* analyzed by multilocus sequence typing (MLST)

(Blignaut et al., 2002; Munro and Hube, 2002) or DNA finger printing with the moderately repetitive oligonucleotide Ca3 (Blignaut et al., 2002; Pujol et al., 2002; Soll and Pujol., 2003) can be assigned to subsets of closely related strain types, referred to, for convenience, as clades (Odds and Jacobsen, 2008; Soll and Pujol, 2003). The main clades to which individuals strains are assigned are essentially the same by both typing approaches (Fidel et al., 2004; Odds et al., 2007), and approximately 70% of the large numbers of *C. albicans* isolates typed so far belong to one of the four largest clades, numbered 1 to 4 (Odds et al., 2007).

In this study, we searched that *C. albicans* MLST database (<http://test1.mlst.net/>) for subsets of isolates from different geographical sources that had the same anatomical origin and compared the MLST results for these three sets of isolates for similarities and differences that reflected either their pathological or their geographical origins.

Materials and Methods

C. albicans strain collection

A total of 94 independent *C. albicans* strains recovered during vaginal colonization. The clinical isolates were geographically diverse; it is from 3 countries representing 3 continents and comprised 60 isolates from healthy Moroccan women (North Africa), 18 isolates from French women representing Western Europe (systematics was set the time of childbirth) and 16 isolates from immunocompromised Brazilian women infected by HIV and/or with a vaginitis (South America). Yeast strains were isolated and identified primarily by test filamentation, *C. albicans* strains were reidentified with use of CHROM Agar *Candida* (CHROMagar).

MLST status determination

MLST was based on seven housekeeping genes, including loci AAT1a, ACC1, ADP1, MPIb, SYA1, VPS13 and ZWF1b (Calderone, 2002). PCRs were carried out with mixtures containing 1 µl (for DNA extraction from culture-positive samples) of DNA template in a final volume of 50 µl. Final concentrations of components of the PCR were: 10x PCR buffer, 5 mM MgCl₂, 100 µM of each primer, 5 mM dNTP mixture and 5 Taq polymerase. The cycling conditions for first round of PCR was Initial denaturation at 94°C for 4 min, followed by 30 cycles of denaturation at 94°C for 45s, annealing at 55°C for 1 min, extension at 72°C for 1min and final extension at 72°C for 10 min. DNA sequencing was performed by using the same primers used in PCR and on both strands.

Results of MLST are retained in the database in the form of genotype numbers, which define unique sequences for pairs of alleles, and diploid strain types (DSTs), which define unique combinations of genotypes. This methodology for determining DSTs by MLST has been described previously in 2003 by Bougnoux et al.

Analysis of MLST data

The results for the variable sites from these vengene fragments sequenced were concatenated into a single sequence. To cope with heterozygous code data, each base in the concatenated sequences of the polymorphic sites was transformed into two bases: the same if the base is homozygous code, so, e.g., the sequence ACGT would emerge as AACCGGTT, and as the component bases for heterozygous codes, so, e.g., AWST would come out as AAATCGTT. These derived sequences

were used to generate a single-linkaged endrogram for the 94 isolates by the unweighted pair-group method using arithmetic averages (UPGMA) and determined by p-distance as implemented by MEGA3 software (Kumar et al., 2004). Assignment of isolates to clades was based on the results of a large-scale analysis of 1391 isolates using a cut-off p-distance of 0.04 in a dendrogram generated by UPGMA (Odds et al., 2007). The isolates were thus each designated as a member of 17 MLST clades or as singleton, not assignable to a UPGMA-based clade.

Results and Discussion

MLST data for isolates from Morocco

MLST of 60 *C. albicans* isolates resulted in delineation of 50 separate DSTs in the study panel. 45 (90%) corresponded to a single isolate whereas 5 were shared by multiple isolates: 6 isolates were type 79 and 4 isolates were type 69 and examples of this type were obtained from geographical sources in all parts of the world. The following three other DSTs were represented by less than 10 isolates: DSTs 1363 by 3 isolates, 423 and 1372 by 2 isolates.

MLST data for isolates from France and Brazil

25 DSTs were identified. DSTs 69 was the most commonly encountered strain type: among the 34 isolates from France and Brazil, he was represented by 5 isolates (three from Brazil and two from France) followed by DSTs 435 for 4 isolates (all from Brazil), 209 (one from Brazil and other from France) and finally 1359 with two isolates from Brazil whereas 62% corresponded to a single isolate.

Clade assignments of isolates

Figure 1 illustrates the UPGMA dendrogram based on MLST data for 94 vaginal isolates (M, Morocco; F, France; B, Brazil). 17 clades were defined; isolates that did not cluster with others within the 0.04 cutoff and cluster containing fewer than 10 isolates were labeled singletons (Table 1). Four isolates received the singleton designation, leaving 89 isolates assigned to clades and 95% of the isolates were assigned to one of the clades previously described in *C. albicans* (Calderone, 2002). A vertical bar to the right of each clade indicates the clade designation and relative size, respectively.

Clade 1, the largest clade, accounted for 47% of the isolates tested (except for the four singletons). It contained the isolates from the three different countries, with DST69 as the most common individual strain type within the clade followed by DST79. Clades 2 and 4 were the next largest clades, with 21% of the all isolates (except for the four singletons) being split among them. Clade 11, previously represented by a small number of isolates (Tavanti et al., 2005b), emerged also in the present study as the fourth most populous clade with its 7 assigned isolates constituting 8% of all isolates sequenced.

No difference in the clade distribution of the 16 vaginal isolates from immunocompromised Brazilian women infected by HIV and/or with a vaginitis and in comparison with other vaginal isolates from healthy women. Full details of the isolate panel are given in the Supplementary online material (Table 2).

Candida species are important human pathogens that are best known for causing opportunistic infections in immunocompromised hosts (e.g. transplant patients,

AIDS sufferers, cancer patients). Most of these infections are opportunistic, depending on environmental changes in the host, for example, a weakening of the immune response, as occurs during transplant procedures, or a radical change in the rest of the microbial population. This behaviour is characteristic of the other medically important *Candida* species, such as *C. albicans*, *C. glabrata*, *C. dubliensis* and *C. parapsilosis*, as well.

Candida albicans is a facultative commensal yeast of the human gastrointestinal and genital mucosa. Medical progress in developing countries has been accompanied by an increase in opportunistic candidiasis. This species is most frequently identified in vaginal specimens. However, the global diversity of *C. albicans* strains isolated from carriage populations remains uncharacterized.

Vaginal thrush remains a frequent problem and most women suffer at least one such infection during the course of their lives. Most of these infections are successfully treated using azole antifungal drugs, for example. However, about 5% of these women suffer recurrent *Candida* infections, the basis of which appears to have more to do with the immune status of the patients than the properties of the infecting *Candida* strain (Fidel et al., 2004). However, given the medical importance of this fungus, it is hardly surprising that most research on *C. albicans* has been directed towards an understanding of this organism as a pathogen and the treatment of *Candida* infections.

To make a geographical strain comparison with isolates as closely matched as possible for non-geographical variables, we chose isolates that came from one anatomical source (vagina) of clinical interest, and

which originated in well separated regions of the planet. Hence, our choice of vaginal isolates was from 3 countries representing 3 continents: Morocco (North Africa: 60 isolates), France (Western Europe: 18 isolates) and Brazil (South America: 16 isolates). To this aim, we constructed a UPGMA tree based on the concatenated MLST sequences of all isolates (Figure 1). 70 DSTs were identified. DSTs 69 and 79 were observed for 9 and 6 isolates, respectively. Interestingly, the 9 isolates with DST 69 have been collected from carriers from the 3 countries. 95% of the isolates were assigned to one of the clades previously described in *C. albicans* (Odds et al., 2007). 68 (72%) fell into 4 of the 5 previously described major clades (1, 2, 4 and 11) (Odds et al., 2007). 47% were assigned to clade 1. The table1 summarizes the clade assignment of all the isolates. The majority of isolates from each of the three countries belonged to clade1 (47%) with DST69 as the most common individual strain type within the clade followed by DST79, but the proportions of clade1 isolates from each country ranged from 28% for France, 42% for Morocco and 75% for Brazil (Table2).

The proportion of French isolates in clade 2 was two times higher than the proportion of Moroccan isolates (11% and 5% respectively). Among all isolates include 4 (14 isolates), 17% came from Morocco and 22% came from France, while 11% of clade 11 isolates came from French and 8% from Morocco.

The proportion of Moroccan isolates that were singletons or assigned to clades 5–17 was substantially higher than for French isolates; any singleton was detected in the Brazilian isolates.

Generally no obvious difference in the clade distribution of the vaginal isolates from the

three different countries could be observed and reflected either their pathological or their geographical origins. We purpose that the selection of the panel was restricted to assign the isolates coming from the three countries into fifth major clade (clade 3) and Brazilian isolates to the other major clades except clade1.

Generally, clade 1 is the most populous strain group in this and the previous study and correlates with the most populous cluster observed in various studies based on DNA fingerprinting and other genome-based technologies (Lott et al., 1999; Pujol et al., 1997; Schmid et al., 1999; Odds, 1988). This demonstrates that the most common, globally distributed *C. albicans* strain types are those of MLST clade1, and DST 69 is both the most frequently encountered member of clade1 and its putative founding type. As the results of MLST of 1,391 *C. albicans* isolates resulted in delineation of 1,005 separate DSTs. DST 69 was the most commonly encountered strain type: 65 isolates were type 69 following by DSTs 37 and DSTs 79, and examples of this type were obtained from geographical sources in all parts of the world (Frank et al., 2007).

In the present study no obvious difference in the clade distribution of the vaginal isolates from the three different countries could be observed and reflected either their pathological or their geographical origins. We purpose that the selection of the panel was restricted to assign the isolates coming from the three countries into fifth major clade (clade3) and Brazilian isolates to the other major clades except clade1.

Shunji Takakura et al. (2008) in their study the choice of blood and vaginal isolates was from two relatively small geographical regions, England/Wales and Japan, and one large region, the USA, whose population includes large numbers of immigrants from

England and Japan. They found geographical differences were strongly indicated, with clade 2 strains clade 3 strains particularly prominent among US isolates, clade 2 strains particularly prominent among English isolates but no statistically significant difference in MLST strain type between all blood and all vaginal isolates. Finally, the sequencing of the genome of *Candida albicans* and those of several other

medically relevant *Candida* species has provided a major impetus for *Candida* comparative and functional genomic analysis. These have provided a fascinating insight into the molecular and cellular biology of these fungi and should pave the way for the development of more sensitive diagnostic strategies and novel antifungal therapies.

Figure.1 UPGMA dendrogram based on MLST data for 94 vaginal isolates (M, Morocco; F, France; B, Brazil). The dashed line shows the cutoff at a P distance of 0.04 used to delineate clusters designated as clades.

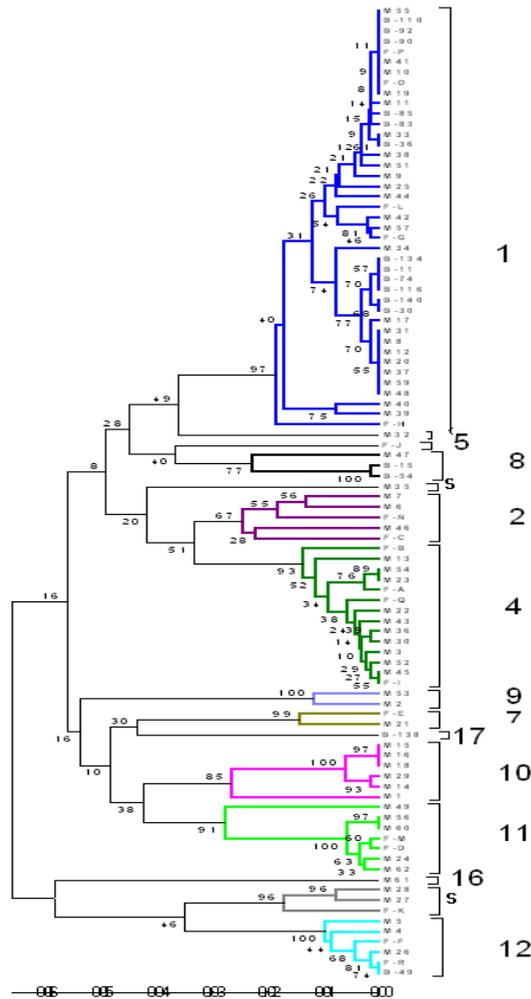


Table.1 Geographical and anatomical origins of *C. albicans* isolates in 17 clades plus single to isolates

Geographical origin	Anatomical origin	Number of isolates	Clades																	
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Singletons
Morocco	Vagina	60	25	3	0	10	0	0	1	1	2	6	5	3	0	0	0	1	0	3
France	Vagina	18	5	2	0	4	1	0	1	0	0	0	2	2	0	0	0	0	0	1
Brazil	Vagina	16	12	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	1	0
Total		94	42	5	0	14	1	0	2	3	2	6	7	6	0	0	0	1	1	4

Table.2 Break down by geographical origin of MLST clades for 94 vaginal isolates

Isolates from Brazil			
	Number in clade	As % of all Brazilian isolates	As % of all isolates in
Clade1 (42isolates)	12	75	29
Clade 2 (5 isolates)	0	0	0
Clade 4 (14 isolates)	0	0	0
Clade 11 (7 isolates)	0	0	0
Other Clades and singletons	4	25	15

Isolates from France			
	Number in clade	As % of all French isolates	As % of all isolates in clade
Clade1 (42isolates)	5	28	12
Clade 2 (5 isolates)	2	11	40
Clade 4 (14 isolates)	4	22	29
Clade 11 (7 isolates)	2	11	29
Other Clades and singletons	5	28	19

Isolates from Morocco			
	Number in clade	As % of all Moroccan isolates	As % of all isolates in clade
Clade1 (42isolates)	25	42	60
Clade 2 (5 isolates)	3	5	60
Clade 4 (14 isolates)	10	17	71
Clade 11 (7 isolates)	5	8	71
Other Clades and singletons	17	28	65

In the present study, we have reported i) the first Molecular Typing data concerning vaginal carriage of *Candida albicans* in healthy women from Morocco, ii) determined the genetic diversity of *C. albicans* strains isolated from vaginal samples in Morocco, iii) examined genetic relationships among *C. albicans* recovered during vaginal colonization from 3 countries representing 3 continents: Morocco (North Africa), France (Western Europe) and Brazil (South America). MLST of *C. albicans* isolates originated from Morocco revealed an extensive diversity. However these isolates were predominantly distributed into 6 clades including 4 of the previously identified 5 major clades and 2 of the previously identified minor clades.

MLST showed that the distribution in clades of *C. albicans* vaginal isolates sampled in Morocco, France and Brazil did not differ significantly and that *C. albicans* isolates had a phylogenetic distribution mostly similar to that observed in developed countries.

References

- Blignaut, E., Pujol, C., Lockhart, S., Joly, S., Soll, D.R. 2002. Ca3 fingerprinting of *Candida albicans* isolates from human immuno deficiency virus-positive and healthy reveals an ewcladein South Africa. *J. Clin. Microbiol.*, 40: 826–836.
- Bougnoux, M.E., Aanensen, D.M., Morand, S., Theraud, M., Spratt, B.G., d'Enfert, C. 2004. Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. *Infect. Genet. Evol.*, 4: 243–252.
- Bougnoux, M.E., Tavanti, A., Bouchier, C., Gow, N.A.R., Magnier, A., Davidson, A.D., Maiden, M.C.J., d'Enfert, Odds, F.C. 2003. Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. *J. Clin. Microbiol.*, 41: 5265–5266.
- Bougnoux, M.-E., Morand, S., d'Enfert, C. 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans*. *J. Clin. Microbiol.*, 40: 1290–1297.
- Calderone, R.A. 2002. *Candida* and Candidiasis. ASM Press, Washington, DC.
- Dodgson, A.R., Pujol, C., Denning, D.W., Soll, D.R., Fox, A.J. 2003. Multilocus sequence typing of *Candida glabrata* reveals geographically enriched clades. *J. Clin. Microbiol.*, 41: 5709–5917.
- Fidel Jr., P.L. 2004. History and new insights into host defense against vaginal candidiasis. *Trends Microbiol.*, 12: 220–227.
- Frank C. Odds, Marie-Elisabeth Bougnoux, Duncan J. Shaw, Judith M. Bain, Amanda D. Davidson, Dorothée Diogo, Mette D. Jacobsen, Maud Lecomte, Shu-Ying Li, Arianna Tavanti, Martin C. J. Maiden, Neil A. R. Gowand Christophe d'Enfert, 2007. Eucaryotic Cell, Vol. 6, Pp. 1041–1052.
- Fridkin, S.K., Jarvis, W.R. 1996. Epidemiology of nosocomial fungal infections. *Clin. Microbiol. Rev.*, 9: 499–511.
- Holmes, K.K., Hansfield, H.H. 1998. Sexually transmitted diseases: Overview and clinical approach. In: Fauci, A.S., Braunwald, E., Isselbacher, K.J., Wilson, J.D., Martin, J.B., Kasper, D.L., Hauserand, S.L., Longo, D.L. (Eds.), Harrison's principles of internal

- medicine, 14th edn. McGraw-Hill, London.
- Jacobsen, M.D., Gow, N.A.R., Maiden, M.C.J., Shaw, D.J., Odds, F.C. 2007. Strain typing and determination of population structure of *Candida krusei* by multilocus sequence typing. *J. Clin. Microbiol.*, 45: 317–323
- Kumar, S., Tamura, K., Nei, M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform*, 5: 150–163.
- Litvintseva, A.P., Thakur, R., Vilgalys, R., Mitchell, T.G. 2006. Multilocus sequence typing reveals three genetic sub populations of *Cryptococcus neoformans* var. *grubii* (Serotype A), including a unique population in Botswana. *Genetics*, 172: 2223–2238.
- Lott, T.J., Holloway, B.P., Logan, D.A., Fundyga, R., Arnold, J. 1999. Towards understanding the evolution of the human commensal yeast *Candida albicans*. *Microbiology*, 145: 1137–1143.
- Martin, G.S., Mannino, D.M., Eaton, S., Moss, M. 2003. The epidemiology of sepsis in the United States from 1979 through 2000. *N. Engl. J. Med.*, 348: 1546–1554.
- Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA.*, 95: 3140–3145.
- Munro, C., Hube, B. 2002. Antifungal therapy at the HAART of viral therapy. *Trends Microbiol.*, 10: 173–177.
- Odds, F.C. 1988. *Candida* and Candidosis, seconded. Bailliere Tindall, London, UK.
- Odds, F. C., Bougnoux, M.-E., Shaw, D.J., Bain, J.M., Davidson, A.D., Diogo, D., Jacobsen, M.D., Lecomte, M., Li, S.-Y., Tavanti, A., Maiden, M.C.J., Gow, N.A.R., d'Enfert, C. 2007. Molecular phylogenetics of *Candida albicans*. *Eukaryotic. Cell*, 6: 1041–1052.
- Pujol, C., Joly, S., Lockhart, S.R., Noerl, S., Tibayrenc, M., Soll, D.R. 1997. Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis, and southern blot hybridization with the moderately repetitive DNA probe Ca3 for finger printing *Candida albicans*. *J. Clin. Microbiol.*, 35: 2348–2358.
- Pujol, C., Pfaller, M., Soll, D.R. 2002. Ca3 fingerprinting of *C. albicans* blood stream isolates from the United States, Canada, South America and Europe reveals a European clade. *J. Clin. Microbiol.*, 40: 2729–2740.
- Schmid, J., Herd, S., Hunter, P.R., Cannon, R.D., Yasin, M.S.M., Samad, S., Carr, M., Parr, D., McKinney, W., Schousboe, M., Harris, B., Ikram, R., Harris, M., Restrepo, A., Hoyos, G., Singh, K.P. 1999. Evidence for a general-purpose genotype in *Candida albicans*, highly prevalent in multiple geographical regions, patient types and types of infection. *Microbiology*, 145: 2405–2413.
- Shunji Takakura, Satoshi Ichiyama, Judith M. Bain, Amanda D. Davidson, Mette D. Jacobsen, Duncan J. Shaw, Neil A.R. Gow, Frank C. Odds 2008. Comparison of *Candida albicans*

- strain types among isolates from three countries. *Int. J. Med. Microbiol.*, 298: 663–668.
- Soll, D.R., Pujol, C. 2003. *Candida albicans* clades. *FEMS Immunol.Med. Microbiol.*, 39: 1–7.
- Odds, F.C., Jacobsen, M.D. 2008. Multilocus sequence typing of pathogenic *Candida* species. *Eukaryotic. Cell*, 7: 1075–1084.
- Tavanti, A., Davidson, A.D., Fordyce, M.J., Gow, N.A.R., Maiden, M.C.J., Odds, F.C. 2005b. Population structure and properties of *Candida albicans*, as determined by multilocus sequence typing. *J. Clin. Microbiol.*, 43: 5601–5613.
- Tavanti, A., Davidson, A.D., Johnson, E.M., Maiden, M.C., Shaw, D.J., Gow, N.A., Odds, F.C. 2005a. Multilocus sequence typing for differentiation of strains of *Candidatropicalis*. *J. Clin. Microbiol.*, 43: 5593–5600.
- Taylor, J.W., Fisher, M.C. 2003. Fungal multilocus sequence typing-it's not just for bacteria. *Curr. Opin. Microbiol.*, 6: 351–356.
- Urwin, R., Maiden, M.C. 2003. Multilocus sequence typing: a tool for global epidemiology. *Trends Microbiol.*, 11: 479–487.