Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

A population structure analysis of *Candida dubliniensis* using multilocus sequence typing and an investigation of the molecular mechanism(s) of clade-specific resistance to 5-fluorocytosine

A thesis submitted to the University of Dublin in fulfilment of the requirements for the degree Doctor of Philosophy by

Brenda McManus

December 2009

Microbiology Research Unit, Division of Oral Bioscience,
Dublin Dental School and Hospital,
University of Dublin,
Trinity College

Declaration

I hereby declare that this thesis has not previously been submitted for a degree at this or any other university, and that it represents my own unaided work, except where duly acknowledged in the text. I agree that this thesis may be lent or copied at the discretion of the librarian, Trinity College Dublin.

Brenda Mc Manus.

Brenda McManus, B.A. (Mod) Microbiology

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" ("I found it!") but rather "hmm....that's funny..."

Isaac Asimov

This thesis is dedicated to the memory of my grandmother Rose Murphy,
a devoted woman of unconditional compassion and unwavering faith.
An inspiration.

Summary

Candida dubliniensis shares many phenotypic and genetic characteristics with its closest relative, *Candida albicans*, which most likely delayed the identification and subsequent classification of *C. dubliniensis* as a distinct taxon. Previous population structure analyses of *C. dubliniensis* were based on DNA fingerprinting using the species-specific probe Cd25, and sequence analysis of the internal transcribed spacer (ITS) region of the ribosomal gene cluster. These studies showed that *C. dubliniensis* is comprised of three major clades (Cd25 groups I–III) comprising four distinct ITS genotypes (1–4). Multilocus sequence typing (MLST) has proved useful for investigating the population biology of *C. albicans*, identifying many distinct clades. In the present study, MLST was used to investigate the population structure of *C. dubliniensis* for the first time. Combinations of ten loci previously tested for MLST analysis of *C. albicans* were assessed for their discriminatory ability with epidemiologically unrelated *C. dubliniensis* isolates from diverse geographic locations, including representative isolates of the previously identified three Cd25-defined major clades and four ITS genotypes. UPGMA dendrograms generated using the data from these loci revealed a population structure which supports that previously suggested by Cd25 fingerprinting and ITS genotyping, identifying three main MLST clades (termed C1–C3) comprising four ITS genotypes. Application of a common MLST scheme to *C. dubliniensis* and *C. albicans* revealed that the population structure of *C. dubliniensis* is significantly less divergent than that of *C. albicans*. On the basis of the highest number of genotypes per variable base the following eight loci are recommended for MLST analysis of *C. dubliniensis*; *AAT1b, ACC1, ADP1, PMIb, RPN2, ALA1, exVPS13,* and *exZWF1b*. When 14 avian-excrement-associated *C. dubliniensis* isolates recovered from two separate locations in Ireland, approximately 150 km apart, and 36 *C. dubliniensis* isolates recovered from humans in Ireland were compared using these recommended loci, all 50 isolates were found to belong to MLST clade C1 and belonged to ITS genotype 1. However, 13/14 of avian-excrement-associated isolates were found to be genetically distinct from the human isolates. Six new diploid sequence types (DSTs) were identified in 13 of 14 avian-associated-excrement isolates. The remaining isolate was identical to a human isolate, suggesting that transmission may occur between humans and birds. In total, the *C. dubliniensis* MLST scheme has identified 37 DSTs from the 78 human and avian-associated *C. dubliniensis* isolates examined to date.

The MLST clade C3 comprises *C. dubliniensis* isolates recovered exclusively from the Middle East exhibiting high level resistance to 5-fluorocytosine (5FC) ($MIC_{50} \geq 128$ μg/ml). *Candida albicans* also exhibits clade-specific 5FC resistance, most commonly mediated by an Arg101Cys substitution in the *FUR1* gene product. The DNA sequences of the *CdFUR1* gene and the *CdFCY2*-encoded cytosine permease genes were compared between 5FC-resistant and 5FC-susceptible *C. dubliniensis* isolates and revealed no amino acid transitions, indicating that

mutations in these genes were unlikely to be associated with 5FC resistance. Broth microdilution assays using 5-fluorouracil (5FU), the toxic deaminated form of 5FC, showed that both 5FC-resistant and 5FC-susceptible *C. dubliniensis* isolates exhibited similar 5FU MICs, suggesting that the *C. dubliniensis* cytosine deaminase (Fca1p) encoded by the *CdFCA1* gene may play a role in mediating *C. dubliniensis* clade-specific 5FC resistance. Amino acid sequence analysis of Fca1p identified a homozygous Ser29Leu substitution in all 12 5FC-resistant ($MIC_{50} \geq 128$ μg/ml) isolates investigated, which was not present in any of the nine 5FC-susceptible ($MIC_{50} \leq 0.25$ μg/ml) isolates examined. Tetracycline-inducible expression of the *CdFCA1* gene from a 5FC-susceptible *C. dubliniensis* isolate in two separate 5FC-resistant isolates restored 5FC susceptibility ($MIC_{50} \leq 0.25$ μg/ml). However, tetracycline-inducible expression of the *CdFCA1* gene from a 5FC-resistant isolate in a separate 5FC-susceptible isolate did not confer 5FC resistance ($MIC_{50} \leq 0.25$ μg/ml), demonstrating that the Ser29Leu substitution is recessive. Real-time PCR analysis showed no significant difference in *CdFCA1* expression between 5FC-susceptible and 5FC-resistant isolates either in the presence or absence of sub-inhibitory concentrations of 5FC, suggesting that the Ser29Leu substitution in the *CdFCA1* gene was the sole cause of clade-specific 5FC resistance in *C. dubliniensis*.

A combination of culture and *C. dubliniensis*-specific PCR was used to accurately determine the prevalence of *C. dubliniensis* in a normal healthy population of students, an immunocompromised human population of cystic fibrosis patients, as well as in avian-excrement-associated samples. The use of oral rinse sampling and culture-based methods as the sole method of prevalence determination resulted in the recovery of *C. dubliniensis* from the oral cavities of 2.7% of the normal healthy population. A similar analysis carried out using oral swab sampling and a combination of both culture and PCR-based identification methods revealed a three-fold higher prevalence rate (8.3%) in the oral cavities of a separate normal healthy student population, suggesting that prevalence rates may be underestimated when solely using culture-based methods. In the present study, persistent *C. dubliniensis* carriage was identified in the respiratory tracts of 12% of the cystic fibrosis patients investigated. *Candida dubliniensis* was recovered from 3/134 (2.2%) of avian-associated-excrement samples, a much lower prevalence than previously reported from the surface of *Ixodes uriae* ticks living in seabird excrement in a previous study.

In conclusion, the present study is the first to demonstrate that MLST can be used effectively to analyse the population structure of *C. dubliniensis*. The application of a common MLST scheme to *C. dubliniensis* and *C. albicans* enabled a comparative population structure analysis, which revealed that the population structure of *C. dubliniensis* is significantly less divergent than that of *C. albicans*. The present study also demonstrated for the first time that the primary cause of clade C3-specific 5FC resistance of *C. dubliniensis* isolates recovered in the Middle East is a recessive Ser29Leu substitution in Fca1p.

Table of Contents

Chapter 2. General materials and methods

Chapter 3. A study of the prevalence of *Candida dubliniensis* in a normal healthy and in an immunocompromised human population, and recovery of *C. dubliniensis* from avian-excrement-associated sources

Chapter 4. Population analysis of *Candida dubliniensis* using multilocus sequence typing (MLST), and a study of the genetic relationship of *C. dubliniensis* isolates recovered from human and avian-excrement-associated sources

Chapter 7. General Discussion

Index of Figures

Page numbers refer to the text page preceding the figure(s)

Index of Tables

Page numbers refer to the text page preceding the table(s)

Acknowledgments

First and foremost I wish to thank my supervisor Professor David Coleman for his constant encouragement, patience, support and understanding over the last four years. Thank you for motivating me both to undertake and (hopefully) complete a Ph.D. I am also deeply indebted to Professor Derek Sullivan for all his valuable help and suggestions throughout the last four years, and to Dr. Gary Moran for always making himself available to help with the more technical aspects of the work. I am especially grateful to Dr. Emmanuelle Pinjon and Dr. Judy Higgins, whose consistently good advice, support and friendship made the day to day aspects of the work so much easier.

I am deeply indebted to Dr. Christophe d'Enfert, Dr. Marie-Elisabeth Bougnoux and Dorothée Diogo for being so welcoming during my stay in Paris, and for all the help they provided me with in the MLST study and publications.

I wish to acknowledge the funding for this project by the Microbiology Research Unit, which ensured that I was never short of anything required for my research, and had every opportunity to present my work at conferences, and to learn techniques and applications abroad when needed. Thank you to Mary O'Donnell for maintaining the Microbiology Research Unit, and ensuring all required instruments and consumables were supplied, thus enabling the laboratory to consistently operate to full capacity. I also wish to acknowledge the Moyne Institute preparation room for supplying all culture media to the laboratory.

I am grateful to Dr. Ronnie Russell and Dr. Philip Murphy for providing me with the clinical samples that were examined in the present study, and to Dr. José Pontón, Dr. Silvia Borecká-Melkusová and Prof. Helena Bujdákova for supplying me with additional *C. dubliniensis* isolates. I wish to acknowledge the help of Dr. Miles Nunn at Oxford, whose own research prompted mine to go in a much unexpected direction. At the beginning of my research project I would never have imagined that its completion would require a pair of wellington boots! Many thanks also to Dr. Oscar Merne who so willingly shared some of his vast knowledge of ornithology with me, and helped me to collect samples on the beach in the middle of an Irish winter.

Many sincere thanks to all of the other staff and postgraduate students in the DDSH, both past and present, who have given me support and made me laugh throughout the last four years. I will never forget the piggy-ducky kidnapping; the funniest week in the lab by far. I am grateful to the masterminds behind it for making

me laugh for an entire week, and then returning the hostages! Thank you especially to Emma and Maria whose constant friendship and communication over the last months made the writing up process so much easier to bear!

I thank my brothers and sister; Stephen, Ciarán and Sinéad for their support and patience throughout the last four years. I also wish to express my gratitude to "the Howthies" Anna, Lorna, James, Mick, Alan, Eva, Alana, Damien, Steven, Lavan and Roslyn, as well as to Rob, Geoff, Claire, Orla, Lucy and Dervla. Thank you all for always being around to "put up" or "act up" with me throughout the last ten years. I look forward to "acting up" with you all for many more years to come!

Finally, my deepest appreciation is for my parents Brendan and Imelda; thank you for passing on your sense of ambition and determination to me, by either genetics or by example. I could not have completed this work without these traits, or you.

Abbreviations

AIDS	acquired immunodeficiency syndrome
A540, A600	absorption at 540 nm or 600 nm, respectively
ATP	adenosine 5' -triphospate
Azole	synthetic *N*-substituted azoles, including the imidazole and triazole antifungal antibiotics
BMD	broth microdilution
bp	base pair
Bq	Becquerel
BURST	based upon related sequence types
cfu	colony forming units
Ci	curie
cm	centimetre
cpm	counts per minute
Δ	deletion (of a gene)
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	dideoxynucleoside
DOX	doxycycline
DST	diploid sequence type
EDTA	ethylenediamine tetraacetic acid
e.g.	for example
et al.	and others
5FC	5-fluorocytosine
5FU	5-fluorouracil
g	gram
g	gravitational force
h	hour
HIV	human immunodeficiency virus

IPTG	isopropyl-ß-D-thiogalactopyranoside
i.e.	that is
kb	kilobase pair
L agar	Luria agar
L broth	Luria broth
l	litre
M	molar
Mb	megabase
mg	milligram
µg	microgram
ml	millilitre
µl	microlitre
MIC	minimum inhibitory concentration
min	minute
NCCLS	National Committee for Clinical laboratory Standards
NJ	neighbour joining
nm	nanometre
no.	number
$O.D_X$	optical density at x nm
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDA	potato dextrose agar
PFGE	pulsed-field gel electrophoresis
p.f.u.	plaque forming units
RAPD	random amplified polymorphic DNA
rDNA	ribosomal DNA
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute

s	second
sarcosyl	*N* - lauroylsarcosine, sodium salt
S_{AB}	similarity coefficient
SDS	sodium dodecyl sulphate
SSC	salt, sodium citrate
ST	sequence type
TBE	tris-borate EDTA
Tris	tris (hydroxymethyl) aminoethane
UK	United Kingdom
UPGMA	unweighted pair group method according to arithmetic averages
USA	United States of America
U.V.	ultraviolet
v/v	% "volume in volume" expresses the number of millilitres of an active constituent in 100 millilitres of solution
w/v	% "weight in volume" expresses the number of grams of an active constituent in 100 grams of solution or mixture
X-gal	5-bromo-4-chloro-indoyl-ß -D-galactoside
~	approximately
>	greater than
<	less than
≥	greater than or equal to
≤	less than or equal to

Publications

Much of the work described in this thesis has been published in international journals and as a book chapter as listed below. Copies of published papers are included at the end of the thesis.

1. **McManus, B.A., Coleman, D.C., Moran, G.P., Pinjon, E., Diogo, D., Bougnoux, M-E., Boreká-Melkusová, S., Bujdáková, H., Murphy, P., d'Enfert, C. & Sullivan, D.J. (2008).** Multilocus sequence typing reveals that the population structure of *Candida dubliniensis* is significantly less divergent than that of *Candida albicans*. *J Clin Microbiol* **46**:652-664. DOI:10.1128/JCM.01574-07. PMID: 18057125.

2. **McManus, B.A., Sullivan, D.J., Moran, G.P., d'Enfert, C., Bougnoux, M-E., Nunn, M.A. & Coleman, D.C. (2009).** Genetic differences between avian and human isolates of *Candida dubliniensis*. *Emerg Infect Dis* **15**:1467-1470. DOI:10.3201/eid1509.081660.

3. **McManus, B.A., Moran, G.P., Higgins, J.A., Sullivan, D.J. & Coleman, D.C.** A Ser29Leu substitution in the cytosine deaminase Fca1p is responsible for clade-specific 5-flucytosine resistance in *Candida dubliniensis*. *Antimicrob Agents Chemother* **53:**4678-4685. DOI:10.1128/AAC.00607-09.

4. **Moran, G.P., McManus, B.A., Coleman, D.C. & Sullivan, D.J.** (In Press) Molecular Epidemiology of *Candida* species. In *Handbook of Pathogenic Yeasts.* Edited by R. Ashbee and E. Bignell. London: Springer.

Chapter 1

General Introduction

1.1. *Candida* and *Candida* infection

In recent decades, fungi have emerged as one of the major causes of human disease, especially amongst the immunocompromised population who particularly susceptible to infection (Wisplinghoff *et al.*, 2003a, 2003b, 2004; Pfaller & Diekema, 2007; Warnock, 2007). The majority of these mycotic infections are associated with *Candida*, *Aspergillus*, and *Cryptococcus* species. Data from the National Nosocomial Infection Survey (NNIS) found that *Candida* species are the fourth most common cause of hospital-acquired bloodstream infection in the USA (Wisplinghoff *et al.*, 2003a; Wisplinghoff *et al.*, 2003b; Wisplinghoff *et al.*, 2004), predominantly caused by *Candida albicans* and *Candida glabrata.*

Candida species normally colonise the human epidermis or mucosae as harmless commensal organisms, but are opportunistic pathogens that cause infections when host immunity is compromised. There are many risk factors that predispose individuals to invasive *Candida* infections. The use of broad-spectrum antibiotics inhibits bacterial growth but permits fungal colonisation – the first step of an invasive fungal infection. Disruption of skin barriers with indwelling medical devices (such as catheters and central venous lines), surgical procedures, the use of prosthetic devices (such as dentures), and burns all provide opportunities for *Candida* species to cause infection. Immunosuppressive medical treatments, such as the use of corticosteroids and immunosuppressive drugs in the treatment of autoimmune diseases and organ transplantation, and chemotherapy and radiation therapy in the treatment of malignancies also predispose patients to *Candida* infection. Diabetes and cystic fibrosis patients and HIV-infected individuals are also predisposed. *Candida* species are one of the most frequent causes of fungal diseases in humans, most commonly resulting in superficial infections such as vulvovaginal and oropharyngeal candidiasis, although more invasive forms of infection such as candiduria, peritonitis, and endocarditis are also very common.

Candida albicans is the most frequent cause of candidiasis, and is the most pathogenic. However, candidiasis can also be caused by species such as *Candida parapsilosis, C. glabrata, Candida tropicalis, Candida krusei,* and *Candida dubliniensis* (Powderly, 1992; Hazen, 1995). *Candida* infection by species other than *C. albicans* is being encountered more frequently than in previous decades. Emerging and inherent

resistance to prophylactic and therapeutic antifungal drugs amongst *Candida* species other than *C. albicans* is a major contributing factor. *Candida krusei* is intrinsically resistant to fluconazole. Fluconazole-resistance has also been encountered in *C. albicans, C. dubliniensis*, and *C. glabrata* in clinical settings (Moran *et al.*, 1998; Ghannoum & Rice, 1999). Increases in life expectancy may also play a role in a higher frequency of infection caused by *Candida* species other than *C. albicans*. For example, *C. glabrata* is often associated with infection in the older and more debilitated population (Lockhart *et al.*, 1999; Warnock, 2007). In contrast, *C. parapsilosis* has been noted as a frequent cause of candidaemia in paediatric patients and neonates (Pasqualotto *et al.*, 2005). Nosocomial *Candida* infection is also common. *Candida parapsilosis* has been reported to be often found on the hands of healthcare workers in European and Latin American countries (Pfaller & Diekema, 2004; Colombo *et al.*, 2006), and is also known to adhere well to materials of indwelling medical devices such as intravascular lines (Weems, 1992).

1.2. Emergence of a novel species; *Candida dubliniensis*

Prior to the development of molecular-based techniques, fungal taxonomy relied heavily on phenotypic and morphological characteristics, which proved problematic due to the variability of certain phenotypic characteristics within *Candida* species. A number of studies published in the early 1990's reported the recovery of atypical *C. albicans* isolates from HIV-infected and AIDS patients. The isolates in question produced germ-tubes and chlamydospores, which were previously considered to be morphological characteristics that were unique to *C. albicans*.

A study from Dublin in 1993 (Sullivan *et al.*, 1993) examined eight putative *C. albicans* isolates that were recovered from HIV-infected and AIDS patients. These isolates produced germ-tubes and chlamydospores, but yielded unusual profile codes with the API ID32C and API 20C AUX yeast identification systems, preventing the definitive identification of these isolates. These eight isolates were thus referred to as "atypical isolates". The same study used oligonucleotide fingerprinting, random amplification of polymorphic DNA (RAPD) analysis and DNA fingerprint analysis using the *C. albicans*-specific complex probe 27A to examine the eight atypical isolates, and found that the atypical isolates yielded RAPD fingerprints that were clearly distinct from the RAPD fingerprints of the typical *C. albicans* isolates. Furthermore, the study found that the atypical isolates hybridised weakly to the *C. albicans* complex DNA fingerprinting probe 27A (Sullivan *et al.*, 1993).

Sullivan *et al.* (1993) also included the former reference type strain (NCPF 3108) for the former species *Candida stellatoidea* in this study. *Candida stellatoidea* was previously thought to be a distinct species that was closely related to *C. albicans.* Isolates of *C. stellatoidea* could be divided into two genetically distinct types, I and II, which did not metabolise sucrose in contrast to typical isolates of *C. albicans*. *Candida stellatoidea* type I isolates differed substantially from *C. stellatoidea* type II and typical *C. albicans* isolates on the basis of several genetic characteristics (Kwon-Chung *et al.*, 1989), whereas the genetic differences between isolates of *C. stellatoidea* type II and *C. albicans* were less distinctive. *Candida stellatoidea* type I isolates are now classified as a true subgroup of *C. albicans*, and type II isolates are now classified as a sucrose-negative variant of *C. albicans* (Kwon-Chung *et al.*, 1989). The atypical isolates described by Sullivan *et al.* (1993) were observed to be more similar to type I than type

II isolates of *C. stellatoidea*, and therefore the investigators hypothesised that the atypical isolates might be sucrose-positive variants of *C. stellatoidea* type I (Sullivan *et al.*, 1993).

A research group in Australia (McCullough *et al.*, 1995a, 1995b) examined 18 similarly atypical *C. albicans* isolates that failed to hybridise with the *C. albicans*-specific DNA fingerprinting probe 27A. Using RAPD analysis these atypical isolates also generated different sized amplimers to those of typical *C. albicans* isolates using *C. albicans*-specific primers. Another study (Boerlin *et al.*, 1995) examined 13 unusual *C. albicans* isolates that had been recovered from HIV-positive intravenous drug users in Switzerland. All of these isolates were β-glucosidase negative. The group used multilocus enzyme electrophoresis (MLEE) to examine 16 enzyme loci, and DNA fingerprinting using the *C. albicans*-specific complex fingerprinting probe Ca3, in order to carry out a population-genetics study. The analysis found that the phenotypically atypical isolates could be distinguished from most phenotypically typical *C. albicans* isolates using the MLEE method, and that only faint bands were produced by the atypical isolates following Southern blotting and hybridisation to the *C. albicans* species-specific Ca3 probe.

A further study from Dublin undertook a detailed comparative investigation of the atypical *C. albicans* isolates (Sullivan *et al.*, 1995). Phenotypic characterisation based on growth temperatures, carbohydrate assimilation profiles, germ-tube and chlamydospore production, as well as genotypic characterisation using DNA fingerprinting analysis with the *C. albicans*-specific complex DNA probe 27A, RAPD analysis, and karyotype analysis were undertaken. Many differences were reported between atypical isolates and typical isolates of both *C. albicans* and *C. stellatoidea.* The authors suggested the presence of another taxon composed of the atypical isolates. This was further confirmed by nucleotide sequence analysis of the V3 variable region of the large subunit ribosomal RNA gene from atypical isolates and typical isolates of *C. albicans*, as well as a reference *C. stellatoidea* isolate, NCPF 3108. Phylogenetic analysis classed the atypical isolates as a separate taxon which was named *C. dubliniensis* (Sullivan *et al.*, 1995).

Since its recognition as a species distinct from *C. albicans*, a widespread geographical distribution for *C. dubliniensis* has been reported (Pujol *et al.*, 1997; Schoofs *et al.*, 1997; Sullivan *et al.*, 1997; Odds *et al.*, 1998; Pinjon *et al.*, 1998; Salkin *et al.*, 1998; Jabra-Rizk *et al.*, 1999; Polacheck *et al.*, 2000; Pontón *et al.*, 2000;

Redding *et al.*, 2001; Gee *et al.*, 2002; Blignaut *et al.*, 2003; Fotedar & Al-Hedaithy, 2003; Montour *et al.*, 2003; Ahmad *et al.*, 2004; Bujdáková *et al.*, 2004; Al Mosaid *et al.*, 2005; Melkusová *et al.*, 2005; Brito Gamboa *et al.*, 2006 Asmundsdóttir *et al.*, 2008; Alvarez *et al.*, 2009). Retrospective identification studies have reported that at least 1.2% of isolates recovered in a clinical setting that were previously identified as *C. albicans* were in fact *C. dubliniensis* (Coleman *et al.*, 1997b; Odds *et al.*, 1998; Jabra-Rizk *et al.*, 2000). Recently the recovery of *C. dubliniensis* from non-human sources was reported for the first time (Nunn *et al.*, 2007).

1.2.1. Phenotypic properties of *C. dubliniensis*

The phenotypic similarities, such as germ-tube and chlamydospore production, of both *C. dubliniensis* and *C. albicans* are the most probable reasons for the late identification of *C. dubliniensis* as a separate species. However, since its classification as a distinct species in 1995, a number of phenotype-based methods have been described in order to distinguish *C. dubliniensis* from *C. albicans*, with varying degrees of success.

1.2.1.1. Production of germ-tubes and chlamydospores

The ability to produce germ-tubes in serum was previously considered a phenotypic characteristic of *C. albicans*. However, *C. dubliniensis* also has this ability, although *C. dubliniensis* is unable to produce germ-tubes following incubation in medium containing *N*-acetylglucosamine, whereas *C. albicans* can (Gilfillan *et al.*, 1998).

The production of chlamydospores on Rice Agar Tween medium (RAT agar) or cornmeal agar is another characteristic shared by *C. albicans* and *C. dubliniensis*. Chlamydospores are spore-like structures that are typically produced at the termini of pseudohyphae, *via* a single suspensor cell. On RAT agar the chlamydospores produced by *C. albicans* usually occur singly, with one chlamydospore attached terminally to pseudohypha *via* a suspensor cell, whereas *C. dubliniensis* produces chlamydospores that are frequently arranged in pairs, triples, or in multiples attached to a terminal suspensor cell (Sullivan *et al.*, 1995; Coleman *et al.*, 1997b) on RAT agar. This characteristic does not hold true for all *C. dubliniensis* isolates however, as has been demonstrated by a number of laboratories (Schoofs *et al.*, 1997; Kirkpatrick *et al.*, 1998).

1.2.1.2. Growth at elevated temperatures

Early comparative analyses of *C. albicans* and *C. dubliniensis* revealed that isolates of *C. albicans* exhibit good growth at 42°C, whereas isolates of *C. dubliniensis* do not (Sullivan *et al.*, 1995). However, *C. dubliniensis* isolates that were able to grow at this temperature were later reported. A further study revealed that *C. dubliniensis* was unable to grow at 45°C, whereas 99% of *C. albicans* isolates demonstrated growth at this elevated temperature (Pinjon *et al.*, 1998). No isolate of *C. dubliniensis* to date has been shown to have the ability to grow at this increased temperature. However, some isolates of *C. albicans* cannot grow at this temperature (Kirkpatrick *et al.*, 1998; Pinjon *et al.*, 1998), which suggests that this is not a 100% accurate method of distinguishing between the two species.

1.2.1.3. Colony colour on CHROMagar Candida™ medium

On chromogenic medium such as CHROMagar Candida™ agar, *C. albicans* colonies are typically pale green/blue and moderately sized following incubation at 37°C for 48 h. The colonies formed by *C. dubliniensis* on this agar are typically darker green/blue in comparison to *C. albicans* and are much smaller in size. However this characteristic is not absolute and intra-species variation does occur. The typical colony colour of *C. dubliniensis* on CHROMagar Candida™ medium (CHROMagar, Paris, France) may be lost after subculture or prolonged storage of isolates, and colonies may appear a slightly paler shade of green, making them more difficult to distinguish from those of *C. albicans* (Schoofs *et al.*, 1997).

1.2.1.4. Carbohydrate assimilation profiles

The analysis of carbohydrate assimilation profiles offered by commercial yeast identification kits is one method for distinguishing the closely related species *C. albicans* and *C. dubliniensis* from each other. One such kit is the API ID32C system (bioMérieux, Paris, France) which consists of a strip of 29 different carbon sources. *Candida albicans* and *C. dubliniensis* differ in their abilities to assimilate four of these carbon sources during incubation for 48 h at 37°C. When the kit is used according to the manufacturer's instructions, *C. albicans* assimilates lactate, xylose, trehalose and methyl-α-D-glucopyranoside, none of which are assimilated by *C. dubliniensis* (Pincus *et al.*, 1999).

1.2.2. Genotypic properties of *C. dubliniensis*

The species that subsequently became known as *C. dubliniensis* was originally observed as atypical germ-tube-positive and chlamydospore-positive isolates of *C. albicans* that yielded unusual fingerprint patterns on hybridisation with oligonucleotide probes that hybridised to eukaryotic microsatellite repeat sequences. These fingerprint patterns were distinct from those of other isolates of both *C. albicans* and *C. stellatoidea* (Coleman *et al.*, 1993; Sullivan *et al.*, 1993). Restriction endonuclease-digested DNA from these isolates also showed weak hybridisation to the complex DNA fingerprinting probe 27A, and yielded distinct profiles with RAPD analysis (Coleman *et al.*, 1993; Sullivan *et al.*, 1993). Distinct hybridisation patterns were also observed between atypical and typical *C. albicans* isolates using a different *C. albicans*-specific complex DNA fingerprinting probe Ca3. Typical and atypical isolates could also be separated into two distinct groups according to MLEE (Boerlin *et al.*, 1995; Pujol *et al.*, 1997).

Comparative sequence analysis of the V3 variable region of the large ribosomal subunit genes and the *ACT1* gene encoding actin identified substantial nucleotide sequence divergence between *C. albicans* and *C. dubliniensis* (Sullivan *et al.*, 1995; Donnelly *et al.*, 1999). This sequence divergence enabled the design of species-specific primers that specifically amplified the *ACT1*-associated intron of *C. dubliniensis*, enabling rapid and definitive identification of the species in as little as 4 h using the polymerase chain reaction (PCR).

Since the classification of *C. dubliniensis* as a distinct species, a species-specific complex DNA fingerprinting probe, known as Cd25, has been developed for DNA fingerprint analyses of *C. dubliniensis* (Joly *et al.*, 1999). This probe is based on semi-repetitive genetic elements that are dispersed throughout the *C. dubliniensis* genome. Following restriction endonuclease digestion of *C. dubliniensis* DNA, the probe is hybridised to the resulting digestion bands that display nucleotide sequence identity to the Cd25 probe, resulting in a Cd25-fingerprint pattern that may be compared amongst different isolates in order to assess genetic relatedness, as discussed further in Section 1.4.5.

The Wellcome Trust Sanger Institute Pathogen Genomics group has undertaken a *C. dubliniensis* genome sequencing project which is now complete (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/). The isolate sequenced was

the *C. dubliniensis* type strain CD36, as this is the isolate that has been studied the most intensively to date. Whole genome shotgun sequencing of the type strain has identified 5,758 genes amongst a total of 13 chromosomes. The entire *C. dubliniensis* genome is 14.6 Mb whereas the *C. albicans* genome is 14.3 Mb. However, 98.1% of putative genes in *C. dubliniensis* display nucleotide sequence composition and positional similarity with those of *C. albicans.*

1.2.3. Virulence of *C. dubliniensis*

Despite the close phylogenetic relationship and phenotypic similarities between the two species, *C. albicans* is significantly more pathogenic than *C. dubliniensis. Candida dubliniensis* is rarely recovered from normal healthy individuals (Coleman *et al.*, 1997b; Pontón *et al.*, 2000; Montour *et al.*, 2003), and is more commonly associated with immunocompromised individuals, such as those undergoing organ transplantation, treatment for malignancies, or those who are infected with HIV or have AIDS (Moran *et al.*, 1997; Sullivan *et al.*, 1997; Sebti *et al.*, 2001). A recent study by Magee *et al.* (2008) compared the karyotypes of *C. albicans* and *C. dubliniensis* using several genomic techniques and concluded that, while the organisation of chromosomes was very similar in both species, the genomic organisation and frequency of translocation differed between the two species. Distinct karyotypes were identified for each *C. dubliniensis* isolate examined in the study, in contrast to *C. albicans* for which half of the isolates had a standard karyotype (Magee *et al.*, 2008). The authors suggested that the higher frequency of genomic reorganisation in *C. dubliniensis* may contribute to a lower genomic stability than that of *C. albicans,* possibly contributing to the lesser virulence of the former species (Magee *et al.*, 2008).

A number of studies have undertaken comparative analyses of the virulence factors of both *C. albicans* and *C. dubliniensis*, examining the kinetics of hypha production (Gilfillan *et al.*, 1998; Stokes *et al.*, 2007), phenotypic switching (Hannula *et al.*, 2000), as well as comparing the genes that encode virulence factors such as phospholipases, proteinases and adhesins in both species (McCullough *et al.*, 1995b; Moran *et al.*, 2004; Hoyer *et al.*, 2001).

One of the most important virulence factors for any microbial species is its ability to adhere to and colonise the mucosae of the host. A number of research groups have carried out studies comparing the adherence of *C. albicans* and *C. dubliniensis* to human buccal epithelial cells (BECs) with inconsistent results. Some studies have

claimed that *C. dubliniensis* is more adherent (Gilfillan *et al.*, 1998; McCullough *et al.*, 1995b; Jabra-Rizk *et al.*, 2001), while other studies have claimed that the opposite is true (Vidotto *et al.*, 2003). *Candida dubliniensis* has been shown to be more adherent than *C. albicans* in the presence of the antifungal agent fluconazole, which may explain the increased prevalence of *C. dubliniensis* in patients who receive prophylactic treatment with this antifungal drug (Borg-von Zepelin *et al.*, 2002). Homologues of the *C. albicans* agglutinin-like sequence (ALS) proteins have been identified in *C. dubliniensis*. However, the nucleotide sequences of this gene family display substantial divergence in both species (Hoyer *et al.*, 2001).

A number of previous studies have revealed that *C. dubliniensis* isolates have lower filamentation rates in comparison to *C. albicans* isolates *in vitro*, in reconstituted human epithelial (RHE) models, and in murine infection models (Gilfillan *et al.*, 1998; Stokes *et al.*, 2007), while *C. albicans* switches to its filamentous form and increases production of proteinases and adhesins, *C. dubliniensis* remains in the yeast form (Gilfillan *et al.*, 1998; Vilela *et al.*, 2002; Stokes *et al.*, 2007). Secreted aspartyl proteinases (Saps) are believed to be another important virulence factor for *Candida* species, and homologues of seven of the ten *C. albicans SAP* genes have been identified in *C. dubliniensis* by Southern hybridisation (Gilfillan *et al.*, 1998). Comparative genomic hybridisation analysis identified the absence of the *SAP5* gene in *C. dubliniensis* and identified only one gene possessing nucleotide sequence similarity to the *SAP4, 5* and *6* genes. This absence has been confirmed by the annotation of the *C. dubliniensis* genome which is now complete. The absence of these *SAP* genes may possibly account for the lower virulence of the species in comparison to *C. albicans* (Moran *et al.*, 2004).

Many previous comparative studies have indicated that *C. dubliniensis* is less tolerant of environmental stresses such as elevated temperatures, high salt concentrations and oxidative stress than *C. albicans* (Pinjon *et al.*, 1998; Alves *et al.*, 2002; Vilela *et al.*, 2002). Recent comparative genomic studies have highlighted that 4.4% of *C. albicans* gene sequences (representing 247 genes) are significantly divergent or absent in *C. dubliniensis* (Moran *et al.*, 2004). More recent investigation using genome-wide gene expression profiling has revealed that *C. albicans* and *C. dubliniensis* display similar core transcriptional responses to environmental stresses. However, differences in the expression of the *ENA21/ENA22* genes were noted between the two species in response to osmotic stress (Enjalbert *et al.*, 2009). These genes were

strongly induced in *C. albicans* but were not in *C. dubliniensis*. The authors suggested that insufficient expression of the Ena21/22 sodium ion efflux pump in *C. dubliniensis* may contribute to the increased salt sensitivity of the species (Enjalbert *et al.*, 2009).

1.3. Identification of *C. dubliniensis* from clinical specimens

In order for the true prevalence of *C. dubliniensis* to be determined, the development of accurate and reliable identification methods is required, with particular regard to the differentiation of *C. dubliniensis* from the closely related *C. albicans*. Previously culture- and phenotype-based methods have been used for this purpose. However, intra-species variation and unreliability have hampered such prevalence studies, most likely underestimating the true prevalence of *C. dubliniensis*. However, over the last decade, a number of PCR-based methods as well as a highly accurate culture-based method have been developed which have improved the ease with which the two species may be differentiated, thus enabling more accurate *C. dubliniensis* prevalence studies to be carried out.

1.3.1. Phenotype-based methods for clinical identification of *C. dubliniensis* in clinical specimens

1.3.1.1. Production of germ-tubes and chlamydospores

Prior to the first description of *C. dubliniensis* as a distinct species in 1995, the formation of germ-tubes and chlamydospores was used in the definitive identification of *C. albicans*. However, *C. dubliniensis* also produces these structures. A number of differential agars have been developed in recent years in an attempt to distinguish *C. dubliniensis* from *C. albicans*, and are based on the production of a hyphal fringe by colonies of *C. dubliniensis*, and the absence of such structures surrounding colonies of *C. albicans* following specific periods of incubation.

Staib agar (Staib & Morschhäuser, 1999) is based on an aqueous extract of *Guizotia abyssinica* seed (nigel seed or blackseed) and is commonly known as "birdseed agar". *Candida dubliniensis* was observed to produce a hyphal fringe composed of hyphae, pseudohyphae and chlamydospores on this agar, in contrast to *C. albicans* which grew as smooth shiny colonies composed of yeast cells, following 48 h of incubation at 30°C. This finding was supported by another study (Al Mosaid *et al.*, 2001) which demonstrated that 97.7% of 130 tested *C. dubliniensis* isolates formed

colonies with surrounding hyphal fringes on this agar, and that all of 166 tested *C. albicans* isolates grew as smooth shiny colonies composed of yeast cells (Fig. 1.1A).

Pal's agar was also examined for the differentiation of the two species (Al Mosaid *et al.*, 2003). This medium contains an aqueous extract of *Helianthus annus* (sunflower) seeds (including kernel and shells) and differentiates the two species on the basis of the production of a hyphal fringe by *C. dubliniensis* after 48 h of incubation at 30°C. This hyphal fringe consists of abundant hyphae, pseudohyphae, and chlamydospores. All 128 tested *C. dubliniensis* isolates produced a hyphal fringe on this agar, in contrast to all 124 tested *C. albicans* isolates, which grew as smooth colonies composed of yeast cells (Fig. 1.1B).

Mosca *et al.* (2003) demonstrated that the milk-based Casein agar can be used to distinguish *C. albicans* from *C. dubliniensis* isolates according to the production of chlamydospores following 48 h incubation at 24°C. All *C. dubliniensis* isolates produced chlamydospores on this medium, in contrast to *C. albicans* of which only 9/120 isolates produced chlamydospores.

Khan *et al.* (2004) investigated the efficacy of tobacco (Nicotiana) agar for the differentiation of *C. dubliniensis* from *C. albicans*. This agar was based on an aqueous extract of tobacco from commercially available cigarette brands and offered high levels of discrimination between the two closely related species with 100% accuracy (Khan *et al.*, 2004). Using this medium, isolates of *C. albicans* formed smooth white-creamy colonies that lacked a hyphal fringe or chlamydospores, in contrast to isolates *C. dubliniensis*, which formed rough yellow-brown colonies with surrounding hyphal fringes consisting of hyphae, pseudohyphae and chlamydospores following 48–72 h incubation at 30°C (Fig. 1.1C) (Khan *et al.*, 2004).

Recently, two studies used CHROMagar Candida™ medium supplemented with Pal's agar to discriminate between different *Candida* species on the basis of colony colour, as well as between isolates of *C. dubliniensis* and *C. albicans* on the basis of hyphal-fringe production by the former species (Sahand *et al.*, 2005; Raut & Varaiya, 2009). Both studies demonstrated that the majority of tested *C. dubliniensis* isolates grew as rough blue/green colonies with surrounding hyphal fringes, in contrast to the *C. albicans* isolates, all of which were light green and smooth, and did not produce hyphal fringes.

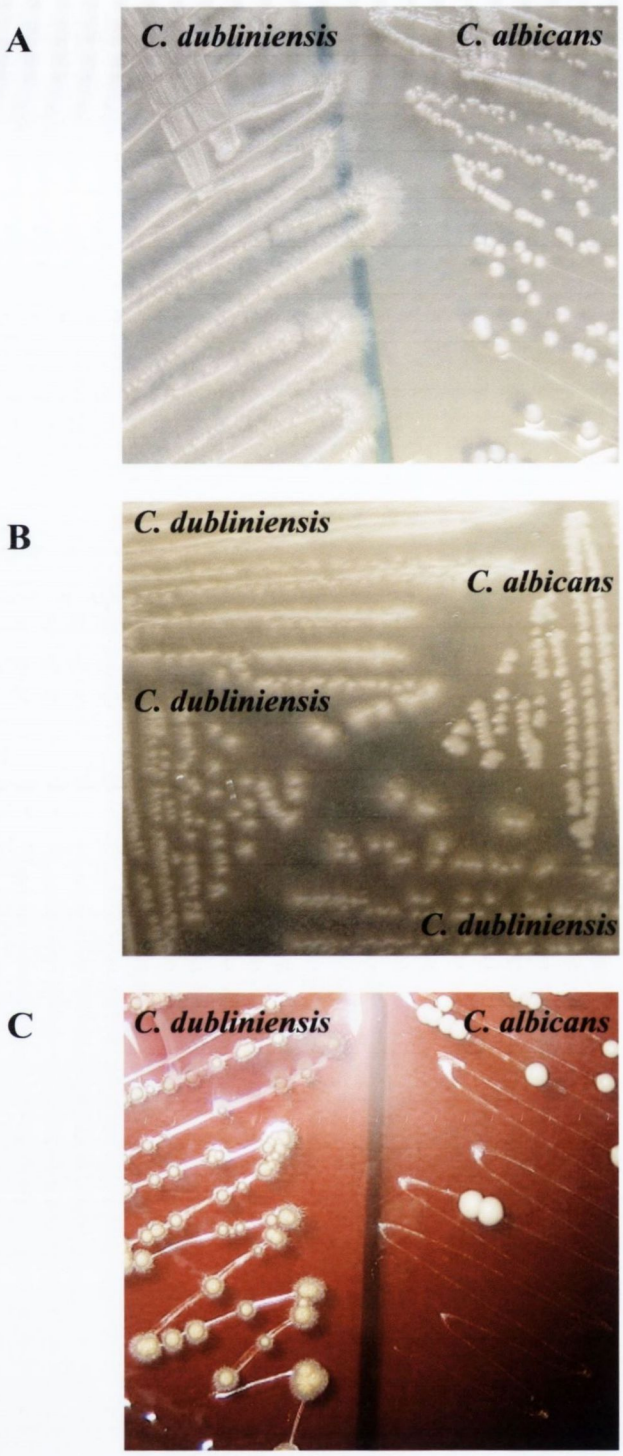


Figure 1.1. Differentiation of *C. dubliniensis* and *C. albicans* on Staib, Pal's and tobacco agar.

Colonies formed by *C. dubliniensis* are surrounded by a hyphal fringe composed of hyphae, pseudohyphae and chlamydospores on Staib agar (panel A) Pal's agar (panel B) and tobacco agar (panel C), in contrast to *C. albicans* colonies which are smooth and shiny on each agar following 48 h incubation at 30°C.

1.3.1.2. Growth at elevated temperatures

The inability of *C. dubliniensis* to grow at 42°C was originally thought to be a useful method of differentiating between isolates of *C. dubliniensis* and *C. albicans*, as isolates of *C. albicans* grow well at this temperature in contrast to *C. dubliniensis* (Sullivan *et al.*, 1995). However, isolates of *C. dubliniensis* that were able to grow poorly or well at this temperature were later reported (Coleman *et al.*, 1997a, 1997b; Sullivan *et al.*, 1997; Pinjon *et al.*, 1998; Sullivan & Coleman, 1998), and further studies suggested that inability or ability of isolates to grow at 45°C offered a higher level of discrimination (Pinjon *et al.*, 1998). No isolate of *C. dubliniensis* has demonstrated an ability to grow at this increased temperature to date, whereas the majority of *C. albicans* isolates are able to grow at this temperature. However, some isolates of *C. albicans* have also demonstrated an inability to grow at 45°C, (Kirkpatrick *et al.*, 1998; Pinjon *et al.*, 1998), which suggests that although this method is a simple and inexpensive method of differentiation, it is not 100% reliable.

1.3.1.3. Colony colour on CHROMagar Candida™ medium

The most commonly used medium for the presumptive identification of medically important *Candida* species is CHROMagar Candida™ medium (Odds & Bernaerts, 1994). Different *Candida* species can be putatively identified according to colony colours formed on this medium that are typical of individual *Candida* species. *Candida albicans* forms large green to blue/green colonies. *Candida tropicalis* grows as colonies that are blue/gray in the centre and surrounded by a dark brown/purple halo in the agar surrounding the colony. *Candida krusei* grows as pale, flat, papillate colonies with spreading edges. *Candida glabrata* grows as dark pink colonies that are paler in colour around the edges of the colony, while *C. parapsilosis* grows as white-pale pink colonies (Odds & Bernaerts, 1994). *Candida dubliniensis* grows as green/blue colonies that are typically darker in colour and smaller than those formed by *C. albicans.* However, paler green colonies can also be formed by *C. dubliniensis* on this agar. Although CHROMagar Candida™ medium is highly useful for the presumptive identification of a several medically important *Candida* species, it is not reliable enough for the definitive differentiation of *C. albicans* and *C. dubliniensis*.

1.3.1.4. Carbohydrate assimilation profiles

A comprehensive study was undertaken by Pincus *et al.* (1999) in order to determine the usefulness of commercial yeast identification systems for the identification of *C. dubliniensis*. These identification systems rely on the differential assimilation of a range of substrates by different species. Individual species have distinct assimilation patterns that give rise to a numerical code or profile, from which a species is identified from a corresponding database. Pincus *et al.* (1999) examined the API 20C AUX, API ID32C, VITEK YBC, VITEK 2 ID-YST (bioMérieux) and the RapID Yeast Plus (Remel Inc., Lenexa, KS, USA) yeast identification systems. The percentage of isolates capable of assimilating each substrate was compared for both *C. albicans* and *C. dubliniensis*. Any substrate that showed > 50% difference in reactivity was considered useful in species differentiation. The study found that following 48 h incubation according to the manufacturer's instructions, the assimilation of four substrates in the API 20C AUX system [glycerol (GLY), D-xylose (XYL), methyl-α-D-glucopyranoside (MDG) and D-trehalose (TRE)] could distinguish the two species, as could four substrates [XYL, MDG, TRE, and lactate (LAT)] in the API ID32C system, as well as four substrates [TRE, MDG, LAT, and 4-methylumbelliferyl phosphate (MUP)] in the VITEK 2 ID system. Only two substrates [MUP and α-D-glucopyranoside] enabled the differentiation of the two species using the RapID Yeast Plus system. The VITEK YBC kit showed differential assimilation of three substrates [TRE, XYL and GLY] after 24 h incubation, and of only one substrate [XYL] after 48 h (Pincus *et al.*, 1999). This study reported that of all the systems analysed, the API ID32C system yielded the most consistent results with *C. dubliniensis*. However, only four different profile numbers were obtained with the 80 isolates tested. Prior to the identification of *C. dubliniensis* as a separate species, the API ID32C codes obtained for such isolates were not present in the API ID32C database, resulting in low identification scores for *Candida sake* or *C. stellatoidea* or no identification (Sullivan *et al.*, 1993; Boerlin *et al.*, 1995; Sullivan *et al.*, 1995, 1997; Coleman *et al.*, 1997a; Pujol *et al.*, 1997). However, the database has since been updated so that several *C. dubliniensis*-specific API ID32C codes are now included.

1.3.2. Identification of *C. dubliniensis* using PCR-based methods

Due to the many phenotypic characteristics shared between *C. albicans* and *C. dubliniensis*, definitive identification of each species is difficult based on phenotype-

based methods alone. Although Pal's agar and tobacco agar offer high levels of discrimination (Al Mosaid *et al.*, 2003; Khan *et al.*, 2004), the most reliable and definitive methods of identification are molecular based methods such as PCR. The genome sequences of *C. albicans* and *C. dubliniensis* display approximately 90% nucleotide sequence identity (Moran *et al.*, 2004), presenting sufficient nucleotide sequence divergence to enable the use of PCR-based techniques, which are easy to perform, rapid, specific, and are highly reproducible.

Elie *et al.* (1998) developed a method enabling the differentiation of five clinically relevant *Candida* species using PCR amplification of the internal transcribed spacer 2 (ITS2) region of the rRNA locus, followed by enzymatic immunoassay. The method enabled identification of *C. dubliniensis* from template DNA in 7 h, thus providing a more rapid means of identification than phenotype-based methods.

Park *et al.* (2000) also used the ITS2 region of reference *Candida* strains to develop molecular beacon probes enabling a rapid and definitive means of *C. dubliniensis* identification. These molecular beacon probes are small single-stranded, nucleic-acid-hairpin probes with a high specificity that fluoresce highly on binding target DNA, due to the presence of fluorophores and quenchers that are covalently linked to each end of the probe. The study identified *C. dubliniensis* isolates with 100% accuracy and reduced identification times to 6 h approximately.

Examination of the *ACT1* exon and intron sequences in ten epidemiologically unrelated *C. dubliniensis* isolates as well as a number of other yeast species revealed sufficient nucleotide sequence divergence of the *ACT1* intron between different species, and little intraspecies variation in this region amongst the *C. dubliniensis* isolates tested (Donnelly *et al.*, 1999). Primers were designed that amplified a 288-bp product from the *ACT1* intron of *C. dubliniensis* template DNA specifically, offering a rapid and definitive method of discrimination between *C. albicans* and *C. dubliniensis* that could be achieved in as little as 4 h (Donnelly *et al.*, 1999).

1.4. Molecular typing of *Candida* species

Molecular typing systems have proved very useful in the epidemiological and population structure analyses of microbial pathogens. The study of microbial population structures facilitates the understanding of the dynamics of infectious organisms in human populations, the complex relationships between commensal and infectious

organisms, the origins of infection, the emergence of drug resistance in populations, and the genetic relatedness of isolates in the same species (Soll, 2000). To provide such information about pathogenic microorganisms, molecular typing systems should be:

- Effective at discriminating between isolates of the same species that are highly related but are non-identical.
- Able to recognise the same strain among different isolates and to generate reproducible data.
- Resistant to high-frequency genome reorganisation and evolutionary pressure so that genetic differences are relatively stable over time and mutate with a medium frequency, thus reflecting evolutionary change only.
- Able to determine the genetic distance between isolates that are more closely related and those that are less so.
- Amenable to computer-based analysis to enable data normalisation, analysis and storage.

1.4.1. Multilocus enzyme electrophoresis (MLEE)

This has been used as a powerful method for assessing genetic diversity, population structures and epidemiological analyses of microorganisms, and has been used for such studies of several yeast species including *C. albicans* (Caugant & Sandven, 1993; Pujol *et al.*, 1993; Boerlin *et al.*, 1995). This technique is based on single nucleotide polymorphisms (SNPs) that give rise to amino acid substitutions in approximately ten house-keeping enzymes. Amino acid changes lead to alterations in the charge and mobility of the encoded protein, which is detected following starch-gel electrophoresis or polyacrylamide-gel electrophoresis of cell extracts and visualised with enzyme-specific stains. Each isolate generates a staining profile that is used in computer based analysis and population structure studies. This method is time consuming, requiring the analysis of approximately ten enzymatic loci. Furthermore, MLEE is based on the indirect analysis of genotypes. It will not detect SNPs that do not result in amino acid substitutions or changes that do not affect the charge or mobility of the encoded protein (Caugant & Sandven, 1993; Soll, 2000; Gil-Lamaignere *et al.*, 2003).

1.4.2. Restriction fragment length polymorphism (RFLP) analysis

This technique was one of the first DNA fingerprinting methods to prove useful in assessment of genetic relatedness amongst fungal species. It relies on the variable patterns generated by restriction endonuclease digestion of chromosomal DNA following agarose gel electrophoresis. These patterns can vary amongst isolates due to SNPs in restriction sites, acquisition/deletion of restriction endonuclease sites, or insertions/deletions of DNA sequence between restriction endonuclease sites. These fragments can be visualised directly by staining with ethidium bromide. This method is rapid, simple and inexpensive. Alternatively, the digested chromosomal DNA can be transferred to a membrane by capillary blotting and probed with species-specific DNA sequence probes by Southern hybridisation (discussed further in Section 1.4.5).

The method has been applied to many *Candida* species such as *C. albicans* (Bart-Delabesse *et al.*, 1993), *C. parapsilosis* (Branchini *et al.*, 1994), *C. tropicalis* (Doebbeling *et al.*, 1991, 1993) and *C. rugosa* (Dib *et al.*, 1996). Sullivan *et al.* (1995) used RFLP to examine separate *Eco*RI and *Hin*fI restriction endonuclease-digested genomic DNA extracts in order to examine the genetic relatedness of typical and atypical isolates of *C. albicans.* The analyses revealed significant genetic differences between the typical and atypical isolates of *C. albicans*, the latter being subsequently reclassified as *C. dubliniensis* (Sullivan *et al.*, 1995).

1.4.3. Electrophoretic karyotyping

Electrophoretic karyotyping has proved useful for the population structure analysis of many *Candida* species including *C. dubliniensis* (Asakura *et al.*, 1991; Doi *et al.*, 1994; Soll, 2000; Shin *et al.*, 2001; Gee *et al.*, 2002). The technique involves the treatment of fungal cells with enzymes, proteases and detergents to degrade cell walls, membranes and proteins and the separation of the remaining chromosome-sized DNA molecules according to size by agarose gel electrophoresis using pulsed-field gel electrophoresis (PFGE) which uses electric fields of alternate orientation. These chromosome-sized molecules can display variation in size amongst different isolates due to unequal chromosomal rearrangements by translocation, transposable elements or recombination events. However, the pattern variations between moderately related isolates can be difficult to interpret, and the method is time consuming and requires expensive equipment.

1.4.4. Random amplified polymorphic DNA (RAPD) analysis

This technique has been widely used for population analyses of *Candida* species (Lehmann *et al.*, 1992; Bostock *et al.*, 1993; Coleman *et al.*, 1997a; Soll, 2000; Gil-Lamaignere *et al.*, 2003). Random amplified polymorphic DNA (RAPD) analysis, also called arbitrarily primed PCR (AP-PCR), uses random single primers that are approximately 10 bp in length in separate PCR amplifications with low annealing temperatures. Whilst one such primer can amplify a complex pattern that shows good variability amongst isolates of the same species, most RAPD primers result in 1–3 intense bands that may or may not differ amongst isolates. To overcome this, most RAPD systems use more than one primer, amplifying the DNA of each isolate with each primer separately, and combining the resulting data. However, RAPD reproducibility is poor due to the many factors that can affect PCR efficiency.

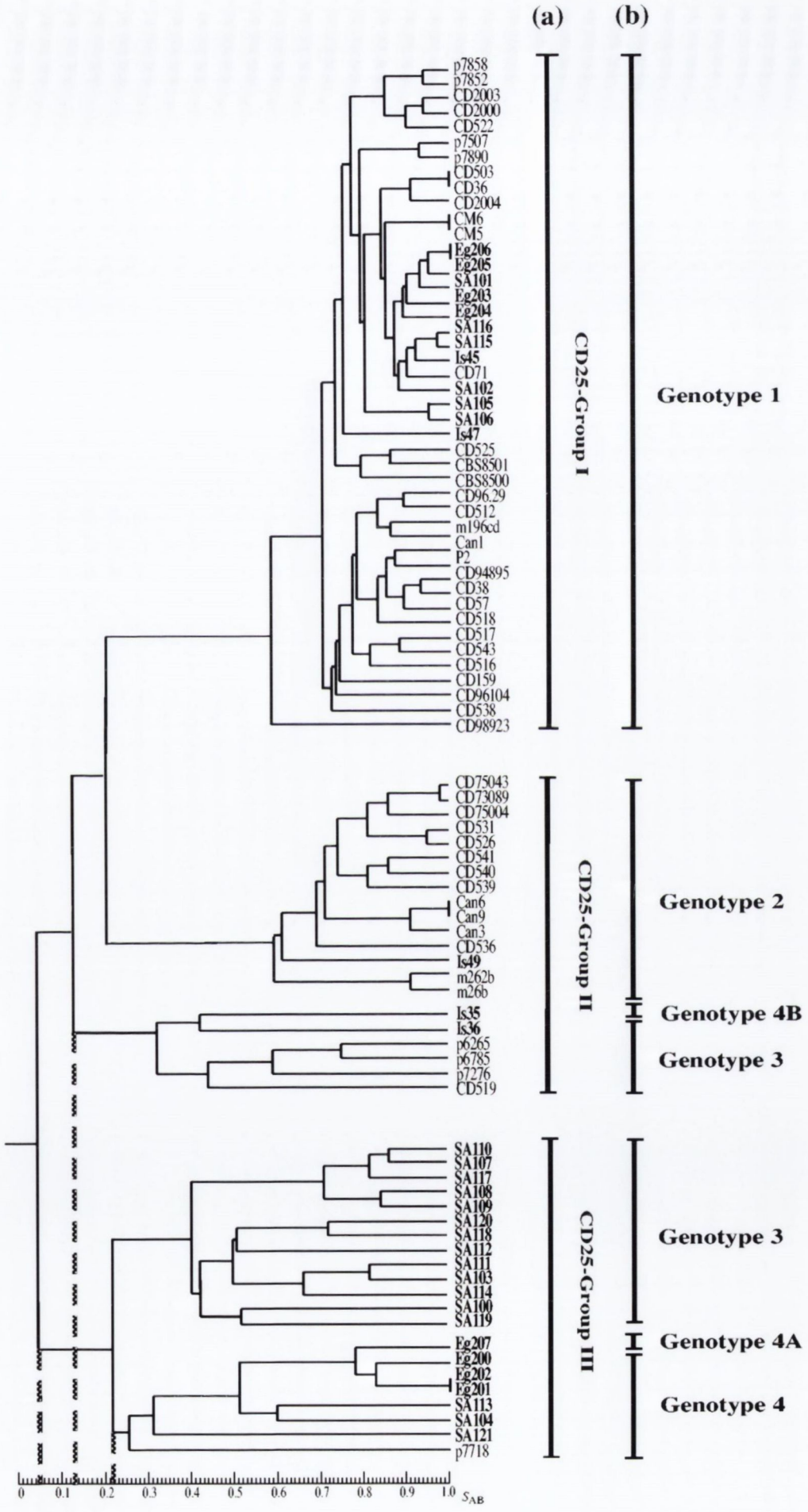
1.4.5. Complex DNA fingerprinting probes

This method is an adaptation of RFLP that incorporates Southern hybridisation of restriction endonuclease-digested chromosomal DNA to a complex species-specific DNA fingerprinting probe that identifies repetitive DNA sequences which are dispersed throughout the genome. Complex probes have been developed for fingerprinting and population structure analysis of *C. albicans* (Scherer & Stevens, 1988; Pujol *et al.*, 2002), *C. glabrata* (Lockhart *et al.*, 1997), *C. dubliniensis* (Joly *et al.*, 1999) and *C. tropicalis* (Joly *et al.*, 1996) with great success due to high levels of discrimination between groups of closely related isolates. In brief, species-specific complex probes that share nucleotide sequence identity to repetitive DNA sequences dispersed across the genome are applied to restriction endonuclease-digested DNA by Southern hybridisation. Hybridisation results in band patterns which then undergo computer-assisted analysis so that average similarity coefficient (S_{AB}) values are calculated for every possible pair of isolates included in the study.

Such studies on *C. dubliniensis* using the complex DNA fingerprinting probe Cd25 have identified the presence of three major clades, termed Cd25 groups I–III (Fig. 1.2). Isolates belonging to Cd25 group I are all closely related despite being recovered in many countries around the world, although mainly from HIV-infected individuals (Joly *et al.*, 1999; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005). Sequence analysis of the ITS region of the rRNA gene cluster revealed that Cd25 group I isolates consist of a single ITS genotype, designated ITS genotype 1 (Gee *et al.*, 2002). In contrast, Cd25

Figure 1.2. The population structure of *C. dubliniensis* determined by DNA fingerprinting analysis with the species-specific complex probe Cd25.

Dendrogram generated from average S_{AB} values computed for every possible pairwise combination of independent *C. dubliniensis* isolates recovered from a broad range of geographical locations, and including all currently known ITS genotypes, fingerprinted with the Cd25 probe. Three distinct clades are visible, termed Cd25 group I–III, where group I consists of isolates of ITS genotype 1 exclusively, group II contains all ITS genotype 2 isolates, along with ITS genotype 3 isolates and a single ITS genotype 4 isolate. Cd25 group III consists of ITS genotype 3 and 4 isolates from the Middle East exclusively, and contains isolates that exhibit high levels of intrinsic resistance to 5-fluorocytosine. Figure taken from Al Mosaid *et al.*, 2005.


(a)
(b)
p7858
p7852
CD2003
CD2000
CD522
p7507
p7890
CD503
CD36
CD2004
CM6
CM5
Eg206
Eg205
SA101
Eg203
Eg204
SA116
SA115
Is45
CD71
SA102
SA105
SA106
Is47
CD525
CBS8501
CBS8500
CD96.29
CD512
m196cd
Can1
P2
CD94895
CD38
CD57
CD518
CD517
CD543
CD516
CD159
CD96104
CD538
CD98923
CD25-Group I
Genotype 1
CD75043
CD73089
CD75004
CD531
CD526
CD541
CD540
CD539
Can6
Can9
Can3
CD536
Is49
m262b
m26b
Is35
Is36
p6265
p6785
p7276
CD519
CD25-Group II
Genotype 2
Genotype 4B
Genotype 3
SA110
SA107
SA117
SA108
SA109
SA120
SA118
SA112
SA111
SA103
SA114
SA100
SA119
Eg207
Eg200
Eg202
Eg201
SA113
SA104
SA121
p7718
CD25-Group III
Genotype 3
Genotype 4A
Genotype 4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
S_{AB}

group II isolates are more diverse and consist of three separate ITS genotypes (ITS genotypes 2–4), which corresponded to distinct subclades within the Cd25 group II fingerprinting clade (Gee *et al.*, 2002; Al Mosaid *et al.*, 2005) (Fig. 1.2). More recently, a third major clade, termed Cd25 group III, was identified among isolates from Egypt, Israel and Saudi Arabia (Al Mosaid *et al.*, 2005). The DNA fingerprints of Cd25 group III isolates are very distinctive relative to isolates from Cd25 groups I and II (Fig. 1.3). Nucleotide sequence analyses of the ITS2 region revealed that Cd25 group III isolates belong to ITS genotypes 3 and 4 (Fig. 1.2) (Al Mosaid *et al.*, 2005). All Cd25 group III isolates examined to date exhibit high-level intrinsic resistance to 5-fluorocytosine (5FC) (Al Mosaid *et al.*, 2005).

1.4.6. Multilocus sequence typing

Multilocus sequence typing (MLST) involves the PCR amplification and DNA sequence analysis of 6–8 housekeeping genes that are under stabilising selection pressure. For each housekeeping gene locus, sequence variations are identified as separate alleles, and each allele is assigned an integer. For each isolate analysed, the integers for each housekeeping locus are then combined to generate an allelic profile. Each distinct allelic profile is then assigned another integer, termed a sequence type (ST).

Although MLST was originally designed for epidemiological and population structure analysis of bacterial species, it has been applied to many diploid fungal species over the last decade, and has become a very popular method for the molecular typing of many *Candida* species (Bougnoux *et al.*, 2002; Dodgson *et al.*, 2003; Robles *et al.*, 2004; Tavanti *et al.*, 2005a, 2005b). The diploid nature of most *Candida* species increases the level of sequence variation due to the presence of heterozygous nucleotide sites, which provide additional genotypes. For most diploid organisms, the ST is known as a diploid sequence type (DST).

As it is based on direct DNA sequence analysis, MLST is highly reproducible, relatively inexpensive, rapid, and has a good discriminatory power. Data are comparable amongst different research groups, enabling collaborative studies *via* online databases. Computer-based analysis of DNA sequence or allelic profile data can be used to generate phylogenetic trees that display the genetic relatedness of the isolates being investigated and can suggest STs or DSTs that may have been founding genotypes for each distinct clade.

1.4.6.1 Comparison of DNA fingerprinting and MLST methods in C. albicans

Extensive fingerprinting analysis has been used to investigate the population biology of *C. albicans* using the Ca3 complex fingerprinting probe, resulting in the division of the species into five distinct clades (Boerlin *et al.*, 1995; Blignaut *et al.*, 2002; Pujol *et al.*, 2002; Soll & Pujol, 2003). Isolates belonging to clades I, II, and III tend to be more geographically dispersed, whilst isolates belonging to the South African clades and European clades (see below) tend to be more geographically localised (Blignaut *et al.*, 2002; Pujol *et al.*, 2002; Soll & Pujol, 2003).

Blignaut *et al.* (2002) noted that 53% of *C. albicans* from black HIV-infected individuals, and 33% of isolates from healthy white individuals from South Africa separated into a previously unrecognised clade, termed the South African or SA clade. Isolates belonging to clades I, II and III were also observed in this location. However, only 2% of North American isolates belonged to the SA clade (Blignaut *et al.*, 2002; Pujol *et al.*, 2002) showing a geographical specificity of the SA clade. Pujol *et al.* (2002) discovered a fifth clade that was most frequently isolated in Europe. Of isolates studied from this location, 26% were found to belong to this newly described clade termed the European or E clade, whilst only 2% of North American isolates, 5% of South American isolates, and 1% of South African isolates were seen to cluster with the European isolates (Pujol *et al.*, 2002). The study also noted the absence of clade II isolates in South America.

A particularly interesting clade-specific feature of *C. albicans* was subsequently reported, when it was observed that isolates that exhibited resistance to 5FC were predominantly from clade I (Pujol *et al.*, 2004), supporting an earlier study that indicated a clonal population structure in the species and the rarity of recombination events between clades (Graser *et al.*, 1996; Cowen *et al.*, 1999; Pujol *et al.*, 2003).

Early MLST analysis of *C. albicans* identified the presence of four main clades of related isolates in the population structure, as well as at least eight minor clades. Similar to the results observed with the Ca3 fingerprinting probe, a predominant association of 5FC-resistant isolates with a specific clade was noted. Further population analysis using 1,391 *C. albicans* isolates identified the presence of 17 clades, of which MLST clades 1–4 showed good correspondence with previously identified Ca3 clades I, II, III and SA respectively, thus identifying geographical enrichment of clades (Odds *et al.*, 2007). Isolates displaying reduced 5FC susceptibility clustered in MLST clade 1 (Tavanti *et al.*, 2005a; Odds *et al.*, 2007; Odds & Jacobsen, 2008).

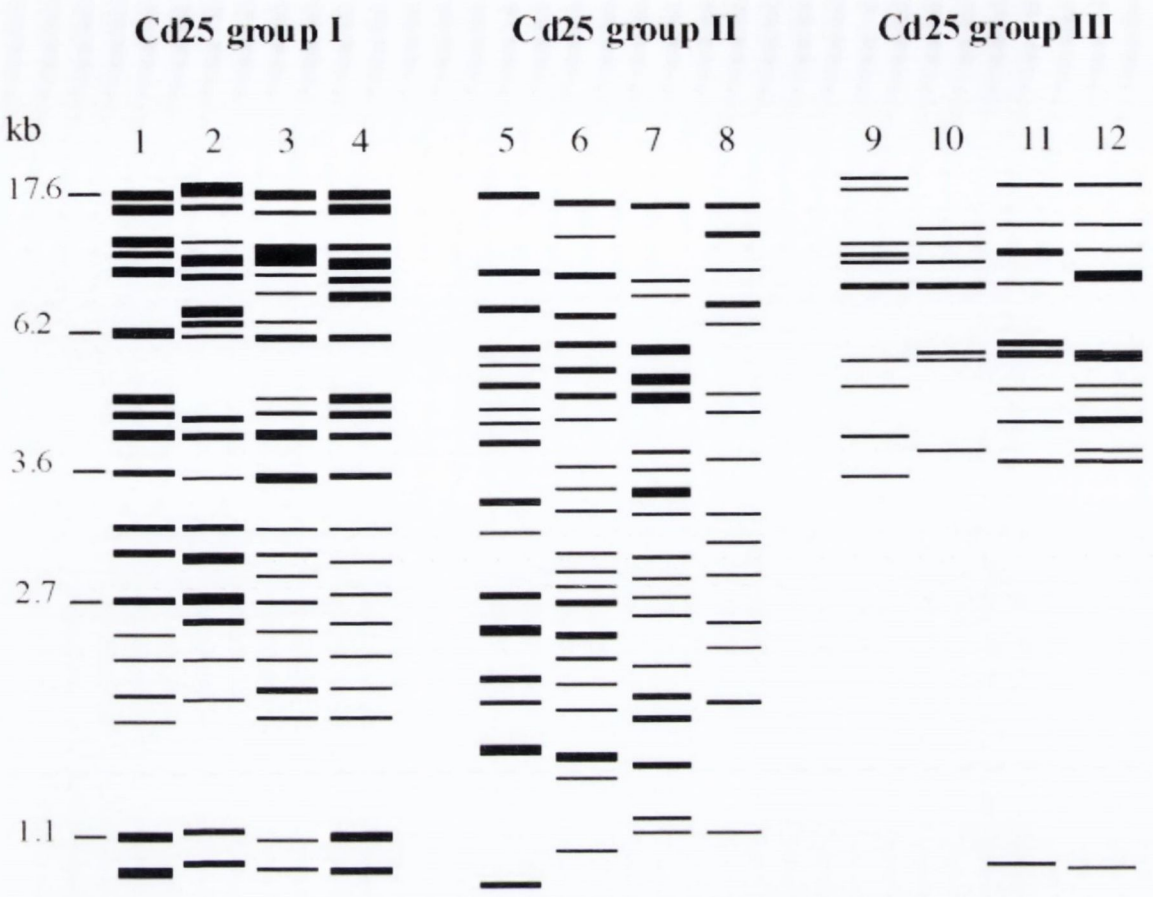


Figure 1.3. Fingerprint patterns obtained by hydridisation of the complex DNA fingerprinting probe Cd25 to the DNA of isolates belonging to Cd25 groups I, II and III.

Model generated using the DENDRON software package showing band positions and intensities of Cd25-generated hybridisation fingerprint patterns of restriction endonuclease-digested genomic DNA of *C. dubliniensis* isolates belonging to the major clades Cd25 group I, Cd25 group II, and the novel 5FC-resistant Saudi Arabian and Egyptian clade Cd25 group III identified in this study. Molecular sizes in kilobases are shown on the left. The *C. dubliniensis* isolates from which the corresponding patterns in the lanes were obtained are as follows (genotypes are shown in parentheses): lane 1, CM6 (1); lane 2, CD518 (1); lane 3, SA102 (1); lane 4, Eg204 (1); lane 5, Can9 (2); lane 6, CD514 (2); lane 7, Is49 (2); lane 8, p6265 (3); lane 9, SA103 (3); lane 10, SA121 (4); lane 11, SA119 (3); lane 12, Eg201 (4). Figure taken from Al Mosaid *et al.*, 2005.

The ability of the MLST, RAPD, MLEE and DNA fingerprinting methods to discriminate between isolates of *C. albicans* has been investigated (Robles *et al.*, 2004). MLST was demonstrated as a highly effective technique that had a comparable discriminatory power to the MLEE and Ca3 fingerprinting systems, but was more discriminatory than RAPD (Robles *et al.*, 2004). While DNA fingerprinting using complex DNA probes has a very high discriminatory power, it is time consuming, reproducibility is poor, and results cannot be compared between different laboratories, which makes MLST a more favourable alternative.

1.4.7. Microarray-based typing

The latest technological advancement combines MLST with DNA microarray technology. Lott & Scarborough (2008) have designed such an array for high-throughput SNP detection in *C. albicans*. This array contains SNPs that have been previously identified by MLST, as well as other SNPs on chromosomes that are not included in the *C. albicans* MLST scheme (Lott & Scarborough, 2008). This technique has not as yet been applied to large numbers of *C. albicans* isolates, and it remains to be seen if it can be used as a reproducible, cost-effective and high-throughput platform for population structure analysis and epidemiological analyses.

1.5. Antifungal agents in clinical use

Most antifungal drugs developed to date target the biosynthetic pathways of fungal nucleic acids or cellular structural components. There are a limited number of antifungals in use to date, and resistance has been reported to most. Resistance is typically defined and quantified according to a minimum inhibitory concentration (MIC), where a drug is titrated according to a standardised protocol and incubated with a test isolate for a defined period of time. The lowest concentration of the drug to inhibit fungal growth is defined as the MIC value.

The allylamines, the morpholines, and the azole-based classes of antifungals all target the biosynthetic pathway of ergosterol, which is a vital component of fungal cell membranes. The polyenes target ergosterol in the cell membrane directly. The echinocandins target the biosynthesis of β-linked glucans, which are a major component of fungal cell walls, and 5FC targets the pyrimidine salvage pathway leading to a disruption in the biosynthesis of nucleic acids and cellular proteins (Groll *et al.*, 1998; Ghannoum & Rice, 1999; Yang & Lo, 2001; Sanglard, 2002).

1.5.1. Ergosterol-targeting antifungal agents

1.5.1.1. Allylamines and thiocarbamates

The allylamines and the thiocarbamates are classes of antifungal agents that target a squalene epoxidase enzyme that acts early on in the biosynthesis of ergosterol, which is encoded by the *ERG1* gene (Fig. 1.4). The allylamines, such as terbinafine and naftifine, act by inhibiting the conversion of the sterol precursor squalene to 2,3-oxidosqualene, leading to the build up of squalene and the depletion of ergosterol in the cell membrane, which may affect cell permeability and organisation (Ryder *et al.*, 1998; White *et al.*, 1998; Ghannoum & Rice, 1999). The drug is fungicidal against dermatophytes and some filamentous fungi and demonstrates high activity against *Cr. neoformans*. Terbinafine is used very successfully in the treatment of cutaneous *Candida* infections, and may be administered topically or orally. However, terbinafine exhibits no activity against *C. tropicalis*, *C. glabrata* or *C. krusei* (Ryder *et al.*, 1998) and only fungistatic activity in *C. albicans* (Ryder *et al.*, 1998).

1.5.1.2. Morpholine derivatives

The morpholine derivatives are a group of totally synthetic compounds that includes fenpropimorph, tridemorph and amorolfine, of which only amorolfine is used clinically, the other two compounds are mainly used in the preparation of agricultural fungicides. Amorolfine is used topically to treat superficial infections, and is highly fungicidal against dermatophytic and filamentous fungi. However, it is only fungistatic against most yeast species. The compound inhibits two enzymes that act in the biosynthetic pathway of ergosterol, namely the C14 sterol reductase encoded by the *ERG24* gene, and the C8 sterol isomerase encoded by the *ERG2* gene (Fig. 1.4). This results in the depletion of ergosterol synthesised by the cell and the build-up of ignosterol. A similar build-up of ignosterol occurs in *S. cerevisiae* cells upon deletion of the *ERG24* gene, disrupting the cell membrane and inhibiting the uptake of glucose and pyrimidines (Vanden Bossche, 1997; Sanglard & Bille 2002; Akins, 2005).

1.5.1.3. Azole-based antifungals

The azoles are a group of broad-spectrum synthetic antifungal compounds that are in widespread therapeutic use against yeasts, filamentous and dermatophytic fungi. There are currently three classes of azoles: the N1-substituted imidazoles (ketonazole, miconazole and clotrimazole), the triazoles (fluconazole and itraconazole) and the new

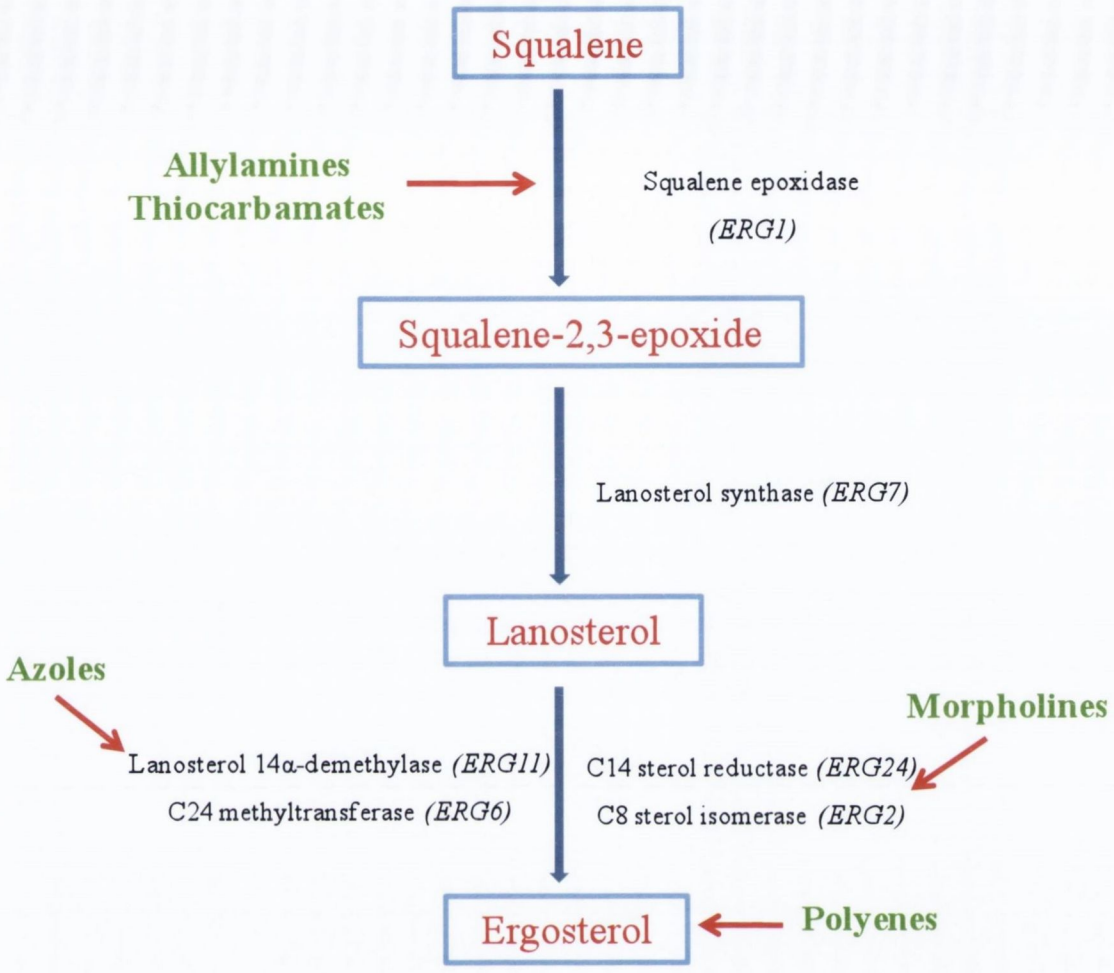


Figure 1.4. The latter stages of the metabolic synthetic pathway of ergosterol in yeast highlighting the most critical enzymes involved.

The biosynthetic enzymes involved in each metabolic step are displayed in black typeface, the genes encoding each enzyme are highlighted in black, italicised typeface, the substrates and products of each step are shown in red typeface, and antifungal agents that act on biosynthetic enzymes or the end product ergosterol are indicated in green typeface.

generation of azoles which are also triazoles (posaconazole, ravuconazole and voriconazole). The azoles target a P450 cytochrome enzyme, lanosterol demethylase, which is encoded by the *ERG11* gene (Fig. 1.4). The lanosterol demethylase ensures that C4-methyl groups are removed from lanosterol to generate ergosterol, which is necessary to maintain the fluidity and integrity of the cell membranes. The haem moiety of the cytochrome enzyme binds the free nitrogen of the azole ring as a sixth ligand, preventing the activation of oxygen which is necessary for the demethylation of lanosterol. This leads to a depletion of ergosterol in the cell membrane and an accumulation of sterol precursors, resulting in a cell membrane with altered structure and function. The azoles are fungistatic against most yeasts and fungi, and are fungicidal against *Cr. neoformans, Aspergillus* and dermatophytic fungi. The azoles are poorly soluble in water, with the exception of fluconazole. They are generally administered as oral preparations, and fluconazole, itraconazole and voriconazole can also be administered intravenously (White *et al.*, 1998; Ghannoum & Rice, 1999; Sanglard, 2002; Sanglard & Bille, 2002).

1.5.1.4. Polyenes

The polyenes are a class of organic antifungal compounds that are produced by some species of *Streptomyces*. The polyenes are characterised by the presence of a macrolide ring which contains a series of at least three conjugated double bonds. One of the most successful is amphotericin B, a potent broad-spectrum fungicidal drug produced by *Streptomyces nodosus* that is used to treat systemic *Candida* infections. It is soluble in both basic and acidic environments. However, it has a poor solubility in water, hampering intramuscular or oral administration (Sanglard & Bille, 2002). The drug binds hydrophobically to ergosterol in the cell membrane of the target cell (Fig. 1.4), resulting in the formation of aqueous pores and thus affecting membrane permeability which causes leakage of cytoplasmic components and leads to cell death. It is also thought to cause oxidative damage to the fungal cell. The higher affinity of the drug for ergosterol in fungal cell membranes rather than to cholesterol in human cell membranes enables this drug to be used clinically. However, binding of the drug to cholesterol can lead to nephrotoxicity. For this reason, amphotericin B is typically formulated in liposomes, an example of which is available clinically as Ambisome. These formulations enable the delivery of higher dosages, as well as reduced toxicity. The drug is transferred to the ergosterol containing target cells by a selective transfer

mechanism from the "donor" liposomes to the "target" ergosterol-containing fungal cell membrane with the help of fungal/human cell membrane phospholipases, thus avoiding toxicity to the host (Ghannoum & Rice, 1999; Sanglard & Bille, 2002).

Nystatin is another polyene antifungal that is used to treat mucosal *Candida* infections. Examples of commercially available preparations include Nystex and Nystan. A liposomal formulation of the drug also exists as Nyotran.

1.5.2. Glucan synthesis inhibitors

1.5.2.1. Cyclic lipopeptides/echinocandins

The most recently developed class of antifungals consists of the acetylated cyclic lipopeptides, of which the echinocandins are in clinical use. These compounds were originally obtained from soil fungi. Semi-synthetic derivatives of these compounds display broad-spectrum and highly fungicidal activity by attacking the biosynthetic pathway of glucans, which play a pivotal role in the organisation of the cell wall (Groll *et al.*, 1998). Echinocandins, such as caspofungin, micafungin and anidulafungin are used successfully in the treatment of cutaneous and invasive *Candida* and *Aspergillus* infections (Baixench *et al.*, 2007). The echinocandins act on β-(1,3)-glucan synthetase, an integral, heterodimeric membrane protein that is responsible for glucan synthesis. Inhibition of glucan synthesis leads to increased chitin levels in the cell wall and decreased ergosterol and lanosterol levels in the cell membrane, eventually leading to a thickening of the cell wall, failure of cells to separate during budding, pseudohyphal growth, and osmotic sensitivity (Ghannoum & Rice, 1999).

1.5.3. Nucleic acid synthesis inhibitors

1.5.3.1. 5-Fluorocytosine

5-Fluorocytosine (Systematic name: 4–amino–5–fluoro–1, 2–dihydropyrimidine–2–one) was first synthesised in 1957 as a fluorinated analogue of cytosine, for use in anti-cancer treatment (Heidelberger *et al.*, 1957, 1958). However, it was used more successfully in the treatment of human candidiasis and cryptococcosis (Tassel & Madoff, 1968). A fluorinated pyrimidine analogue, it is highly water soluble enabling oral and intravenous administration, and is fungicidal to susceptible yeasts and fungi. The compound is taken into the cell by a purine-cytosine permease as a prodrug, and is converted to its toxic form 5-fluorouracil (5FU) by the activity of a cytosine

deaminase. After deamination, 5FU is phosphorylated by a uracil phosphoribosyl tranferase (UPRT) and two specific kinases, eventually converting it to 5-fluorouridine triphosphate (5FUTP), which in turn gets incorporated into the cells' RNA, causing miscoding and leading to an inhibition of fungal protein synthesis. The administration of 5FC as a non-toxic prodrug prevents the toxicity of this pyrimidine compound for mammalian cells. Mammalian cells do not possess a cytosine deaminase to catalyse the deamination of the antifungal drug to its toxic form, and therefore it is not toxic to the cells of the host. However, the conversion of 5FC to 5FU by intestinal bacteria is possible which can result in toxicity in oral formulations of the drug (Sanglard, 2002).

1.6. Resistance mechanisms to antifungal agents.

Antifungal drug resistance is defined as persistence of a fungal pathogen despite antifungal therapy. An isolate is classified as resistant to a specific antifungal agent when the MIC value exhibited by the isolate exceeds the average MIC values for the rest of the population. Both intrinsic and acquired resistance mechanisms to antifungal therapeutics have been described for many yeast pathogens. Intrinsic resistance (i.e., primary resistance prior to any drug exposure) to fluconazole exists in *C. krusei* (Orozco *et al.*, 1998). Resistance to antifungal drugs is usually acquired by the stepwise modification of target genes in order to accumulate resistance determinants amongst clonal populations. Genetic exchange and recombination can occur across species that are very closely related and can mate together. However, horizontal transfer of genetic elements is uncommon, in contrast to the methods of resistance determinant acquisition that occur in bacterial populations (Anderson, 2005). The rate of mutation or microevolution in a species will affect its ability to acquire resistance, as will the fungistatic or fungicidal activity of the drug in question (Anderson, 2005). The major mechanisms by which fungal cells exhibit resistance are as follows:

- By reduction in drug accumulation *via* a decrease in drug uptake.
- By increase in drug efflux, alteration in the expression or genetic sequence of the drug target.
- By modification or degradation of the drug.

1.6.1. Ergosterol-targeting antifungal agents

1.6.1.1. Resistance to allylamines and thiocarbamates

Terbinafine resistance has been observed in fluconazole-susceptible *C. albicans* isolates, but terbinafine cross-resistance can occur in fluconazole-resistant *C. albicans* isolates. It is thought that there may be two separate mechanisms at work in order to produce resistance. Multidrug efflux transporters have been described previously that can confer resistance to a variety of compounds. It has been shown in *C. albicans* that the *CDR1* and *CDR2* genes encoding such transporters may confer resistance to terbinafine as well as to fluconazole (Sanglard *et al.*, 1996, 1997; Ryder *et al.*, 1998).

1.6.1.2. Resistance to morpholine derivatives

Acquired resistance to the morpholine group of antifungal compounds has yet to be reported clinically due to its limited use in the treatment of superficial fungal infections to date. However, resistance has been generated *in vitro* in *S. cerevisiae* through the overexpression of the *ERG24* gene (Fig. 1.4). The multidrug efflux pumps encoded by the *CDR1* and *CDR2* genes have been observed to bind the azole-derivatives as well as the morpholines and terbinafine, conferring resistance in *C. albicans* (Sanglard *et al.*, 1997; Sanglard, 2002). Thus it appears that the acquisition of resistance to this drug in clinical conditions is likely in the future.

1.6.1.3. Resistance to azole-based antifungals

Resistance to the azole class of antifungals has been well documented. Both intrinsic and acquired forms of resistance have been described in many different species. *Mucor* species are instrinsically resistant to the azoles, whereas *C. krusei* and *A. fumigatus* are intrinsically resistant to fluconazole but are susceptible to itraconazole and voriconazole (Orozco *et al.*, 1998; Bille, 2000; Sanglard, 2002; Sanglard & Bille, 2002). The acquisition of fluconazole resistance has been reported in *C. albicans, C. dubliniensis* and *C. glabrata* in clinical settings following treatment with the drug (Ghannoum & Rice, 1999). Resistance to fluconazole in *C. albicans* was first associated with AIDS patients receiving azole prophlaxis, however, it is now also found in patients who have not previously received treatment using azoles, as well as in HIV-negative patients (White *et al.*, 1998).

Azole resistance is achieved by an alteration in drug transport, alteration of the *ERG11*-encoded cellular target, or modification of the ergosterol biosynthesis pathway (Fig. 1.4) (White *et al.*, 1998; Ghannoum & Rice, 1999; Sanglard & Bille, 2002). In species such as *C. glabrata, C. albicans, C. krusei* and *C. dubliniensis,* fluconazole resistance has been linked with a failure to accumulate intracellular quantities of the drug, most commonly by overexpression of the *CDR1*- and *MDR1*-encoded multidrug efflux transporters or their homologues (Sanglard *et al.*, 1995; Moran *et al.*, 1998; White *et al.*, 1998; Ghannoum & Rice, 1999; Sanglard *et al.*, 1999; Sanglard, 2002; Sanglard & Bille, 2002).

Conformational alteration of the lanosterol demethylase target by various mutations of the *ERG11* gene (Fig. 1.4) has been reported as a method of azole-resistance in *C. tropicalis*, *C. dubliniensis* and *C. albicans* (Lamb *et al.*, 1997; White, 1997; Perea *et al.*, 2002; Vandeputte *et al.*, 2005). Some 83 amino acid changes in the *ERG11* gene product have previously been associated with azole-resistance in *C. albicans* (Sanglard, 2002). Upregulation of the lanosterol demethylase target has also been reported in *C. albicans, C. dubliniensis* and *C. glabrata* as a method of resistance (Marichal *et al.*, 1997; Perea *et al.*, 2002; Sanglard, 2002).

Modification of the ergosterol biosynthetic pathway has been previously noted in two distinct clinical *C. albicans* isolates exhibiting azole-resistance. These isolates also exhibited resistance to amphotericin B, correlating with alteration of the sterol content in the cell walls, namely an accumulation ergosta-7, 22-dienol-3β-ol instead of ergosterol. The altered cell wall content was linked with a defect in the *ERG3*-encoded C5 sterol desaturase, which catalyses one of the steps in the conversion of lanosterol to ergosterol. This defect has also been noted in *S. cerevisiae* (Kelly *et al.*, 1997).

1.6.1.4. Resistance to polyenes

Intrinsic resistance to amphotericin B is a relatively rare event, but it has been reported in *Candida lusitaniae*, *Candida guillermondii* and in some moulds such as *Fusarium* spp. (Bille, 2000). Acquired resistance has been reported in *Candida, Aspergillus* and *Cryptococcus* species *in vitro* (White *et al.*, 1998). Cross-resistance to amphotericin B has been reported in clinical *C. albicans* isolates that were recovered from HIV-infected and leukaemic individuals who were receiving azole prophylaxis. A defect in sterol desaturation leads to modification of the sterol content in cell membranes of these resistant isolates, namely replacement of ergosterol with other

sterols such as 3β-ergost-7, 22-dienol and 3β-ergosta-8-enol (Kelly *et al.*, 1996, 1997; Nolte *et al.*, 1997; Ghannoum & Rice, 1999; Sanglard, 2002). Increased catalase activity is another method used by the resistant cells in order to diminish the oxidative damage caused by amphotericin B (Sanglard, 2002; Sanglard & Bille, 2002).

Resistance to nystatin has been reported rarely, occurring as a reduction or a substitution in the ergosterol content of the cell membrane, thus reducing the affinity of the drug for its target (Ghannoum & Rice, 1999).

1.6.2. Glucan synthesis inhibitors

1.6.2.1. Resistance to cyclic lipopeptides/echinocandins

Resistance to this class of antifungal agents has been reported infrequently to date. However, it has been observed clinically in isolates of *C. albicans, C. glabrata, C. parapsilosis* and *C. krusei* (Baixench *et al.*, 2007), and resistant mutants have also been generated *in vitro*. Resistance in all cases has been attributed to point mutations resulting in amino acid substitutions in the gene encoding β-(1,3)-glucan synthetase (Baixench *et al.*, 2007; Ghannoum & Rice, 1999). Isolates of *C. parapsilosis* that exhibited high caspofungin MIC values also exhibited elevated micafungin and anidulafungin MICs, indicating cross resistance (Baixench *et al.*, 2007).

1.6.3. Nucleic acid synthesis inhibitors

1.6.3.1. Resistance to 5-fluorocytosine

Intrinsic resistance to this drug has been reported in many clinically relevant species such as *C. albicans, C. glabrata, C.krusei, C. tropicalis,* and *Cr. neoformans* (Coleman *et al.*, 1998; Ghannoum & Rice, 1999; Sanglard, 2002), and secondary resistance can be acquired rapidly (Stiller *et al.*, 1983). For this reason, 5FC is usually administered in combination with other antifungal drugs, primarily amphotericin B (Abele-Horn *et al.*, 1996; Sanglard, 2002). In *C. dubliniensis, C. albicans* and *C. tropicalis,* 5FC resistance has been associated with clonal clusters of isolates. A reduced susceptibility to 5FC is exhibited by 72.7% of *C. albicans* isolates belonging to MLST clade 1 and Ca3 fingerprint group I (Pujol *et al.*, 2004; Odds *et al.*, 2007). In the *C. dubliniensis* population, the majority of isolates from the Middle East belonging to Cd25 fingerprint group III exhibit high levels of intrinsic 5FC resistance (Al Mosaid *et al.*, 2005), and a clonal group of 5FC-resistant isolates has been noted in Paris, France

amongst a population of *C. tropicalis* that was otherwise genetically diverse (Desnos-Ollivier *et al.*, 2008).

Resistance mechanisms that have been reported to date include, reduced drug uptake by mutation of purine-cytosine permeases, or alterion of the metabolic pathway by which the drug is metabolised such as by mutations of the cytosine deaminase or the UPRT enzymes (Erbs *et al.*, 1997; Stolz & Vielreicher, 2003; Hope *et al.*, 2004; Chapeland-Leclerc *et al.*, 2005; Paluszynski *et al.*, 2006; Papon *et al.*, 2007).

Aims of the present study

Candida dubliniensis was first described in 1995 as a species that is genetically and phenotypically very similar to the most pathogenic of the *Candida* species, *C. albicans*, so that misidentification of *C. dubliniensis* as *C. albicans* has most likely underestimated the true prevalence of the species. Furthermore, the true prevalence of *C. dubliniensis* in non-human sources is yet to be determined. In recent years both culture- and PCR-based methods have been developed to discriminate between the two species with high levels of accuracy.

The population structure of *C. dubliniensis* has been defined previously by DNA fingerprinting using the species-specific complex probe Cd25. Three main clades (termed Cd25 groups I–III) have been reported (Joly *et al.*, 1999; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005), of which Cd25 group III consists of isolates recovered in the Middle East, most of which exhibit high-level intrinsic resistance to the antifungal agent 5FC. Although DNA fingerprinting analysis is informative and offers high levels of discrimination, poor reproducibility and labour intensity make a highly reproducible, rapid technique such as MLST an appealing alternative for population structure analysis of this species.

The objectives of the present study were:

- To determine the prevalence of *C. dubliniensis* in the normal healthy population, and in an immunocompromised human population, as well as the prevalence of the species in avian sources.
- To determine if MLST could be used as an alternative method to investigate the population structure of *C. dubliniensis,* and to examine the correlation between studies using the complex Cd25 fingerprinting probe and using MLST.
- To investigate the application of a common MLST scheme to both *C. albicans* and *C. dubliniensis* as a method to enable a comparison of their respective population structures, as well as to assess the evolutionary relatedness of the two species.
- To investigate the molecular mechanism(s) of 5FC resistance in Cd25 group III *C. dubliniensis* isolates by comparative DNA and amino acid sequence analysis of the enzymes involved in 5FC metabolism in both 5FC-resistant and 5FC-susceptible isolates recovered in the Middle East.

Chapter 2

General Materials and Methods

2.1. General microbiological methods

2.1.1. *Candida* isolates and culture conditions

All *Candida* isolates were routinely cultured on potato dextrose (PDA) agar (Oxoid, Basingstoke, Hants., UK) at pH 5.6, at 37°C for 48 h. Liquid cultures were grown overnight at 37°C in yeast peptone "dextrose" (YPD) broth (per litre: 10 g yeast extract [Sigma-Aldrich Ltd., Tallaght, Dublin, Republic of Ireland] 20 g bactopeptone [Difco, Detroit, MI, USA] and 20 g glucose pH 5.5) 37°C in a Gallenkamp model G25 orbital incubator (Leicester, UK) at 200 rpm.

The *Escherichia coli* strain DH5α (Hanahan, 1983) [F^- ϕ80d, *lacZ*Δm15 *endA1 recA1 relA1 hsdR17* (rK-mK$^+$) *supE44 thi-1* λ- *gyrA96* Δ (*laclZYA-argF*) U169] was routinely grown on Luria-Bertani agar (LA) pH 7.4 at 37°C (Lennox 1955), or in Luria-Bertani broth (LB) pH 7.4 at 37°C with shaking at 200 rpm for liquid cultures.

2.1.2. Chemicals, enzymes, radioisotopes, and antifungal drugs

All chemicals used in the study were of analytical grade or molecular biology grade and were purchased from Sigma-Aldrich Ltd., Fisher Scientific Ltd. (Bishop Meadow Rd. Loughborough, UK) or from Roche Diagnostic Ltd. (Lewes, East Sussex, UK).

Enzymes for molecular biology procedures as well as dNTPs, DNA loading dye and DNA molecular markers were purchased from the Promega Corporation, (Madison, WI, USA), Roche Applied Science, (Mannheim, Germany), NewEngland Bioloabs Inc. (Ipswich, MA, USA), Ambion (Applied Biosystems, Warrington, UK) or Invitrogen (Biosciences Ltd., Dun Laoghaire, Dublin, Republic of Ireland). Custom synthesised oligonucleotides were purchased from Sigma Genosys Biotechnologies Europe Ltd. (Pampisford, Cambridgeshire, UK). DNA purification kits were purchased from Qiagen (Qiagen Science, Crawley, West Sussex, UK) (DNeasy® blood and tissue kit), Sigma-Aldrich Ltd. (GenElute™ PCR kit) or Promega (Wizard® SV Gel and PCR Clean up system).

Zymolyase 20T (20,900 U/g) was purchased from Seikagaku Corporation (Tokyo, Japan) and was stored at 4°C. Proteinase K (Roche) solutions were prepared in Milli-Q® Biocel-purified water (resistivity 18.2 MΩ•cm) (Millipore™, Carrigtwohill, Cork, Republic of Ireland) at a concentration of 20 mg/ml and were stored at –20°C.

Radiolabelled 5-fluorocytosine (Sigma-Aldrich Ltd.) was prepared in aqueous ethanol (1:1) at a concentration of 78.7 μM and was stored at 4°C. Stock solutions of 5-fluorocytosine and 5-fluorouracil (Sigma-Aldrich Ltd.) were prepared in Milli-Q® Biocel-purified water at a concentration of 10 mg/ml, and stored at –80°C.

2.1.3. Buffers and solutions

Tris-EDTA (TE) buffer was routinely used in many experiments and consisted of 10 mM Tris-HCl, 1 mM EDTA, pH 8.0. TBE buffer was prepared at a 5 × concentration and consisted of 0.45 M trizma base, 0.45 M boric acid, and 0.01 M EDTA. This was diluted to a 0.5 × concentration in Milli-Q® Biocel-purified water water for use as a buffer in agarose gel electrophoresis.

Liquefied phenol washed in Tris-buffer was purchased from Fisher Scientific Ltd. and was used in the preparation of phenol:chloroform:isoamyl alcohol (24:24:1) using 24 ml of phenol, 24 ml of chloroform, and 1 ml of isoamyl alcohol. This solution was stored at 4°C in the dark.

Cell breaking buffer consisted of 2% (v/v) Triton X-100, 1 mM EDTA, 1% (w/v) sodium dodecyl sulphate (SDS), 100 mM NaCl, and 10 mM Tris-HCl, pH 8.0. The solution was stored at 4°C.

2.1.4. Storage of *Candida* isolates

Candida isolates were maintained on plastic beads in Microbank cryogenic vials (Pro-lab Diagnostics, Cheshire, UK) at –80°C. Stored isolates were reactivated by sub-culture on PDA plates and incubated for 48 h at 37°C.

2.2. Identification of *Candida* species

2.2.1. Chromogenic media

Candida Selective Chromogenic agar (Oxoid Ltd. Hampshire, UK) and CHROMagar Candida™ medium (CHROMagar) are commercially available agars, each containing chromogenic substrates which allow colonies of several medically important *Candida* species to be presumptively identified on the basis of colony colour and morphology. Colonies of *C. albicans* (light green-blue/green colonies), *C. glabrata* (dark pink colonies with paler pink edges), *C. krusei* (pale, flat, papillate colonies with spreading edges) and *C. tropicalis* (blue/gray in the centre surrounded by a dark brown/purple halo in the agar surrounding the colony) can be distinguished from each

other upon primary isolation. The medium has been shown to be clinically useful in the presumptive identification of these species (Odds & Bernaerts, 1994). *Candida dubliniensis* colonies from clinical samples are typically dark green-blue on this medium but the colours yielded from the colonies can fade following sub-culture or storage (Coleman *et al.*, 1997a; Schoofs *et al.*, 1997; Kirkpatrick *et al.*, 1998).

2.2.2. Chlamydospore and hyphal fringe production

Pal's agar was prepared freshly using the extract of unsalted sunflower seeds (including shells and kernels). First, an aqueous extract of sunflower seeds was prepared by pulverising 50 g of seeds in a domestic blender for 5 min and then adding the ground seeds to 1 litre of distilled water, followed by boiling for 30 min. The seed extract was then cooled and filtered, and the following ingredients were added; glucose (1 g), KH_2PO_4 (1 g), and creatinine (1 g). The pH was adjusted to 5.5, the volume was readjusted to 1 litre, and 15 g of agar (Difco) was added before the mixture was autoclaved at 110°C for 20 min. Putative *C. dubliniensis* isolates were plated on this medium and incubated at 30°C for 48–72 h. Following incubation, isolates were examined for the presence of a hyphal fringe composed of hyphae, pseudohyphae and blastospores that is a specific characteristic of *C. dubliniensis* on this medium (Al Mosaid *et al.*, 2003).

2.2.3. Carbohydrate assimilation profiles

Biotyping was carried out using the API ID32C yeast identification system (bioMérieux) which identifies *Candida* isolates to the species level using a series of standard substrates that are each contained in separate cupules along a plastic strip with a specially adapted database (Pincus *et al.*, 1999). Tests were carried out according to the manufacturer's instructions. For each isolate, an inoculum was prepared from 24–48 h old colonies cultured on PDA medium. Four colonies of 3–4 mm in diameter were resuspended in Milli-Q® Biocel-purified water to a turbidity equivalent to a 2 McFarland standard. This suspension was then used to inoculate an aliquot of 'C medium' which was supplied by the manufacturers. Each of the cupules in the strip was then inoculated with 135 µl of the C suspension medium and incubated for 48 h at 30°C. Readings were made at 24 h and 48 h by visually assessing the growth of the test isolates in each of the cupules in comparison to that in the negative control cupule. The presence or absence of growth was recorded for each cupule on a result sheet supplied

by the manufacturers, and the substrate assimilation pattern for each isolate was converted into a ten-digit numerical profile. These profiles were then cross-referenced in the APILAB ID32C software package (version 3.3.3) (bioMérieux, Paris, France). In the database, each profile is listed along with the percentage of identification, which is an estimate of how closely the profile corresponds to that of a particular taxon, relative to all the other taxa in the database and the T index, which is an estimate of how closely the profile corresponds to the most typical set of reactions for a particular taxon. Based on these parameters, a set of reactions which closely resemble those of a particular taxon will be classed as an 'excellent' or 'good' identification, and will yield an identification to the species level, whereas atypical results will be classed as having 'poor' or 'low' discriminatory powers and are usually unable to yield a reliable identification.

2.3. Preparation of template DNA from *C. dubliniensis* isolates

Three different methods were used to prepare template DNA from *Candida* isolates. The boiling method was used to identify putative *C. dubliniensis* colonies on solid medium by means of PCR. The bead-beating method was used to prepare DNA that could then be stored at 4°C and could be used in many different PCR amplification experiments on separate occasions. Template DNA was prepared using the Qiagen DNeasy® Blood and Tissue kit for isolates that were to undergo PCR amplification followed by direct sequencing of purified amplimers. The concentration of DNA samples was assessed by measuring their absorbance at 260 nm using a spectrophotometer (Genosys 2, ThermoSpectronic; AGB Dublin, Ireland) and calculating the concentration using the following formula: 1 unit of A_{260} = 50 µg DNA.

2.3.1. Boiling method

A single colony from a culture grown for 48 h at 37°C on PDA was suspended in 50 µl of Milli-Q® Biocel-purified water (Millipore™). Cell suspensions were boiled for 10 min and the lysed cells were subjected to a clearing spin at 20,000 × *g*, for 10 min in an Eppendorf (model 5417C) microfuge (Eppendorf, Hamburg, Germany). Template DNA contained in 25 µl of the supernatant fluid was used for amplification using PCR (Donnelly *et al.*, 1999).

2.3.2. Bead-beating method

Isolates were grown overnight in 5 ml of YPD broth at 37°C with shaking at 200 rpm in an orbital incubator (Gallenkamp, Leicester, UK). Cells were harvested from 1.5 ml of culture by centrifugation at 14,000 × *g*, and the resulting pellet was resuspended in 200 µl of cell breaking buffer (Section 2.1.3) and transferred to a 2 ml screw-capped tube (Sarstedt, Drinagh, Wexford, Republic of Ireland) which contained 0.3 g acid washed glass beads (Sigma-Aldrich Ltd.). Following the addition of 200 µl of a mixture of phenol:chloroform:isoamyl alcohol (24:24:1) the cells were disrupted in a BIO101FastPrep instrument (Qbiogene, Cambridge, UK) for 1 min at maximum speed, followed by centrifugation at 14,000 × *g* for 10 min. The aqueous phase was removed and extracted twice with an equal volume of chloroform:isoamyl alcohol (24:1), and nucleic acids were precipitated by the addition of 400 µl 70% (v/v) ethanol at –20°C. Purified DNA was pelleted by centrifugation for 10 min, washed in 70% (v/v) ethanol, dried, and resuspended in 50 µl of Milli-Q® Biocel-purified water.

2.3.3. DNA extraction for PCR amplification and DNA sequencing

Template DNA was prepared from isolates using the Qiagen DNeasy® blood and tissue kit for amplification by PCR followed by direct sequencing of purified amplimers. Briefly, isolates were grown overnight in 5 ml of YPD broth at 37°C with shaking at 200 rpm in an orbital incubator (Gallenkamp). Cells were harvested from 1.5 ml of culture by centrifugation at 14,000 × *g* and were resuspended in 600 µl sorbitol buffer (1 M sorbitol, 100 mM EDTA, 14 mM β-mercaptoethanol). Cell walls were digested by addition of 10 mg Zymolyase 20T and incubation at 37°C for 30 min. Cell protoplasts were harvested by centrifugation at 300 × *g* for 10 min and were treated with 20 µl of 10 mg/ml Proteinase K followed by incubation at 56°C for 15 min in order to lyse the cells. DNA was recovered from the lysed cells using the DNeasy® mini spin columns and collection tubes according to the protocol outlined by the manufacturers.

2.4. Identification and characterisation of *C. dubliniensis* using PCR amplification

2.4.1. Identification of *C. dubliniensis* using PCR amplification

Isolates of *C. dubliniensis* were identified using the species-specific PCR primers (DUBF/DUBR), which target the intron of the *ACT1* gene in *C. dubliniensis*

(Donnelly *et al.*, 1999) (Table 2.1). These PCR amplifications were carried out in a final reaction volume of 50 µl, containing 10 pmol of each of the forward and reverse primers, 2.5 mM $MgCl_2$, 10 mM Tris/HCl (pH 9.0 at 25°C) 10 mM KCl, 0.1% (v/v) Triton X-100, 1.25 U GoTaq polymerase (Promega) and 25 µl of template DNA. Each reaction mixture also contained 10 pmol each of the universal fungal primers RNAF/RNAR (Fell, 1993) which amplify approximately 614 bp from all fungal large-subunit rRNA genes and were used as an internal positive control (Table 2.1). Cycling conditions consisted of 6 min at 95°C followed by 30 cycles of 30 s at 94°C, 30 s at 58°C, 30 s at 72°C, and finally followed by 72°C for 10 min. Amplification products were visualized on a UV transilluminator which emits light at 345 nm (Ultra Violet Products Ltd., Cambridge, UK) following electrophoresis through 2.0% (w/v) agarose gels containing 0.5 µg ethidium bromide/ml.

2.4.2. Genotyping of *C. dubliniensis* isolates using PCR amplification

Template DNA extracted by the bead-beating method described above (Section 2.3.1) was used to determine the genotypes of all isolates. Genotypes were defined according to the sequences of the ITS region as previously described (Gee *et al.*, 2002). Primers specific to each genotype were used in separate PCRs to amplify sequences of the ITS regions 1 and 2 and their intervening 5.8S rRNA gene as described previously (Gee *et al.*, 2002). Each PCR was carried out in a 50 µl volume and contained 10 pmol of the universal fungal primer pair RNAF/RNAR, and 10 pmol of one pair of the four genotype specific primers G1F/G1R, G2F/G2R, G3F/G3R, and G4F/G4R (Table 2.1). Each reaction also contained 2.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 9.0 at 25°C), 10 mM KCl, 0.1% (v/v) Triton X-100, 1.25 U GoTaq polymerase (Promega), 200 µM of each dNTP (Promega) and 100 ng of template *C. dubliniensis* DNA. The PCR cycling conditions for the genotype 1 and 2 primer pairs were as follows: 95°C for 3 min, followed by 30 cycles of 95°C for 1 min, 50°C for 1 min, 72°C for 20 s, and followed by a final 10 min extension step at 72°C. Amplifications using the genotype 3 specific primers were carried out with a separate set of reaction conditions which used an annealing temperature of 63°C. Amplifications using the genotype 4 specific primers were also carried out separately, using 20 µM dNTPs and an annealing temperature of 55°C. Amplification products were separated by electrophoresis through 2% (w/v) agarose gels containing 0.5 µl ethidium bromide/ml and were visualised on a UV transilluminator.

2.5. Recombinant DNA techniques

2.5.1. Polymerase chain reaction (PCR)

Oligonucleotide primers were synthesised commercially by Sigma-Aldrich Ltd. and were used in many separate PCR amplification experiments each requiring a separate set of reaction conditions, which will be described separately.

2.5.2. Purification of PCR amplimers

PCR-amplified template DNA was purified from remaining PCR reagents using the silica-gel-binding properties of the Sigma GenElute® PCR CleanUp kit (Sigma-Aldrich Ltd.) and the Qiaquick® Multiwell PCR purification kit (Qiagen Ltd. West Sussex, UK). Both kits were used according to the manufacturers' instructions. PCR amplimers were purified from agarose gels using a Qiaex® Gel Extraction Kit (Qiagen Ltd.) or the Wizard SV Gel and PCR clean up system (Promega) according to the manufacturers' instructions.

2.5.3. Small-scale isolation of plasmid DNA from *Escherichia coli*

Small-scale preparations of plasmid DNA from *E. coli* strain DH5α were made using the GenElute™ (Sigma-Aldrich Ltd.) plasmid miniprep kit. Cells were subjected to a modified alkaline-SDS lysis procedure, followed by adsorption of the DNA onto silica in the presence of high salts, according to the manufacturers' instructions.

2.5.4. Ligation of DNA fragments

Purified DNA fragments were ligated to P-GEM T-Easy Vector System I vector DNA (Promega) in 10 µl reaction volumes with a 3:1 ratio of insert to vector DNA in 1 × ligase buffer with 1 U of T4 DNA ligase. Reactions were carried out at 4°C for 18 h. Ligations of purified DNA fragments to the tetracycline-inducible expression plasmid pNIM1 were carried out via restriction sites which had been designed within oligonucleotide primers used in PCR amplifications. Ligation reactions were carried out in a 13 µl reaction volume with a 4:1 ratio of insert to vector in 0.77 × ligase buffer, with 1 U of T4 DNA ligase (Promega). Ligation reactions were carried out at 4°C for 18 h.

Table 2.1. Oligonucleotide primers used in the identification and ITS genotyping of *C. dubliniensis* isolates

Primer	Amplicon size	Primer Sequence (5′→3′)	Primer Target	GenBank Accession Number	Amplicon Coordinates[a]	Reference
DUB F	288 bp	GTATTTGTCGTTCCCCTTTC	*ACT1* intron	AJ236897	+ 251→ + 270	(Donnelly *et al.*, 1999)
DUB R		GTGTTGTGTGCACTAACGTC			+ 519 → + 538	
RNA F	614 bp	GCATATCAATAAGCGGAGGAAAAG	Large ribosomal subunit gene	X83718	+40 → + 63	(Fell, 1993)
RNA R		GGTCCGTGTTTCAAGCAG			+637 → + 654	
G1F	331 bp	TTGGCGGTGGGCCCCTG	ITS region 1	AJ311895	+ 80 → + 96	(Gee *et al.*, 2002)
G1R		AGCATCTCCGCCTTATA			+ 394 → + 410	
G2F	323 bp	CGGTGGGCCTCTACC	ITS region 2	AJ311896	+ 54 → + 68	(Gee *et al.*, 2002)
G2R		CATCTCCGCCTTACC			+ 362 → + 376	
G3F	329 bp	TTGGTGGTGGGCTTCTG	ITS region 3	AJ311897	+ 50 → + 66	(Gee *et al.*, 2002)
G3R		GCAATCTCCGCCTTACC			+ 362 → + 378	
[b]G4F	321 bp	GGCCTCTGCCTGCCGCCAGAGGATG	ITS region 4	AJ311898	+ 59 → + 83	(Gee *et al.*, 2002)
[b]G4R		AGCAATCTCCGCCTTACT			+ 362 → + 379	

Abbreviations: ITS region, Internal transcribed spacer region and the intervening 5.8S rRNA gene.

[a]Nucleotide coordinates for amplicons are as indicated, with the first base of the ATG start codon designated +1 for the *ACT1* and *RNA* genes, and with the first base of the ITS region sequence being designated +1 for the ITS amplicons.

[b]The G4F and G4R primers each differed from the ITS genotype 4 sequence by a single base-pair change at the 3′ end to improve specificity.

2.5.5. Transformation of competent *E. coli* prepared using $CaCl_2$

Transformation of competent *E. coli* strain DH5α prepared with $CaCl_2$ was carried out using a modification of the method of Sambrook *et al.* (1989). *Eschericia coli* strain DH5α was inoculated from an overnight broth culture into 200 ml of fresh LB and grown at 200 rpm in an orbital incubator at 37°C for 3 h to an A_{550} of 0.4–0.5. The culture was then decanted into two 100 ml volumes in ice-cold Sorvall tubes. The cells were pelleted by centrifugation in a Sorvall RC-5C plus centrifuge (GMI Inc. Ramsey, MN) at 6,000 × g at 4°C for 10 min. Each pellet was resuspended in 45 ml of ice-cold 0.1 M $CaCl_2$, and left on ice for 15 min with occasional shaking. The cells were then pelleted by centrifugation as before. The pellets were then resuspended in a volume of 5 ml 50 mM $CaCl_2$ for each 100 ml of original culture. A 300 µl aliquot of this cell suspension was transferred to a sterile 1.5 ml microfuge tube on ice for each transformation experiment. Plasmid DNA (up to 50 ng) was added to each tube and incubated on ice for 30 min. A known amount of a standard plasmid preparation was added to a separate tube as a positive control, and a second negative control tube was also included which contained no plasmid DNA. The tubes were then heat shocked at 42°C for exactly 2 min and rapidly transferred to an ice bath. The cells were then incubated at 37°C in a water bath in the presence of 500 µl LB medium to allow the cells to recover and express the antibiotic resistance marker (ampicillin resistance in the case of pGEM T-Easy Vector). A 200 µl aliquot of this suspension was then spread on LA plates containing selective antibiotic (100 µg ampicillin/ml), 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG, Boerhringer Mannheim) and 100 µg (5-bromo-4-chloro-3-indoyl-(-D-galactopyranoside (X-gal, Roche) and incubated for 20 h at 37°C. Recombinant colonies were identified using blue–white selection as described by Sambrook *et al.* (1989).

2.6. DNA sequencing

Sequencing reactions were carried out commercially by Co:Genics (Essex, UK) or by Eurofins MWG Operon (Ebersberg, Germany). Both commercial sequencing companies use the dideoxy chain termination method of Sanger *et al.* (1992), using automated Applied Biosystems ABI 3739*xl* DNA analyzers (Foster City, CA) and dye-labelled terminators. Primers used for amplification of DNA fragments were also used for direct DNA sequencing.

2.7. Susceptibility testing

The MICs for 5FC and 5FU (Sigma-Aldrich Ltd.) were determined according to a modification of the broth microdilution method of the Clinical and Laboratory Standards Institute (CLSI; formerly the National Committee for Clinical Laboratory Standards) document M27-A2 (Clinical and Laboratory Standards Institute, 2002). Assays were carried out in sterile 96-well microtitre plates (Corning, NY, USA). The medium used was RPMI-1640 (Sigma-Aldrich Ltd.) with added 2% (w/v) glucose, and buffered with 0.165 M morpholinepropanesulphonic acid (MOPS) (Sigma-Aldrich Ltd.). The pH was adjusted to 7 using 10 M NaOH. Wells 1 and 12 of each row were dispensed with 100 μl of RPMI-1640-2% (w/v) glucose for use as sterility and growth controls, respectively. In wells 2 to 11, drugs were diluted 1:2 in RPMI-1640-2% (w/v) glucose from a final concentration of 256 μg/ml to 0.5 μg/ml in a final volume of 100 μl. Inocula were prepared from 24 h-old PDA agar plates by suspension of a colony in 2 ml of molecular grade Milli-Q® Biocel-purified water (resistivity 18.2 MΩ/cm) (Millipore™) and adjustment of the cell density to 2–4 × 10^6 cfu/ml using a haemocytometer (Neubauer bright line haemocytometer, Hausser Scientific, Horsham, USA). Inocula were then diluted 1:1000 in RPMI-1640-2% (w/v) glucose and 100 μl was added to each well from columns 2 to 12 in order to achieve a final drug concentration range of 128 μg/ml to 0.25 μg/ml in a final volume of 200 μl.

Microtitre plates were wrapped in aluminium foil and were incubated at 37°C for 48 h. End points were determined spectrophotometrically by measuring the turbidity in each well at 540 nm with an automated plate reader (Tecan GENios multidetection microplate reader; Theale, Reading, UK). An endpoint (MIC_{50}) was determined as the lowest drug concentration which fulfilled the criterion $\%T \geq \%\ Tk + 0.5(100 - \%Tk)$, where T is the transmission from a microtitre plate well containing the drug and the organism being tested, and Tk is the transmission from the drug-free growth control well containing the organism being tested (Galgiani & Stevens, 1976). The MIC_{50} represents the drug concentration that inhibited growth by 50%, as determined by transmission, compared with the growth of controls as described previously (Galgiani & Stevens, 1976). All isolates and transformant derivatives were tested in duplicate and on two separate occasions.

Chapter 3

A study of the prevalence of *Candida dubliniensis* in a normal healthy and in an immunocompromised human population, and recovery of *C. dubliniensis* from avian-excrement-associated sources

3.1. Introduction

The accurate identification of *C. dubliniensis* has been hampered in the past due to its close phenotypic similarity with *C. albicans*, the most commonly isolated *Candida* species. Both species produce germ-tubes and chlamydospores, the most commonly used traits to discriminate these species from the rest of the *Candida* genus. *Candida albicans* and *C. dubliniensis* are often recovered in co-culture (Coleman *et al.*, 1997b; Sullivan *et al.*, 1997, 1999; Pontón *et al.*, 2000). The phenotypic and genotypic similarities that exist between the two species are the most probable reasons for the delayed identification of *C. dubliniensis* as a separate species in 1995, and also made early prevalence and epidemiology studies problematic. Retrospective identification studies have reported that at least 1.2% of isolates recovered in a clinical setting that were previously identified as *C. albicans* were in fact *C. dubliniensis* (Coleman *et al.*, 1997b; Odds *et al.*, 1998; Jabra-Rizk *et al.*, 2000).

3.1.1. Discrimination of *C. dubliniensis* from *C. albicans*

The most commonly used culture medium for the identification of different *Candida* species is CHROMagar Candida™ medium (Odds & Bernaerts, 1994). Following incubation at 37°C for 48 h, *C. dubliniensis* grows as green colonies that are typically darker in colour and smaller than those formed by *C. albicans*. However, paler green colonies can also be formed by *C. dubliniensis* on this agar following storage and subculture (Schoofs *et al.*, 1997; Tintelnot *et al.*, 2000). Although it is highly useful for the presumptive identification of a several medically significant *Candida* species, it is not reliable enough to be used as a sole method of differentiation between *C. albicans* and *C. dubliniensis*.

The analysis of carbohydrate assimilation profiles provided by commercial yeast identification kits such as the API ID32C system (bioMérieux) is one reliable method to distinguish between the closely related *C. albicans* and *C. dubliniensis* species. This kit consists of a strip of 29 different carbon sources, four of which *C. albicans* and *C. dubliniensis* differ in their abilities to assimilate. When used according to the manufacturers' instructions, the kit offers a high level of discrimination between the two species. *Candida albicans* assimilates lactate, xylose, trehalose and methyl-D-glucose, whereas these substrates are not assimilated by *C. dubliniensis* following 48 h

incubation at 37°C, although it can assimilate these substrates following prolonged incubation (Pincus *et al.*, 1999).

A number of differential agar media based on aqueous extracts of specific plant seeds or leaves have been developed to facilitate discrimination between *C. albicans* and *C. dubliniensis*. One such medium known as Staib agar or birdseed agar, is based on an extract of *Guizotia abyssinica* seeds (Staib & Morschhäuser, 1999). Another such medium, tobacco agar, is based on an extract of tobacco leaves (Khan *et al.*, 2004), and a third such agar based on an extract of sunflower seeds (Al Mosaid *et al.*, 2003) is known as Pal's agar. Following 48 h incubation at 30°C on these plant-extract-based agars, *C. dubliniensis* typically grows as rough colonies that produce a hyphal fringe containing abundant chlamydospores, in contrast to *C. albicans*, which typically grows as smooth, non-hyphal colonies. Pal's agar has been shown to be the most reliable of these media for the differentiation of *C. albicans* and *C. dubliniensis* (Al Mosaid *et al.*, 2003). This medium contains an extract of sunflower seeds (including kernel and shells) and distinguishes between the two species on the basis of the production of a hyphal fringe by *C. dubliniensis* after 48 h incubation at 30°C. This hyphal fringe is composed of abundant hyphae, pseudohyphae, and chlamydospores. A previous study showed that 100% of 128 *C. dubliniensis* isolates produced a hyphal fringe on this agar, in contrast to 0% of 124 *C. albicans* isolates, which grew as smooth colonies (Al Mosaid *et al.*, 2003). Mosca *et al.* (2003) demonstrated that the milk-based Casein agar can be used to distinguish *C. albicans* from *C. dubliniensis* isolates according to the production of chlamydospores following 48 h incubation at 24°C. All *C. dubliniensis* isolates produced chlamydospores on this medium, in contrast to *C. albicans* of which only 9/120 isolates produced chlamydospores.

Definitive identification of *C. dubliniensis* is commonly achieved using PCR-based methods. The most extensively used PCR-based method specifically amplifies the *ACT1*-associated intron of *C. dubliniensis* (Donnelly *et al.*, 1999).

3.1.2. Prevalence of *C. dubliniensis*

When *C. dubliniensis* was first identified, it was mainly associated with HIV-infected individuals, many of whom were intravenous drug users, as well as AIDS patients (Sullivan *et al.*, 1995). The species has been recovered from 32% of AIDS patients presenting with oral candidiasis, and from 25% of asymptomatic AIDS patients (Coleman *et al.*, 1997b; Pontón *et al.*, 2000) in Ireland. *Candida dubliniensis* has been

shown to predate the AIDS epidemic (Sullivan *et al.*, 1995; Odds *et al.*, 1998; Meis *et al.*, 1999) and has also been recovered from individuals who are HIV-negative; however it is more often recovered from HIV-negative individuals with other underlying conditions, rather than from the normal healthy population.

The prevalence of *C. dubliniensis* has been studied in a variety of immunocompromised patient cohorts. It has been recovered from cancer patients with candidaemia (Sebti *et al.*, 2001) as well as from patients who were colonised with the opportunistic pathogen (Meis *et al.*, 1999; Sebti *et al.*, 2001). It has been isolated from 11.1% of patients with cystic fibrosis (Peltroche-Llacsahuanga *et al.*, 2002), and from 14% of insulin-using diabetes mellitus patients (Willis *et al.*, 2000). Coleman *et al.* (1997b) reported that 14.6% of HIV-negative individuals with denture-associated candidiasis harboured *C. dubliniensis*, and Polacheck *et al.* (2000) recovered five isolates from HIV-negative patients. Carr *et al.* (2005) reported *C. dubliniensis* as the cause of a case of endocarditis in a HIV-negative intravenous drug user, who died as a result.

Candida dubliniensis has also been recovered from normal healthy individuals, though its prevalence has been reported to be substantially lower than in the immunocompromised population. It is usually recovered from 3–5% of the normal healthy population (Coleman *et al.*, 1997b; Pontón *et al.*, 2000; Montour *et al.*, 2003). In contrast, a study carried out in South Africa (Blignaut *et al.*, 2003) yielded data that varied significantly from this trend. The authors reported the presence of *C. dubliniensis* in the oral cavities of 16% of the normal healthy white population and in 0% of the normal healthy black population, whereas in the HIV-positive population it was present in the oral cavities of 1.5% of black individuals and 9% of white individuals. These findings indicated a higher prevalence of the species in the healthy population than in the HIV-infected population (Blignaut *et al.*, 2003). It is worth noting that no other study to date has produced data to support these findings. *Candida dubliniensis* is able to cause disease in healthy individuals and has been identified as the cause of denture stomatitis in a normal healthy teenager (Mosca *et al.*, 2005).

Candida dubliniensis is most commonly recovered from the oral cavity. However, it has also been recovered from sputum, faeces, urine, blood, vaginal swabs, and wounds of HIV-infected individuals as well as of HIV-negative individuals (Odds *et al.*, 1998; Meis *et al.*, 1999; Jabra-Rizk *et al.*, 2000; Gee *et al.*, 2002; Peltroche-Llacsahuanga *et al.*, 2002).

3.1.3. Geographic distribution of *C. dubliniensis*

The species appears to be globally distributed (Fig. 3.1) and has been recovered from many European countries, Saudi Arabia, Egypt, Israel, Kuwait, South Africa, U.S.A., Canada, Brazil, Colombia, Chile, Venezuela, Argentina, Japan, Thailand, India, Iceland and Australia (Pujol *et al.*, 1997; Schoofs *et al.*, 1997; Sullivan *et al.*, 1997; Odds *et al.*, 1998; Pinjon *et al.*, 1998; Salkin *et al.*, 1998; Jabra-Rizk *et al.*, 1999; Polacheck *et al.*, 2000; Pontón *et al.*, 2000; Redding *et al.*, 2001; Gee *et al.*, 2002; Blignaut *et al.*, 2003; Fotedar & Al-Hedaithy, 2003; Montour *et al.*, 2003; Ahmad *et al.*, 2004; Bujdáková *et al.*, 2004; Al Mosaid *et al.*, 2005; Melkusová *et al.*, 2005; Brito Gamboa *et al.*, 2006 Asmundsdóttir *et al.*, 2008; Alvarez *et al.*, 2009).

Previous studies identified the presence of four different genotypes in the *C. dubliniensis* population. These genotypes are based on the genetic sequences of the ITS region of the rRNA gene cluster, and are known as ITS genotypes 1–4 (Gee *et al.*, 2002). To date the majority of *C. dubliniensis* isolates examined have been designated as ITS genotype 1. This group contains isolates from Europe, Saudi Arabia, Egypt, Israel, India, Canada, Australia, Argentina, Japan and the U.S.A. (Gee *et al.*, 2002; Al Mosaid *et al.*, 2005). Isolates belonging to the ITS genotype 2 group have been recovered in Europe, Canada and Israel (Gee *et al.*, 2002; Al Mosaid *et al.*, 2005). Isolates of ITS genotypes 3 and 4 have been most commonly recovered from countries in the Middle East to date (Gee *et al.*, 2002; Al Mosaid *et al.*, 2005), suggesting the possibility of geographical enrichment of such isolates in this region.

3.1.4. Recovery of *C. dubliniensis* from environmental sources

In 2007 the recovery of *C. dubliniensis* from environmental sources was reported for the first time (Nunn *et al.*, 2007). The study reported the high prevalence of *C. dubliniensis* on the surface of *Ixodes uriae* (ticks) on the Great Saltee Island off the South-East coast of Ireland. These ticks were living in cracks filled with common guillemot excrement, and the authors hypothesised that the supposed source of the isolates was excrement from the seabirds present on the island. All of these isolates were reported as belonging to ITS genotype 1. The authors reported the recovery of *C. dubliniensis* from 13/59 (22%) tick samples recovered at one cliff (Labour-in-Vain, LIV-1), and from 9/43 (21%) tick samples taken at the same cliff on a separate day (LIV-2) (Nunn *et al.*, 2007).

Although this study (Nunn *et al.*, 2007) was the first to report the recovery of *C. dubliniensis* from non-human sources, the closely related species *C. albicans* has previously been associated with many different types of animals including birds, primates, reptiles, amphibians, as well as both wild and domesticated animals (Odds, 1988; Buck, 1990; Pressler *et al.*, 2003; Bougnoux *et al.*, 2004; Edelmann *et al.*, 2005; Tavanti *et al.*, 2005a; Cafarchia *et al.*, 2006; Jacobsen *et al.*, 2008; Wrobel *et al.*, 2008). The presence of *C. dubliniensis* on the surface of ticks found at a location with abundant seabird excrement, and the recovery of the closely related *C. albicans* from a broad range of animal sources, suggested that a separate niche(s) may exist for *C. dubliniensis* in non-human hosts.

3.1.5. Objectives

There were three main aims to this part of the study:

- To investigate the prevalence of *C. dubliniensis* in a normal healthy population of students, and in an immunocompromised population of individuals with cystic fibrosis, using the methods of identification outlined above.
- To examine the prevalence of *C. dubliniensis* from avian sources, in order to determine if a natural reservoir for *C. dubliniensis* exists in birds.
- To determine the prevalence of each ITS genotype amongst the recovered isolates from the normal healthy human population, the immunocompromised human population, and from avian-excrement-associated sources.

Figure 3.1. Disparate geographical locations from where *C. dubliniensis* isolates have been recovered to date.

Countries from which *C. dubliniensis* isolates have been recovered from humans and included in the present study are indicated with blue labels. The avian-excrement-associated *C. dubliniensis* isolates recovered in Ireland are represented by a red label. Countries from which isolates of *C. dubliniensis* were recovered from humans and were not included in the present study are indicated with green labels. Where *C. dubliniensis* isolates were recovered in a particular country, not all specific locations from which isolates were recovered in that country are shown.

3.2. Materials and methods

3.2.1. Prevalence of *C. dubliniensis* in the normal healthy population, and in the immunocompromised population

3.2.1.1. Normal healthy population of students attending Trinity College Dublin, Ireland

A total of 24 healthy students, each of whom were in their early twenties anonymously submitted swab samples from their oral cavities. The Nitrogen-gassed VI-PAK sterile swabs (Sarstedt-Drinagh, Wexford, Ireland) swabs were used to sample the dorsum of the tongue, and were plated immediately onto CHROMagar Candida™ medium in order to preliminarily screen for the dark-green colonies typical of *C. dubliniensis*. Swabs were also were incubated in YPD broth containing 20 µg chloramphenicol/ml (Sigma-Aldrich Ltd.) at 37°C overnight with shaking at 200 rpm and DNA was extracted from all cells present using the bead-beating method (Chapter 2, Section 2.3.2) in order to be screened for the presence of *C. dubliniensis* by PCR (Donnelly *et al.*, 1999). Oral rinses were collected anonymously from 263 students, none of whom wore dentures or were using antibiotics. Oral rinses were collected using 10 ml of sterile distilled water (SDW), of which 500 µl was plated out directly onto CHROMagar Candida™ medium. Suspect *C. dubliniensis* colonies were tentatively identified according to colour and template DNA was prepared from these colonies following sub-culture using the boiling method as outlined in Chapter 2, Section 2.3.1 and identification was confirmed by PCR (Donnelly *et al.*, 1999).

3.2.1.2. Immunocompromised population of patients with cystic fibrosis in Dublin, Ireland

Forty-two patients with cystic fibrosis who were attending the Adelaide and Meath Hospital incorporating the National Children's Hospital in Dublin underwent bronchial lavage over a two month period as part of their routine clinical management, providing 98 samples in total. The patient group sampled consisted of 26 males and 16 females, all of whom were less than 20 years of age (Table 3.1). All patients from whom samples were provided for use in this study remained anonymous, only gender and age information was made available for the purposes of the present study. The sputum samples were plated directly onto CHROMagar Candida™ medium in 100 µl

aliquots in order to screen for the dark-green colonies typical of *C. dubliniensis*. A further 100 µl was incubated in YPD broth containing 20 µg chloramphenicol/ml at 37°C overnight with shaking at 200 rpm in order to be screened for the presence of *C. dubliniensis* DNA by PCR (Donnelly *et al.*, 1999).

3.2.1.3. Oral Candida *carriers attending the Dublin Dental Hospital*

A total of 138 patients attending the Dublin Dental Hospital suffering from suspected oral candidiasis were sampled from the oral cavity using Nitrogen-gassed VI-PAK sterile swabs (Sarstedt-Drinagh, Wexford, Ireland) as part of routine clinical management, in order to diagnose candidiasis. All patients from whom samples were collected gave informed consent and remained anonymous. Only gender and age information was made available for the purposes of the present study. Swabs were plated immediately onto CHROMagar Candida™ medium in order to determine the *Candida* species present in the oral cavity of the patient. Swabs were also were incubated in YPD broth containing 20 µg chloramphenicol/ml at 37°C overnight with shaking at 200 rpm in order to be screened for the presence of *C. dubliniensis* by PCR (Donnelly *et al.*, 1999). In total 168 swabs were collected from 138 patients, and each was assessed for the presence of *C. dubliniensis* using both culture- and PCR-based methods.

3.2.2. Prevalence of *C. dubliniensis* in the avian population

Fresh seabird excrement was collected from damp sand using Nitrogen-gassed VI-PAK sterile swabs (Sarstedt-Drinagh, Wexford, Ireland) and was plated on CHROMagar Candida™ medium within 2 h of collection. In total, 124 faecal samples were collected from a number of beaches along the east coast of Dublin, Ireland on separate occasions. Ten samples were taken at Dún Laoghaire Harbour, 72 samples were taken at Sandymount Strand, and 25 samples were taken at Bray Harbour. All three of these sites are located on the south side of Dublin. Seventeen samples were taken at Claremount Strand on the northside of Dublin. Ten samples were also taken from seabird excrement on the campus of Trinity College Dublin, in the city centre. Birds present at these sites at the time of sampling included *Larus argentatus* (herring gulls) and *L. ridibundus* (black-headed gulls). *Cygnus olor* (swans) were also present at the Bray harbour site. In total, 134 swabs of fresh seabird excrement were examined for the presence of *C. dubliniensis*.

Table 3.1 Prevalence of *C. dubliniensis* in normal healthy and immunocompromised populations

	Total number of individuals sampled [a]	Age range of individuals sampled [b]	Individuals from whom *C. dubliniensis* carriage was identified	Age range of *C. dubliniensis* carriers
		Trinity College Dublin		
	287	17 – 25	9/287 (3.1%)	17 – 25
Female	*N/A* [a]	*17 – 25*	*6/9 (66.7%)*	*17 – 25*
Male	*N/A* [a]	*17 – 25*	*3/9 (33.3%)*	*17 – 25*
		Adelaide and Meath Hospital incorporating the National Children's Hospital in Dublin		
	42	1 – 19 39/42 over 10 years old	10/42 (23.8%)	1 – 19 9/10 (90.0%) over 10 years old
Female	*26*	*23 over 10 years old*	*7/26 (26.9%)*	*1– 19* *6/7 (85.7%) over 10 years old*
Male	*16*	*16 over 10 years old*	*3/16 (18.8%)*	*18 – 19* *3/3 (100.0%) over 10 years old*
		Dublin Dental Hospital		
	138	22 – 97 (66 individuals) [b] 38/66 over 60 years old	24/138 (17.4%)	22 – 78 11/24 (45.8%) over 60 years
Female	*85*	*28/66 over 60 years old*	*15/85 (17.6%)*	*22 – 78* *8/15 (53.3%) over 60 years*
Male	*53*	*10/66 over 60 years old*	*9/53 (17.0%)*	*25 – 74* *3/9 (33.3%) over 60 years*

[a] Gender data was not available for all students sampled in the present study.
[b] Age data was only available for 66 individuals examined from the Dublin Dental Hospital. However, age data for all *C. dubliniensis* carriers identified in the Dublin Dental Hospital is included.

3.2.3. Culture-based identification of *C. dubliniensis*

Oral swabs, oral rinse, sputum and avian faecal samples were plated on CHROMagar Candida™ medium and were incubated at 37°C for 48 h. This was used to tentatively identify the *Candida* species present in the samples according to colony colour and morphology as described in Chapter 2, Section 2.2.1.

3.2.4. Identification of *C. dubliniensis* by PCR amplification and recovery of isolates

DNA was extracted from putative *C. dubliniensis* colonies recovered on CHROMagar Candida™ medium following subculture on fresh PDA agar and using the boiling method as described in Chapter 2, Section 2.3.1. Definitive identification of the isolate was confirmed by PCR as described in Chapter 2, Section 2.4.1.

Swabs of oral and sputum samples were incubated in YPD broth containing 20 µg chloramphenicol/ml at 37°C overnight with shaking at 200 rpm. The bead-beating method was used (as described in Chapter 2, Section 2.3.2) to extract the DNA from broth cultures. The presence or absence of *C. dubliniensis* DNA in the extracted culture DNA was determined by PCR as described in Chapter 2, Section 2.4.1.

Broth cultures that were PCR-positive for the presence of *C. dubliniensis* DNA were diluted 1:100 in sterile distilled water. A 100 µl volume of the diluted culture was plated out on Pal's agar (Al Mosaid *et al.*, 2003) in order to recover *C. dubliniensis* from the culture. *Candida dubliniensis* isolates were identified on the basis of hyphal fringe production on Pal's agar (Al Mosaid *et al.*, 2003) as described in Chapter 2, Section 2.2.2, and the identities of the recovered isolates were confirmed by PCR (Donnelly *et al.*, 1999).

3.2.5. Determination of *C. dubliniensis* ITS genotypes

Template DNA from *C. dubliniensis* isolates was prepared using the bead-beating method as described in Chapter 2, Section 2.3.2, and ITS genotypes were assigned to all isolates by PCR testing (Gee *et al.*, 2002) as described in Chapter 2, Section 2.4.2.

3.2.6. Identification of other *Candida* species

Oral swab samples, sputum and avian faecal samples were plated on CHROMagar Candida™ medium and were incubated at 37°C for 48 h. This was used to

tentatively identify the *Candida* species present in the samples as described in Chapter 2, Section 2.2.1. Definitive identification of other *Candida* species was undertaken using the API ID32C yeast identification system, based on carbohydrate assimilation patterns as described in Chapter 2, Section 2.2.3.

3.3. Results

3.3.1. Prevalence of *C. dubliniensis* in a normal healthy population and in an immunocompromised population

3.3.1.1. Normal healthy population of students attending Trinity College Dublin, Ireland

A total of 24 healthy students, all of whom were in their early twenties, voluntarily submitted swabs of their oral cavities. These swabs were used to inoculate CHROMagar Candida™ plates, which were examined for the presence of typical *Candida* colonies following incubation at 37°C for 48 h. Of the 24 plates examined, 10 plates contained green colonies that are typical of *C. albicans* or *C. dubliniensis*. These colonies were sub-cultured on PDA for 24 h and the boiling method was used (as described in Chapter 2, Section 2.3.1) to extract DNA from these isolates. These DNA samples were tested in separate PCR amplifications with the *ACT1* gene primers specific for *C. dubliniensis* (Donnelly *et al.*, 1999). No amplimers were obtained with these primers, and all *Candida* species present in the 10/24 (41.7%) swabs were tentatively identified as *C. albicans*. This was confirmed in a number of cases using the API ID32C system as described in Chapter 2, Section 2.2.3. Oral swabs were also used to inoculate YPD broth cultures supplemented with 20 μg chloramphenicol/ml, which were incubated at 37°C overnight. The bead-beating method was used (as described in Chapter 2, Section 2.3.2) to extract DNA from the fungal cells present in the overnight culture. This DNA was examined using the same *ACT1* gene primers specific for *C. dubliniensis* (Donnelly *et al.*, 1999). Interestingly, two of the swabs that did not yield any *Candida* growth on CHROMagar showed amplification with the *ACT1* gene *C. dubliniensis*-specific primers following broth enrichment, showing the presence of *C. dubliniensis* in 2/24 (8.3%) of the subjects tested. These *C. dubliniensis* isolates were recovered from the YPD broth cultures using Pal's agar (Al Mosaid *et al.*, 2003) as described in Section 3.2.4.

Oral rinse samples were collected from 263 students (not including the 24 students detailed above) and were plated directly onto CHROMagar Candida™ medium in order to screen for *C. dubliniensis* colonies. Of these samples, 81/263 (30.8%) yielded *Candida* isolates, and 68/263 (25.9%) samples yielded green colonies that are typical of *C. albicans* or *C. dubliniensis*. The isolates yielding green colonies were

tested with the *ACT1* primers by PCR as previously described. *Candida dubliniensis*-specific amplimers were obtained from 7/68 (10.3%) of these green *Candida* colonies. *Candida albicans* isolates were identified using the API ID32C system (as described in Chapter 2, Section 2.2.3) from the remaining 61/68 (89.7%) green *Candida* colonies. Other *Candida* species were also isolated from the student population but were not further characterised in the present study.

In total 287 students were examined using either the oral rinse method or swabs of the oral cavity (Table 3.1). *Candida dubliniensis* isolates were recovered from six female and three male students. The exact ages of the students examined in the present study were unavailable. However, all students examined were aged between 17 and 25 years (Table 3.1). Of these 287 students, 91 (31.7%) were found to harbour oral *Candida* species, of whom 71 (24.7%) harboured *C. albicans* and nine (3.1%) harboured *C. dubliniensis*. Seven of these *C. dubliniensis* isolates (77.8%) were confirmed as ITS genotype 1, one (11.1%) as ITS genotype 2, and one (11.1%) as ITS genotype 3 by PCR amplification using ITS-genotype specific primer pairs (Gee *et al.*, 2002).

3.3.1.2. Population of patients with cystic fibrosis in Dublin, Ireland

Sputum samples were collected from 42 distinct patients suffering from cystic fibrosis and undergoing routine broncheolar lavage over a two-month period. A total of 98 sputum samples were collected from these patients. One individual patient underwent bronchial lavage on seven occasions, thus providing seven sputum samples. Two patients underwent bronchial lavage on six occasions, thus providing six sputum samples each, and 23 patients underwent bronchial lavage on between two and five occasions, providing between two and five sputum samples each, respectively. Of the 42 patients, 16 underwent bronchial lavage on one occasion only.

Of these samples, 18/98 (18.4%) were PCR-positive using *C. dubliniensis*-specific primers (Donnelly *et al.*, 1999) following overnight incubation of the samples in YPD broth supplemented with 20 µg chloramphenicol/ml as described in Section 3.2.2. All of the samples that yielded *C. dubliniensis*-specific PCR products also formed dark-green colonies on CHROMagar Candida™ medium.

Of the 42 patients who provided sputum samples, 10 (23.8%) were found to harbour *C. dubliniensis* in the respiratory tract (Table 3.1). Seven of the *C. dubliniensis*-carriers were female, and three were male (Table 3.1). Four (11.9%) of the 10 *C.*

dubliniensis-carrying patients harboured *C. dubliniensis* of the same ITS genotype in their respiratory tracts persistently, of whom 3 carriers were female (aged 1-19 years), and one was male (aged 18 years). One male patient (aged 19 years) yielded *C. dubliniensis* isolates of both ITS genotypes 1 and 2 on separate occasions. Isolates of *C. dubliniensis* ITS genotype 1 were recovered from this individual on two occasions which were four days apart, whereas on a third occasion a week later only *C. dubliniensis* ITS genotype 2 isolates were recovered. The 18 *C. dubliniensis* isolates recovered from these 10 patients consisted of ITS genotype 1 (66.7%), ITS genotype 2 (22.2%), and ITS genotype 3 (11.1%).

Candida albicans was the most frequently recovered species from sputa on CHROMagar Candida™ (Fig. 3.2A) and was detected in 51/98 (52%) samples recovered from 26/42 (61.9%) patients included in the study. Other *Candida* species were also isolated from the student population but were not further characterised in the present study. However, these latter samples were all taken at different time intervals from the same two patients. Mycelial fungi were also frequently recovered from sputum samples (Fig. 3.2A).

3.3.1.3. Oral Candida *carriers attending the Dublin Dental Hospital*

Sterile swabs were used to sample the oral cavities of patients attending the Dublin Dental Hospital with suspected oral candidiasis; 168 swabs were collected from 138 patients in total. Each swab was examined for the presence of *C. dubliniensis* using both culture- and PCR-based methods as previously described in Sections 3.2.3 and 3.2.4, respectively. *Candida dubliniensis* was recovered from 24/138 (17.4%) patients, 15 were female and nine were male (Table 3.1). Clinical information was limited for the patients in this study. However, of the 24 patients from whom *C. dubliniensis* was isolated, 8 were known to be immunocompromised due to HIV infection, autoimmune diseases such as Sjögren's syndrome or systemic inflammatory rheumatic disease (CREST syndrome), or were taking systemic steroids. Another patient from whom *C. dubliniensis* was recovered had diabetes mellitus.

Two of the samples from which *C. dubliniensis* isolates was recovered did not yield any *Candida* isolates following culture on CHROMagar Candida™ medium. In these two cases, *C. dubliniensis* DNA was detected by PCR. *Candida dubliniensis* was subsequently recovered from these swabs following broth enrichment as described in Section 3.2.4. Of the 24 *C. dubliniensis* isolates recovered from oral swab samples, 20

(83.3%) belonged to ITS genotype 1, and the remaining four (16.7%) belonged to ITS genotype 2. Other *Candida* species were also isolated from the student population but were not further characterised in the present study.

3.3.2. Prevalence of *C. dubliniensis* in the avian population

Fresh faecal deposits left by feeding seabirds were recovered from damp sand on sterile swabs at a number of sampling sites on the east coast of Dublin, Ireland, in an effort to assess the prevalence of *C. dubliniensis* in the avian population. In total, 134 swab samples were taken and plated on CHROMagar Candida™ medium within 2 h of collection. Isolates were tentatively identified according to colony colour on CHROMagar Candida™ medium, and identities were confirmed using the API ID32C yeast identification system.

C. albicans was the *Candida* species most frequently recovered and was isolated from 24/134 (17.9%) of the samples tested. Nine of the 24 (37.5%) and 7/24 (29.2%) *C. albicans* isolates were recovered from seabirds at Dún Laoghaire Harbour and Trinity College Dublin, respectively. The remaining eight *C. albicans* isolates (33.3%) were recovered from birds feeding at the other three sample sites collectively.

Candida dubliniensis was recovered from 3/134 (2.2%) of the samples tested. These three isolates were all recovered from seabirds at Trinity College Dublin and all were identified as belonging to ITS genotype 1 by PCR (Gee *et al.*, 2002).

Other yeast species including *C. glabrata, C. parapsilosis, C. tropicalis, Candida zeylanoides, Cr. neoformans, R. glutinis* were also frequently recovered from faecal samples, as were mycelial fungi (Fig. 3.2).

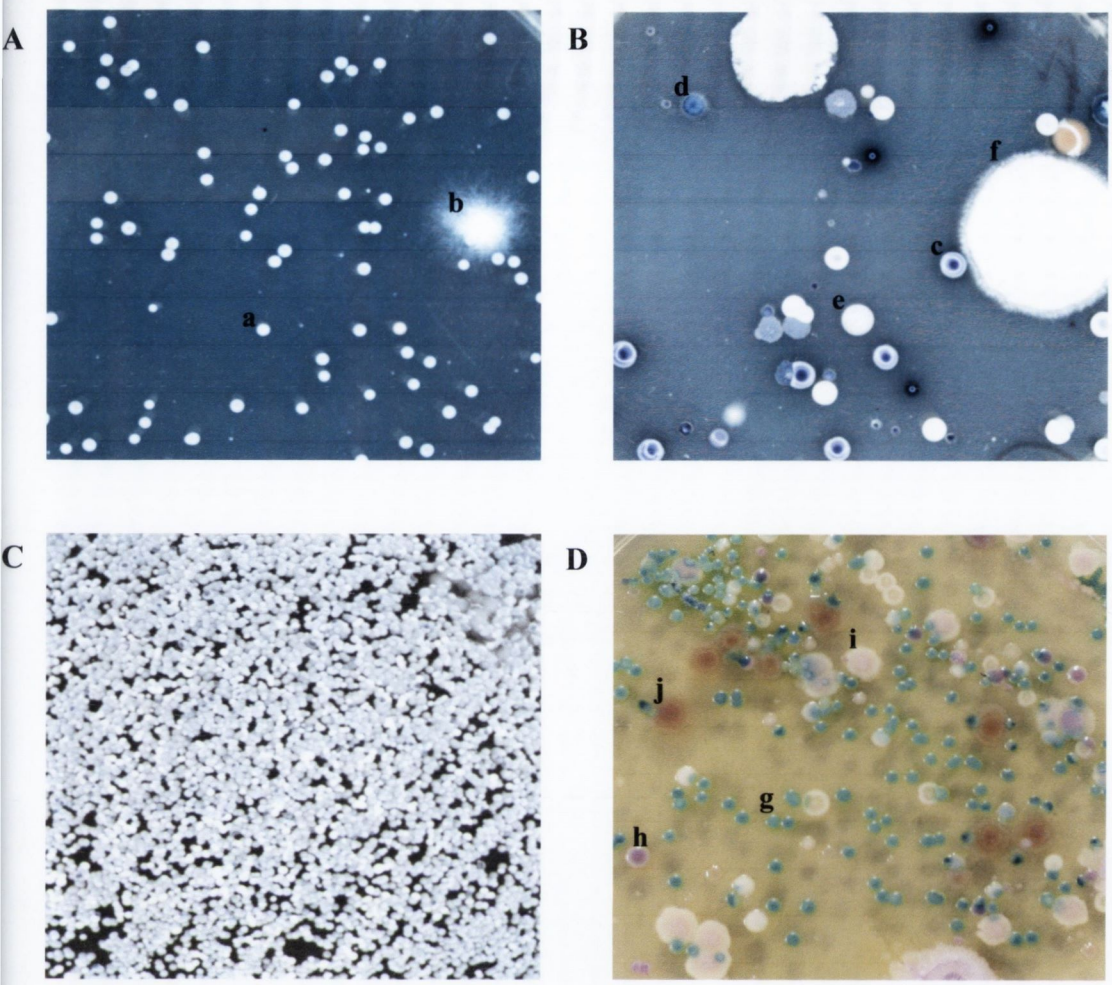


Figure 3.2. Yeast species frequently recovered on CHROMagar Candida™ medium from human sputum samples and avian-excrement samples.

Panel A; sample of yeast species recovered from the sputum of a patient with cystic fibrosis; (a) pale green *C. albicans* colony and (b) a white hyphal colony formed by an aerial dermatophytic fungus. Panel B; sample of yeast species recovered from the faecal swab of a seabird; (c) a purple *C. tropicalis* colony (d) a blue *C. zeylanoides* colony (e) a white *C. parapsilosis* colony and (f) aerial dermatophytic fungal colony. Panel C; sample of yeast species recovered from the faecal swab of a seabird; displaying confluent growth of *C. parapsilosis* which forms white colonies and *C. glabrata* which forms pale pink colonies. Panel D; sample of yeast species recovered from the faecal swab of a seabird displaying the most common species recovered from avian excrement samples; (g) *C. albicans* (green) (h) *C. tropicalis* (purple) (i) *C. krusei* (pale pink with spreading edges), and (j) *Rhodotorula glutinis* (red).

3.4. Discussion

Most previous *C. dubliniensis* prevalence studies have relied solely on culture-based methods. The results of the present study demonstrate that reliance on these methods alone underestimates the true prevalence of this species, and that PCR-based analysis of culture-enriched samples is the most accurate method of prevalence determination. The main objective of the present study was to accurately determine the prevalence of *C. dubliniensis* in the normal healthy, and immunocompromised human population as well as in avian-excrement-associated samples using a combination of culture-based and PCR-based methods, including growth on CHROMagar Candida™ medium or Pal's agar (Al Mosaid *et al.*, 2003), carbohydrate assimilation pattern analysis, and *ACT1*-directed PCR analysis.

Previous culture-based studies reported the prevalence of *C. dubliniensis* in the oral cavities of the normal healthy population in the range 3-5% (Coleman *et al.*, 1997b; Pontón *et al.*, 2000; Montour *et al.*, 2003). In the present study, the use of oral rinse sampling and culture-based methods as a sole method of prevalence determination resulted in the recovery of *C. dubliniensis* from the oral cavities of 2.7% of the normal healthy population, in agreement with previous studies (Pontón *et al.*, 2000; Montour *et al.*, 2003).

Patients suffering from cystic fibrosis are highly susceptible to secondary colonisation of lung tissues by bacteria and fungi, due to dehydration of respiratory secretions and defective mucociliary clearance. A previous study examined the sputa of 54 patients with cystic fibrosis and reported that yeasts (mainly *C. albicans*) were observed in sputum samples of 36/54 (66.7%) patients (Peltroche-Llacsahuanga *et al.*, 2002). The detection of *C. dubliniensis* in 11.1% of the patients studied was also reported. In the present study, *C. dubliniensis* was recovered from 10/42 (24%) patients examined. In 5/42 (12%) cases, *C. dubliniensis* carriage was found to be persistent in the respiratory tract, being consistently recovered from sequential samples taken at different time intervals over several weeks. The presence of *C. albicans* was identified in the sputa of 62% of the patients studied. Although the number of patients suffering from cystic fibrosis were higher in the previous study (n=54) (Peltroche-Llacsahuanga *et al.*, 2002), a similar prevalence of both *C. albicans* and *C. dubliniensis* was determined in the present study. All of the *C. dubliniensis*-specific PCR-positive sputum

samples were also culture-positive for *C. dubliniensis,* which suggests that data from the present study is probably an accurate representation of the true prevalence of *C. dubliniensis* in this patient cohort.

The higher prevalence (17.4%) of oral *C. dubliniensis* amongst patients attending the Dublin Dental Hospital is perhaps unsurprising, given that 9/24 patients from whom *C. dubliniensis* was recovered were either immunocompromised or suffering from conditions such as diabetes mellitus, and all patients had suspected oral candidiasis. A previous study (Lockhart *et al.*, 1999) showed that the frequency of *C. albicans* and *C. glabrata* carriage increases with age, and the study showed that this increase in colonisation occurred independently of denture use. The authors suggested that the increased rate of carriage may be due to differences in the relative levels of yeast inhibitors such as histatins and defensins, or a reduced rate of salivary flow in the oral cavities of the elderly (Lockhart *et al.*, 1999). Age data was available for the 24 patients from whom *C. dubliniensis* was recovered, almost half (11/24 patients) were over 60 (Table 3.1). *Candida dubliniensis* carriage was not significantly associated ($p = 0.25$) with increased age (over 60 years) in the study of patients attending the Dublin Dental Hospital.

Candida dubliniensis isolates were recovered from female individuals more frequently from all three populations examined in the present study (Table 3.1). A two-tailed Student's *t*-test showed that the prevalence did not differ significantly between the two sexes, but was suggestive of a trend ($p = 0.04$).

A previous study reported the recovery of *C. dubliniensis* from 21–22% of *I.uriae* (tick) samples obtained from avian-excrement-filled cracks on the cliff-face at two separate locations on the Great Saltee Island (Nunn *et al.*, 2007). The authors hypothesised that the supposed source of these isolates was the excrement of seabirds in which these ticks were living. The present study examined 134 faecal samples from seabirds, and found the prevalence of *C. dubliniensis* to be very low in comparison to that observed on the surface of ticks in the previous study (Nunn *et al.*, 2007). *Candida dubliniensis* was recovered from only 3/134 (2.2%) samples, all three of which were recovered from bird excrement on the campus of Trinity College in Dublin, Ireland. No *C. dubliniensis* isolates were recovered from faecal samples obtained at seashores, although *C. albicans* isolates were obtained at these sites. It has been previously shown that *C. dubliniensis* is less well able to tolerate osmotic stress than *C. albicans* (Enjalbert *et al.*, 2009); this may explain why *C. dubliniensis* was more frequently

recovered at sites such as a college campus and cracks in the cliff-faces of the Great Saltee Island. From the present study it appears that isolates of *C. dubliniensis* are more abundant in the excrement of seabirds from areas that are heavily populated by humans, as all avian-excrement-associated *C. dubliniensis* isolates recovered in the present study were recovered from birds on the campus of Trinity College Dublin. It is probable that, if transmission between the two hosts occurs it is more likely to occur from humans to birds, due to the scavenging nature of these birds. This directly contrasts with the findings of Nunn *et al.* (2007) however, as the birds inhabiting the Great Saltee Island would rarely come into contact with humans and yet the prevalence of *C. dubliniensis* was reported to be 21–22% in this study. It is possible that the birds inhabiting the Great Saltee Island came into contact with human waste somewhere off the Irish coast, which could explain the increased recovery rate in this location (Nunn *et al.*, 2007). It is also possible that the true source of *C. dubliniensis* is the surface of the *I. uriae* ticks rather than the avian-excrement in which they reside.

To date, the most heavily populated group of *C. dubliniensis* isolates belong to that of ITS genotype 1, and have been recovered from both human and avian sources, and isolates of this ITS genotype have been reported from disparate locations throughout the world (Gee *et al.*, 2002; Al Mosaid *et al.*, 2005; Nunn *et al.*, 2007). These findings are supported by the current study, as the isolates recovered from the student population, the cystic fibrosis patients, and the patients attending the Dublin Dental Hospital were predominantly ITS genotype 1. The second most frequently encountered isolates belong to ITS genotype 2, which correlates with the results of the current study from each of the three cohorts studied. Isolates belonging to ITS genotypes 3 and 4 are encountered less frequently worldwide, however they appear to be more abundant in Middle Eastern countries (Al Mosaid *et al.*, 2005), which suggests geographic enrichment of such isolates in this region, although isolates of ITS genotype 1 have also been recovered in these areas (Gee *et al.*, 2002; Al Mosaid *et al.*, 2005). The current study recovered three *C. dubliniensis* ITS genotype 3 isolates and did not recovery any ITS genotype 4 isolates, correlating with the findings of previous studies which suggest a lower prevalence of such isolates in European countries (Gee *et al.*, 2002; Al Mosaid *et al.*, 2005).

Chapter 4

Population analysis of *Candida dubliniensis* using multilocus sequence typing (MLST), and a study of the genetic relationship of *C. dubliniensis* isolates recovered from human and avian-excrement-associated sources

4.1. Introduction

In order to be able to perform meaningful and informative epidemiological studies of *Candida* isolates it is essential to be able to discriminate between unrelated strains of the species of interest. This enables the source of nosocomial outbreaks to be identified and the population structure of the organism in question to be determined. Prior to advances in molecular techniques, phenotypic methods such as carbohydrate assimilation profiling, serotyping and morphology were used for this purpose. However, difficulties can arise due to phenotypic variability and poor discriminatory powers. Ideally, strain differentiation methods should be highly discriminatory, reproducible and suitable for the analysis of large numbers of isolates as part of inter-laboratory epidemiological studies.

4.1.1. Population structure of *C. dubliniensis* according to DNA fingerprint analysis

To date, DNA fingerprinting using the species-specific, semi-repetitive sequence-containing DNA probe Cd25 has been the most widely applied and informative tool used for *C. dubliniensis* epidemiology and population analysis. In brief, the Cd25 probe (which is highly homologous to repetitive DNA sequences which are dispersed across the *C. dubliniensis* genome) is applied to restriction endonuclease-digested DNA by Southern hybridisation. Hybridisation results in band patterns which then undergo computer-assisted analysis. Average S_{AB} values are calculated for every possible pair of *C. dubliniensis* isolates included in the study. This enables the construction of a dendrogram to infer genetic relationships across isolates of the same species. When first developed, data generated using this probe showed that *C. dubliniensis* is comprised of two distinct major clades, termed Cd25 fingerprint groups I and II, respectively (Joly *et al.*, 1999; Gee *et al.*, 2002). Cd25 group I isolates are all closely related with an S_{AB} value of approximately 0.8 (range 0.8 to 0.86). These isolates comprise the majority of isolates investigated to date and were recovered in many countries around the world, mainly from HIV-infected individuals (Joly *et al.*, 1999; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005). Furthermore, ITS sequence analysis revealed that Cd25 group I isolates consist of a single ITS genotype, namely ITS genotype 1 (Gee *et al.*, 2002). In contrast, Cd25 group II isolates are more diverse, with an average S_{AB} value of 0.52 (range 0.07 to 0.67) (Al Mosaid *et al.*, 2005) and consist of three

separate ITS genotypes (ITS genotypes 2–4), which corresponded to distinct subclades within the Cd25 group II fingerprinting clade (Gee *et al.*, 2002). More recently, a third major clade, termed Cd25 group III, was identified among isolates from Egypt and Saudi Arabia and displayed an average S_{AB} value of 0.35 (range 0.16 to 0.54) (Al Mosaid *et al.*, 2005). The DNA fingerprints of Cd25 group III isolates are very distinctive relative to isolates from Cd25 groups I and II (Fig. 1.3). Nucleotide sequence analysis of the ITS region of these Cd25 group III isolates revealed that they belong to ITS genotypes 3 and 4 (Al Mosaid *et al.*, 2005). Almost all Cd25 group III isolates examined to date exhibit resistance to 5FC (Al Mosaid *et al.*, 2005).

Although DNA fingerprinting has been shown to be a very useful tool in the molecular epidemiological analysis of *C. albicans* and *C. dubliniensis* populations, it is time consuming, expensive and not conducive to inter-laboratory comparisons. There are many other molecular strain-typing techniques that have been applied to the analysis of *Candida* species [e.g. karyotyping and randomly amplified polymorphic DNA fingerprinting (Soll, 2000)]. However, all of these methodologies also suffer from drawbacks, particularly in relation to reproducibility.

4.1.2. Multilocus sequence typing (MLST)

In the late 1990s MLST, a technique based on the nucleotide sequence analysis of a set of housekeeping genes, was developed for the population analysis of several bacterial species (Maiden *et al.*, 1998). This technique has also been applied to the analysis of the diploid yeast *C. albicans* (Bougnoux *et al.*, 2002, 2003; Tavanti *et al.*, 2003) and to other *Candida* species (Dodgson *et al.*, 2003; Tavanti *et al.*, 2005b; Jacobsen *et al.*, 2007). Using the MLST method different sequences obtained at each housekeeping locus are assigned as distinct genotypes. Each distinct genotype is assigned a corresponding integer, and for each isolate tested the combination of integers for each of the housekeeping loci gives rise to the isolates allelic profile. Each unique allelic profile is then assigned a ST number, or, in the case of diploid organisms, a DST number. In haploid organisms there are four possible variations at each polymorphic locus. However, in diploid organisms heterozygosity may present 10 possible variations, thus increasing the number of different genotypes that may occur at each housekeeping locus and increasing the discriminatory power of the method.

Bougnoux *et al.* (2002) identified six housekeeping gene loci that allowed accurate and reproducible discrimination between unrelated *C. albicans* isolates. A

study by Tavanti *et al.* (2003) used four of these loci and a further four loci and also showed high levels of discrimination. These two groups of researchers subsequently revised the combination of loci used in MLST analysis of *C. albicans*, aiming to identify the minimum number of loci required to maintain the high discriminatory power of the scheme (Bougnoux *et al.*, 2003). An agreed consensus scheme between the two different laboratories examined seven loci: *AAT1a, ACC1, ADP1, PMIb* [formerly *MPIb*], *ALA1* [formerly *SYA1*], *VPS13,* and *ZWF1b*. These seven loci allowed the application of MLST to the analysis of *C. albicans* epidemiology and population structure (Bougnoux *et al.*, 2002, 2004, 2006; Tavanti *et al.*, 2005a; Chen *et al.*, 2006; Odds *et al.*, 2006). MLST analysis has since been demonstrated to be as sensitive as DNA fingerprinting (Robles *et al.*, 2004) and due to the nature of the data (i.e., DNA sequences of specific loci) these can be used to create a large database generated by multiple laboratories.

4.1.3. Population and mating structure analysis using MLST

Population structure analysis can be carried out using concatenated MLST sequences, or on allelic profiles and resulting ST data. One of the most commonly used techniques for this type of analysis are dendrograms based on concatenated sequences and clustered using the unweighted-pair group method based on an arithmetic averages (UPGMA) algorithim. This method has shown that the strain groupings of *C. albicans* identified by MLST correlate with clades of *C. albicans* identified using the species-specific DNA fingerprinting probe Ca3 (Tavanti *et al.*, 2005a). The UPGMA method provides useful information on the genetic relatedness between closely related isolates. However, it delivers little information on patterns of evolutionary descent or the identification of founding STs. Another algorithim that is based upon related sequence types (eBURST) can divide large MLST datasets into non-overlapping groups of closely related STs or DSTs resulting in the formation of clonal complexes that consist of closely related isolates. The founder genotypes of each clonal complex can be hypothesised according to their establishment in the population, and closely related isolates are then identified as single locus or double locus variants (SLVs or DLVs).

Previously *C. albicans* was believed to lack a sexual cycle. However, Hull & Johnson (1999) identified a set of genes in this species that corresponded to the mating type locus of *S. cerevisiae*, which regulates the sexual cycle of the latter yeast species. This locus is referred to as the mating type-like locus (*MTL*) in *C. albicans*, and since its

discovery, it has been observed that mating occurs in cells that are in the "opaque" mating competent form. These mating-competent isolates are naturally homozygous at the *MTL* locus, displaying either a/a or α/α genotypes (Hull & Johnson 1999; Lockhart *et al.*, 2003). The mating structure of a population may be investigated using the MLST method. Combination of the specific alleles at each of the biallelic sites on the same chromosome for each gene on the same chromosome enables the definition of haplotypes. Previous analysis of such haplotypes in *C. albicans* has suggested that the species may undergo sexual reproduction, although its primary means of reproduction is clonal (Tavanti *et al.*, 2004).

4.1.4. Genetic relatedness of human and animal isolates of *C. albicans*

Candida albicans has previously been recovered from many different types of animals including birds, primates, reptiles, amphibians, as well as both wild and domesticated animals (Odds, 1988; Buck, 1990; Pressler *et al.*, 2003; Bougnoux *et al.*, 2004; Edelmann *et al.*, 2005; Tavanti *et al.*, 2005a; Cafarchia *et al.*, 2006; Jacobsen *et al.*, 2008; Wrobel *et al.*, 2008). A number of previous studies have analysed the genetic relationships between *C. albicans* isolates from these sources in comparison to those from humans using both Ca3 DNA fingerprinting and MLST (Bougnoux *et al.*, 2004; Edelmann *et al.*, 2005; Tavanti *et al.*, 2005a; Jacobsen *et al.*, 2008; Wrobel *et al.*, 2008). All of these studies concurred that genetic separation is evident between *C. albicans* isolates recovered from animals and humans. Studies using MLST analysis has suggested that MLST clade 8 is enriched with isolates of animal origin, whereas MLST clade 1 is devoid of such isolates, with the exception of those recovered from primates (Jacobsen *et al.*, 2008; Wrobel *et al.*, 2008). It has been suggested that MLST clade 1 isolates may be better adapted for colonisation and infection in humans, as isolates belonging to this clade are most frequently recovered from humans (Jacobsen *et al.*, 2008).

A study by Nunn *et al.* (2007) reported the recovery of a large number of *C. dubliniensis* isolates from the surface of *I. uriae* (ticks) living in seabird excrement at a seabird colony on the Great Saltee Island off the South-Eastern Irish coast. This study was the first to report the recovery of *C. dubliniensis* from non-human sources. All of these isolates were reported as belonging to ITS genotype 1 and were recovered from one location on the island on two separate occasions. The study suggested that *C. dubliniensis* may inhabit the gastrointestinal tracts of birds, and presented an interesting

opportunity to investigate the genetic relationship between human and environmental isolates of *C. dubliniensis*.

4.1.5. Objectives

There were three main objectives for this part of the study:

- To develop MLST as an informative alternative to previously used DNA fingerprinting and ITS genotyping approaches for investigating the population structure of *C. dubliniensis*. At the outset of the study, MLST had several considerable potential advantages over fingerprinting using species-specific probes, including improved reproducibility, rapidity and cost effectiveness. This study aimed to compare the population structure of *C. dubliniensis* defined by MLST, Cd25 fingerprinting and ITS genotyping.
- To investigate if the application of an identical set of MLST loci to *C. albicans* and *C. dubliniensis* isolates could be used to validly compare the population structures of the two species and to provide insights into their evolutionary relatedness. This was undertaken by applying the previously established *C. albicans* MLST scheme to isolates of *C. dubliniensis*.
- To investigate the genetic relationships between *C. dubliniensis* isolates recovered from humans and from seabird excrement using MLST to determine the likelihood that isolates could be transmitted between the two hosts.

4.2. Materials and methods

4.2.1. *Candida dubliniensis* isolates

The *C. dubliniensis* isolates used in this study are shown in Table 4.1. Isolates were selected from diverse geographical locations (Fig. 3.1) and included representatives of the three Cd25 major fingerprinting clades and all four ITS genotypes. The majority of the isolates studied had been described previously. However, a number of new clinical isolates were also included (Table 4.1). These new isolates were identified using phenotypic and genotypic methods as described in Chapter 2 Sections 2.2 and 2.4. A selection of three *C. dubliniensis* isolates recovered from seabirds as previously described in Section 3.2.2 of this study, and 11 isolates recovered by Nunn *et al.* (2007) were also included. A number of *C. dubliniensis* isolates with unusual phenotypes such as inability to produce chlamydospores on Pal's agar following 48 h incubation at 30°C (isolates 98-277, 95-677 and 94-234), or growth as white/pale pink colonies on CHROMagar Candida™ (isolates 17P, 18P, 22BP, 25P, 110P and 140P) medium following 48 h incubation at 37°C were also included in the present study (Table 4.1). These nine isolates were sent to the laboratory by Dr. José Pontón at the Universidad del País Vasco, Bilbao, Spain.

4.2.2. *Candida albicans* isolates

A selection of 50 human *C. albicans* isolates (Table 4.2) were chosen to represent a range of the 17 MLST clades recently described by Odds *et al.* (2007). Sequence data for each of the 50 *C. albicans* isolates were available at http://test1.mlst.net/ for the seven collaborative consensus MLST loci; *AAT1a, ACC1, ADP1, PMIb, ALA1, VPS13, ZWF1b* (Odds *et al.*, 2007), and at http://calbicans.mlst.net/ for the *RPN2* locus (Bougnoux *et al.*, 2002). The seven recommended MLST loci for *C. albicans* and the additional *RPN2* locus were amplified and sequenced as recommended (Bougnoux *et al.*, 2002; Bougnoux *et al.*, 2003) for each of 17 *C. albicans* isolates recovered from seabirds as described in Chapter 3, Section 3.3.2. All locus sequences were treated in an identical manner as the *C. dubliniensis* sequence data as described below.

Table 4.1. *Candida dubliniensis* isolates investigated by MLST analysis

Human isolates	Origin[c]	Yr of isolation	Underlying patient condition	Sample	ITS Genotype	Cd25 group	Source/ Reference
CD36[ab]	Ireland	1988	HIV Pos	Oral	1	I	(Sullivan *et al.*, 1995; Gee *et al.*, 2002)
CD06033[ab]	Ireland	2006	CF	Sputum	1	ND	This study
CD06031[a]	Ireland	2006	CF	Sputum	1	ND	This study
CD060213[a]	Ireland	2006	CF	Sputum	1	ND	This study
CD06041[a]	Ireland	2006	CF	Sputum	1	ND	This study
CD06038[ab]	Ireland	2006	CF	Sputum	1	ND	This study
CD06045[a]	Ireland	2006	CF	Sputum	1	ND	This study
CD06037[ab]	Ireland	2006	CF	Sputum	2	ND	This study
CD06036[ab]	Ireland	2006	CF	Sputum	2	ND	This study
CD060215[ab]	Ireland	2006	CF	Sputum	3	ND	This study
CD604[ab]	France	2000	HIV Pos	Throat	1	ND	This study
CD603[ab]	France	2000	HIV Pos	Throat	1	ND	This study
CM1a[b]	Australia	1991	HIV Pos	Oral	1	I	(Sullivan *et al.*, 1995; Gee *et al.*, 2002)
SA105[a]	S. Arabia	2002	Diabetes	Oral	1	I	(Al Mosaid *et al.*, 2005)
SA115[a]	S. Arabia	2002	Diabetes	Oral	1	I	(Al Mosaid *et al.*, 2005)
SA116[a]	S. Arabia	2002	Diabetes	Oral	1	I	(Al Mosaid *et al.*, 2005)
SA108[ab]	S. Arabia	2002	Diabetes	Oral	3	III	(Al Mosaid *et al.*, 2005)
SA100[ab]	S. Arabia	2002	Leukaemia	Oral	3	III	(Al Mosaid *et al.*, 2005)
SA121[ab]	S. Arabia	2002	S/P Renal Tx	Oral	4	III	(Al Mosaid *et al.*, 2005)
Eg202[a]	Egypt	2002	Cancer	Oral	4	III	(Al Mosaid *et al.*, 2005)
Eg203[a]	Egypt	2002	Cancer	Oral	1	I	(Al Mosaid *et al.*, 2005)
Eg204[a]	Egypt	2002	Cancer	Oral	1	I	(Al Mosaid *et al.*, 2005)
Eg207[a]	Egypt	2002	Diabetes	Oral	4	III	(Al Mosaid *et al.*, 2005)
p6785[ab]	Israel	1999	HIV Neg	Urine	3	II	(Al Mosaid *et al.*, 2005; Polacheck *et al.*, 2000)
p7276[ab]	Israel	1999	HIV Neg	RT	3	II	(Al Mosaid *et al.*, 2005)
p7718[a]	Israel	1999	HIV Neg	Wound	4	III	(Gee *et al.*, 2002; Al Mosaid *et al.*, 2005)
CD71[a]	Argentina	1994	HIV Pos	Oral	1	I	(Sullivan *et al.*, 1997; Gee *et al.*, 2002;)
CD98923[a]	India	1998	HIV Pos	Oral	1	I	(Al Mosaid *et al.*, 2001)
B1324[a]	USA	1998	N/A	Tongue	1	ND	This study
B341[a]	USA	1998	N/A	Throat	1	ND	This study
PM6-2[a]	Chile	2006	N/A	N/A	1	ND	This study
P2[a]	Switzerland	1993	HIV Pos	Oral	1	I	(Gee *et al.*, 2002; Al Mosaid *et al.*, 2005)
1504[a]	Slovakia	2005	N/A	Tonsular swab	1	ND	(Melkusova *et al.*, 2005)
8882[a]	Slovakia	2005	CH	Tonsular swab	1	ND	(Melkusova *et al.*, 2005)
9097[a]	Slovakia	2005	Leukaemia	Tonsular swab	1	ND	(Melkusova *et al.*, 2005)
966/3(1)[ab]	Slovakia	2005	N/A	Sputum	1	ND	(Melkusova *et al.*, 2005)
966/3(2)[a]	Slovakia	2005	N/A	Sputum	1	ND	(Melkusova *et al.*, 2005)
CCY29-177-1[a]	Slovakia	2005	Pelvic organ inflammation	Cervial swab	1	ND	(Melkusova *et al.*, 2005)
MLNIH0479[a]	Thailand	2005	Leukaemia	Blood	1	ND	(Bujdakova *et al.*, 2004)
MLNIH0720[a]	Thailand	2005	TB,Anaemia, Diabetes	Oral	1	ND	(Bujdakova *et al.*, 2004)
49831[a]	Japan	2005	Anaemia	Sputum	2	ND	(Bujdakova *et al.*, 2004)
IFM49883[a]	Japan	2005	N/A	N/A	2	ND	(Bujdakova *et al.*, 2004)
IFM0492[a]	Thailand	2005	Cancer	Blood	1	ND	(Bujdakova *et al.*, 2004)
IFM49832[a]	Japan	2005	Diabetes	Sputum	1	ND	(Bujdakova *et al.*, 2004)
Can4[a]	Canada	1996	HIV Pos	Oral	2	II	(Pinjon *et al.*, 1998; Gee *et al.*, 2002;)
CD539[a]	U.K	1994	AIDS	Oral	2	II	(Pinjon *et al.*, 1998; Gee *et al.*, 2002;)
CBS2747[a]	Netherlands	1952	HIV Neg	Sputum	2	II	(Meis *et al.*, 1999 ; Gee *et al.*, 2002)
CD514[a]	Ireland	1995	HIV Neg	Oral	3	II	(Gee *et al.*, 2002)

Continued overleaf

Table 4.1. ***Continued***

Human isolates	Origin[c]	Yr of isolation	Underlying patient condition	Sample	ITS Genotype	Cd25 group	Source/ Reference
CD519[a]	Ireland	1997	AIDS	Oral	3	II	(Gee *et al.*, 2002)
CD75004[a]	UK	1975	Diabetes	Oral	2	II	(Pinjon *et al.*, 1998; Gee *et al.*, 2002)
49831[a]	Japan	2005	Anaemia	Sputum	2	ND	(Bujdakova *et al.*, 2004)
IFM49883[a]	Japan	2005	N/A	N/A	2	ND	(Bujdakova *et al.*, 2004)
IFM0492[a]	Thailand	2005	Cancer	Blood	1	ND	(Bujdakova *et al.*, 2004)
IFM49832[a]	Japan	2005	Diabetes	Sputum	1	ND	(Bujdakova *et al.*, 2004)
98-277	Spain	1998	N/A	Oral	1	ND	(Brena *et al.*, 2004)
95-677	Spain	1995	HIV Pos	Oral	1	ND	This study
94-234	Spain	1994	HIV Pos	Oral	1	ND	(Brena *et al.*, 2004)
CD505	Ireland	1989	HIV Pos	Oral	1	I	(Gee *et al.*, 2002)
CD06032	Ireland	2006	Healthy	Oral	1	ND	This study
CD06027	Ireland	2006	Healthy	Oral	1	ND	This study
CD0512	Ireland	2005	Systemic steroids	Oral	1	ND	This study
CD524	Ireland	1997	HIV Neg	Oral	1	I	(Gee *et al.*, 2002)
17P	Argentina	N/A	Gingivitis	Oral	2	ND	This study
18P	Argentina	N/A	Periodontitis	Oral	2	ND	This study
22BP	Argentina	N/A	Periodontitis	Oral	2	ND	This study
25P	Argentina	N/A	Periodontitis	Oral	2	ND	This study
110P	Argentina	N/A	Periodontitis	Oral	2	ND	This study
140P	Argentina	N/A	Gingivitis	Oral	2	ND	This study
Avian *C. dubliniensis* isolates							
SL411	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL422	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL370	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL410	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL375-I	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL375-II	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL397	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL414	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL495	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL509	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
SL522	GSI	2007	*I. uriae*	S/E	1	ND	(Nunn *et al.*, 2007)
AV5	TCD	2008	*L. argentatus*	S/E	1	ND	This study
AV6	TCD	2008	*L. argentatus*	S/E	1	ND	This study
AV7	TCD	2008	*L. argentatus*	S/E	1	ND	This study

[a] Isolates used in the development of the *C. dubliniensis* MLST scheme.

[b] Isolates included in assessment of sequence stability and MLST reproducibility.

[c] Geographical location from which isolates were recovered. GSI, Great Saltee Island on the SouthEast coast of Ireland; TCD, campus of Trinity College Dublin.

Abbreviations: HIV Pos; Human immunodeficiency virus-infected; HIV Neg; Human immunodeficiency virus negative; CF; Cystic Fibrosis; S. Arabia; Saudi Arabia; S/P Renal Tx; stage-post renal transplantation; N/A; not available; CH; Congenital Hydrocephalus; RT; Respiratory tract; TB; Tuberculosis; AIDS; Acquired immunodeficiency syndrome patient; *I. uriae; Ixodes uriae*; *L. argentatus; Larus argentatus* (herring gull); S/E; Seabird excrement; isolates recovered at sites where black headed gulls (*Larus ridibundus*), *Larus argentatus* (herring gull) and swans (*Cygnus olor*) were present.

Table 4.2. *Candida albicans* isolates included in the MLST analysis

Human isolate	Origin[a]	Yr of isolation	Source	Sample	DST[b]	MLST Clade	Reference
APRHCRC1	France	2001	Candidaemia	Blood	915	4	(Odds *et al.*, 2007)
APRURC3	France	2001	Candiduria	Urine	529	6	(Odds *et al.*, 2007)
APRURM3	France	2002	Candiduria	Urine	917	1	(Odds *et al.*, 2007)
APRURM6	France	2002	Candiduria	Urine	918	9	(Odds *et al.*, 2007)
APRURM7	France	2002	Candiduria	Urine	919	10	(Odds *et al.*, 2007)
BCHURC10	France	2002	Candiduria	Urine	921	10	(Odds *et al.*, 2007)
BCHURC15	France	2002	Candiduria	Urine	922	11	(Odds *et al.*, 2007)
BCHURS24	France	2002	Candiduria	Urine	923	S	(Odds *et al.*, 2007)
BCHURS3	France	2001	Candiduria	Urine	924	4	(Odds *et al.*, 2007)
BCHURS5	France	2001	Candiduria	Urine	925	11	(Odds *et al.*, 2007)
BCHURS9	France	2001	Candiduria	Urine	926	4	(Odds *et al.*, 2007)
CCHHCRM11	France	2002	Candidaemia	Blood	927	4	(Odds *et al.*, 2007)
CCHHCRM4	France	2001	Candidaemia	Blood	928	14	(Odds *et al.*, 2007)
CCHHCRM7	France	2001	Candidaemia	Blood	918	9	(Odds *et al.*, 2007)
CCHURM1	France	2001	Candiduria	Urine	929	15	(Odds *et al.*, 2007)
CCHURM12	France	2001	Candiduria	Urine	306	9	(Odds *et al.*, 2007)
CCHURM4	France	2001	Candiduria	Urine	931	11	(Odds *et al.*, 2007)
CCHURM9	France	2001	Candiduria	Urine	932	16	(Odds *et al.*, 2007)
CCHURS10	France	2002	Candiduria	Urine	933	9	(Odds *et al.*, 2007)
CP04	USA	NA	NA	NA	951	1	(Odds *et al.*, 2007)
CP05	USA	NA	NA	NA	952	S	(Odds *et al.*, 2007)
CP06	USA	NA	NA	NA	953	1	(Odds *et al.*, 2007)
CP08	USA	NA	NA	NA	954	1	(Odds *et al.*, 2007)
CP12	USA	NA	NA	NA	956	4	(Odds *et al.*, 2007)
CP15	USA	NA	NA	NA	957	3	(Odds *et al.*, 2007)
CP54	USA	NA	NA	NA	976	2	(Odds *et al.*, 2007)
CP58	USA	NA	NA	NA	659	4	(Odds *et al.*, 2007)
CP85	USA	NA	NA	NA	659	4	(Odds *et al.*, 2007)
DPC111	Belgium	2000	Carriage	Mouth	869	9	(Bougnoux *et al.*, 2006)
DPC118	Belgium	2000	Carriage	Mouth	871	1	(Bougnoux *et al.*, 2006)
DPC168	Belgium	2000	Carriage	Mouth	874	11	(Bougnoux *et al.*, 2006)
DPC18	Belgium	2000	Carriage	Feces	883	12	(Bougnoux *et al.*, 2006)
DPC2	Belgium	2000	Carriage	Feces	538	11	(Bougnoux *et al.*, 2006)
DPC206	Belgium	2000	Carriage	Feces	885	1	(Bougnoux *et al.*, 2006)
DPC208	Belgium	2000	Carriage	Feces	887	5	(Bougnoux *et al.*, 2006)
DPC22	Belgium	2000	Carriage	Feces	889	12	(Bougnoux *et al.*, 2006)
DPC25	Belgium	2000	Carriage	Mouth	890	3	(Bougnoux *et al.*, 2006)
DPC28	Belgium	2000	Carriage	Mouth	891	11	(Bougnoux *et al.*, 2006)
DPC35	Belgium	2000	Carriage	Mouth	896	17	(Bougnoux *et al.*, 2006)
DPC37	Belgium	2000	Carriage	Mouth	572	1	(Bougnoux *et al.*, 2006)
DPC44	Belgium	2000	Carriage	Mouth	897	11	(Bougnoux *et al.*, 2006)
DPC47	Belgium	2000	Carriage	Feces	898	3	(Bougnoux *et al.*, 2006)
DPC5	Belgium	2000	Carriage	Mouth	899	11	(Bougnoux *et al.*, 2006)
DPC55	Belgium	2000	Carriage	Feces	900	1	(Bougnoux *et al.*, 2006)
DPC6	Belgium	2000	Carriage	Feces	901	11	(Bougnoux *et al.*, 2006)
DPC65	Belgium	2000	Carriage	Feces	904	11	(Bougnoux *et al.*, 2006)
DPC66	Belgium	2000	Carriage	Mouth	905	9	(Bougnoux *et al.*, 2006)
DPC70	Belgium	2000	Carriage	Mouth	906	11	(Bougnoux *et al.*, 2006)
DPC81	Belgium	2000	Carriage	Digestive tract	910	2	(Bougnoux *et al.*, 2006)
DPC90	Belgium	2000	Carriage	Mouth	912	4	(Bougnoux *et al.*, 2006)
Avian *C. albicans* isolates							
AV101	TCD	2008	*L. argentatus*	S/E	1241	3	This study
AV102	Dún Laoghaire	2008	*L. argentatus*	S/E	1242	11	This study
AV103	Dún Laoghaire	2008	*L. argentatus*	S/E	538	11	This study
AV104	Dún Laoghaire	2008	U/S	S/E	1243	2	This study
AV105	Dún Laoghaire	2008	U/S	S/E	1243	2	This study

Continued overleaf

Table 4.2. ***Continued***

Avian isolate	Origin[a]	Yr of isolation	Source	Sample	DST[b]	MLST Clade	Reference
AV106	Dún Laoghaire	2008	U/S	S/E	1243	2	This study
AV107	Dún Laoghaire	2008	U/S	S/E	1243	2	This study
AV108	Dún Laoghaire	2008	U/S	S/E	1244	2	This study
AV109	Dún Laoghaire	2008	U/S	S/E	1243	2	This study
AV113	Sandymount	2008	U/S	S/E	1246	4	This study
AV114	Sandymount	2008	U/S	S/E	1247	S	This study
AV115	Bray	2008	U/S	S/E	1248	16	This study
AV116	Howth	2008	U/S	S/E	52	1	This study
AV118	Howth	2008	U/S	S/E	1249	4	This study
AV119	TCD	2008	*L. argentatus*	S/E	52	1	This study
AV122	TCD	2008	*L. argentatus*	S/E	52	1	This study
AV123	TCD	2008	*L. argentatus*	S/E	1245	S	This study

[a] Origin; Geographical location from which isolates were recovered.

[b] DST; Diploid Sequence type assigned to each isolate according the consensus MLST scheme for *C. albicans*.

Abbreviations; *L. argentatus; Larus argentatus* (herring gull); U/S; Unknown species; S/E; Seabird excrement, isolates recovered at sites where black-headed gulls (*Larus ridibundus*), herring gulls (*Larus argentatus*) and swans (*Cygnus olor*) were present; S; singletons, i.e., isolates that were not assigned to any MLST clade.

4.2.3. Preparation of template DNA

Candida isolates were grown overnight in 5 ml of YPD broth and cells from 1.5 ml of culture were harvested by centrifugation. DNA was extracted from the cells as described in Chapter 2 Section 2.3.3.

4.2.4. Genotyping of *C. dubliniensis* using PCR amplification

Template DNA was tested in separate PCR amplification experiments with each of the primer pairs G1F/G1R, G2F/G2R, G3F/G3R, and G4F/G4R to identify the ITS genotype of the isolate (Gee *et al.*, 2002) as outlined in Chapter 2 Section 2.4.2.

4.2.5. Selection of loci for MLST analysis

Due to the high level of sequence homology between the majority of *C. albicans* and *C. dubliniensis* open reading frames (ORFs) (Moran *et al.*, 2004), all loci previously examined for the purpose of MLST analysis in *C. albicans* were also investigated for their potential use with *C. dubliniensis*. The usefulness of the six genes examined in *C. albicans* as described by Bougnoux *et al.* (2002) (i.e., *ACC1, VPS13, GLN4, ADP1, RPN2* and *ALA1*), and the additional four loci described by Tavanti *et al.* (2003) (i.e., *AAT1a, AAT1b, PMIb* and *ZWF1b*) was assessed. The *C. albicans* MLST loci sequences were used in separate BLAST searches against the *C. dubliniensis* genome sequence database (the Wellcome Trust Sanger Institute *C. dubliniensis* genome sequence project (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/). For each MLST locus, sequences were aligned using the CLUSTAL W sequence alignment computer program (Thompson *et al.*, 1994), and the *C. albicans* primer-binding regions were identified in the alignment. Any nucleotide differences in the primer-binding regions between the species were adjusted in the corresponding *C. dubliniensis* oligonucleotides to facilitate optimum amplification. All of the *C. dubliniensis*-optimised primer pairs yielded a single PCR product of the expected size (ranging in size from 400 bp to 700 bp) (Table 4.3).

4.2.6. PCR amplification and sequence determination

The DNA of the *C. dubliniensis* isolates underwent PCR amplification in 50 µl reaction volumes containing 200 µM of each dNTP, 1.25 U of GoTaq polymerase (Promega) and 10 µl (1 ×) of GoTaq FlexiBuffer (Promega), 3 µM $MgCl_2$, 100 pmol of each primer, and 100 ng DNA template. Amplification products were purified using a

QIAquick 96-well PCR purification kit (Qiagen Science, MD) and were sequenced on both strands using the same primers that were previously used for amplification. DNA sequencing reactions were performed commercially as described in Chapter 2 Section 2.6.

4.2.7. Sequence analysis and sequence type determination

Sequence analysis was performed by examination of chromatogram files using the ABI prism Seqscape software version 2.0 (Applied Biosystems, Foster City, CA). Numbers were assigned to unique genotypes for each locus, and genotype numbers were then combined to yield a DST number. Allelic profiles and DST numbers are available for each MLST scheme in Appendices A–E in the final Section of this thesis. The sequences of all of the loci examined are provided in Appendix F in the final section. Maximum parsimony trees and unweighted-pair group method with arithmetic averages (UPGMA) dendrograms were constructed using Bionumerics version 5.1 (Applied Maths NV, Sint-Martens-Latem, Belgium) based on concatenated *C. albicans* and *C. dubliniensis* MLST sequences. Neighbour-joining trees were constructed using Mega software program version 3.1 (Kumar *et al.*, 2004). The START2 software program (Jolley *et al.*, 2001) was used to construct UPGMA dendrograms that were based on allelic profiles and DSTs, and to display clonal complexes and hypothesise founder genotypes following BURST analysis.

4.2.8. Linkage disequilibrium and clonality

Linkage disequilibrium was assessed using the Index of Association (Ia), as described by Smith *et al.* (1993) as calculated by the Multilocus 1.3 software package, available at http://www.agapow.net/software/multilocus/ (Agapow & Burt, 2001) using genotype numbers for all loci (scheme C) from all *C. dubliniensis* isolates. The levels of significance for non-random association between loci were computed under the null hypothesis of a freely recombining population (panmixia). Primary founding genotypes and clonal complexes were predicted by BURST analysis using the START2 software program (Jolley *et al.*, 2001) with genotype numbers for all loci (scheme D) from all *C. dubliniensis* isolates.

Table 4.3. Oligonucleotide primers and sequences used for *C. dubliniensis* MLST analysis

Locus	Amplicon size (bp)	No. of bases analysed (bp)	Co-ordinates of *C. albicans* amplicon[b]	Co-ordinates of *C. dubliniensis* amplicon[b]	% sequence identity of analysed segment[c]	Primer (5′→3′) [d]	Analysed segment start (5′) and end (3′) points in *C. dubliniensis*[e]
AAT1a	480	373	+31 → +508	+35→ +497	91	F-A**TCAAACTA**CTAAATTTTTGAC R-C**G**GCAACATGATTAGCCC	5′-ATTGAAA 3′-CG**A**TT**T**
AAT1b	491	341	+736 → +1,226	+742→ +1,233	95	F-ATGGCTTATCAAGGTTTTGC R-GT**A**GCATAAACTGAATA**A**TC**A**	5′-TT**A**ACTAA 3′-TTGGGA**T**CA
ACC1	519	407	+3,184 → +3,702	+3,190→ +3,708	93	F-GC**C**AGAGAAATTTT**G**AT**C**CAATGT R-TTCATCAACATCATCCAAGTG	5′-TTTTGAG**A**T 3′-**T**ACAAGA
ADP1	537	443	+868 → +1,404	+868→ +1,404	90	F-GAGCCAAGTATGAATGA**C**TTG R-TTGATCAACAAACCCGATAAT	5′-**T**ACGTTGCAA 3′-GGAAATCCAA
GLN4	483	404	+82 → +564	+82 → +564	100	F-GAGATAGT**T**AAGAATAAAAAAGTTG R- **G**TCTCTTTC**G**TCTTT**A**GGACCCAATC	5′-TC**T**GCTTTA 3′-TTCAAACC
PMIb	486	375	+406 → +891	+406→ +892	90	F-ACCAGAAATGGCC R-**G**CAGCCAT**A**CATTCAATTAT	5′-TTTAA**G**C 3′-GGGAAGCA
RPN2	447	306	+1,012 → +1,458	+1,015 → +1,461	90	F-TT**T**ATGCATGCTGGTACTAC**T**GATG R-TAA**C**CCCATAC**T**CAAAGCAGCAGCCT	5′-TTGGTC**C**AA**G** 3′-GTCTTTACGA
ALA1	543	391	+2,284 → +2,826	+2,284→ +2,826	90	F-AGAAGAAT**A**GTTGCTCTTACTG R-GTT**G**CC**C**TTACCACCAGCTTT	5′-TAAATCCAAG 3′-AG**T**CTGTATC**T**
VPS13 (ex*VPS13)*[a]	741	403 (675)[a]	+4,854 → +5,594	+4,858→ +5,598	88	F-CGTTGAGAGATATTCGACTT R-ACGGAT**C**GATC**G**CCA**A**TCC	5′-CCTTGATATG 3′-**A**AAATC**T**TGG (5′-AGAGC**A**AA**C**G (3′-AAACCTTGG)
ZWF1b (ex*ZWF1b)*[a]	702	491 (621)[a]	+787 → +1,469	+766→ +1,468	93	F-GTTTCATTTGATCCTGAAGC R-GCCATTGATAAGTACCTGGAT	5′-**C**AAACCAGG 3′-T**A**GAATTAC (5′-AAAGTTTTAAAA) (3′-GAAAATATTTGAA**A**)

[a] Numbers and loci in parenthesis denote extended sequence fragments analysed for polymorphisms.

[b] Nucleotide coordinates for MLST amplicons are numbered based on the adenine residue of the ATG start codon at the 5′ end of the gene being designated +1.

[c] Sequence homology between *C. dubliniensis* and *C. albicans* at the sequenced fragments.

[d] Nucleotide bases shown in bold typeface and underlined denote nucleotide bases that vary between *C. albicans* and *C. dubliniensis.*

[e] Loci in which additional sequence data was analysed in both 5′ and 3′ directions of the original sequence fragment (range; 100 bp to 280 bp) due to the presence of additional polymorphic sites.

4.2.9. Stability and reproducibility of MLST method

The stability and reproducibility of the sequence data at each MLST locus was assessed by carrying out the sequence analysis in duplicate on two randomly selected isolates. For each locus and isolate, the duplicate DNA extractions, PCR reactions and sequencing reactions were carried out independently. Resulting sequence duplicates for each isolate were compared to each other.

4.2.10. PCR amplification of *CdCDR1* alleles

In order to investigate the prevalence of a common point mutation in the *CDR1* gene of *C. dubliniensis* isolates included in the study, alleles of the gene were amplified using the 2F (5′-ATCCTGTTGGTTATGTGTTCG-3′) and 2R (5′-GGGAAATCAACACTTCCAGTC-3′) primers described by Moran *et al.* (2002). The presence of the mutation was identified by restriction fragment length polymorphism analysis (RFLP) of the amplicons. Digestion of the amplicons in isolates carrying a mutant allele with the restriction endonuclease *SspI* (New England Biolabs) yields restriction fragments of 200 and 290 bp in length, while in wild-type strains the approx. 500 bp fragment remains intact. Digestion products were visualised on a UV transilluminator following electrophoresis through 2.0% (w/v) agarose gels containing 0.5 μg ethidium bromide/ml.

4.2.11. Determination of mating types by PCR amplification

The mating types of *C. dubliniensis* isolates included in the study were determined in order to investigate the prevalence of each mating type in the *C. dubliniensis* population. This was determined by PCR using specific primers that were designed to specifically amplify a portion of either the a/a or α/α *MTL* loci of a homozygous isolate, or both in the case of a heterozygote. Each PCR reaction contained specific primers MTLa1-F (5′-TGAAAATGAAGACAATGCGA-3′) and MTLa1-R (5′-CATCTTTTTCTGCTATCAATTC-3′) which amplified 535 bp of the *MTL* type a locus, and specific primers MTLα2-F (5′-ATGAATTCACATCTGGAGGC-3′) and MTLα2-R (5′-CTGTTAATAGCAAAGCAGCC-3′) which amplified 615 bp of the *MTL* type α locus. Thus, a homozygous isolate resulted in the formation of only one PCR product and a heterozygous isolate resulted in the amplification of both loci and the formation of two PCR products. These primers were originally described by Rustad *et al.* (2002) for determination of mating types in isolates of *C. albicans*, but were

modified to enable binding to *C. dubliniensis* DNA. Amplification reactions were carried out in a final reaction volume of 50 µl containing 10 pmol of each of the forward and reverse primers, 2.5 mM $MgCl_2$, 10 mM Tris/HCl (pH 9.0 at 25 °C) 10 mM KCl, 0.1% (v/v) Triton X-100, 1.25 U GoTaq polymerase (Promega) and 25 µl of template DNA. Cycling conditions consisted of a 10 min denaturation step at 94°C, followed by 30 cycles of 1 min at 94°C, 2 min at 55°C, and 3 min at 72°C, followed by a final elongation step of 10 min at 72°C. Amplification products were separated by electrophoresis through 2% (w/v) agarose gels containing 0.5 µl ethidium bromide/ml and were visualised on a UV transilluminator.

4.2.12. Statistical analysis

The discriminatory power of MLST schemes were determined according to the probability that two unrelated isolates sampled from a test population will be placed into different typing groups and was calculated as previously outlined (Hunter, 1990). Fisher's exact test (http://www.exactoid.com/fisher/index.php) was used to determine the distribution of isolates within a population.

4.3. Results I: Establishing an MLST scheme for *C. dubliniensis*

4.3.1. Development of an MLST scheme for *C. dubliniensis*

Multilocus sequence typing has been demonstrated to be an effective tool to analyse the epidemiology and population biology of *C. albicans* (Tavanti *et al.*, 2005a). Since *C. albicans* and *C. dubliniensis* are so closely related we investigated if the same set of locus sequences could also be applied to MLST analysis of *C. dubliniensis*. To investigate this the *C. dubliniensis* genome sequence (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/) was examined to identify the presence of genes homologous to those used previously in the MLST analysis of *C. albicans* (Bougnoux *et al.*, 2002, 2003; Tavanti *et al.*, 2003). Homologous genes were found in each case and comparison of the sequences of the complete ORFs of the orthologous pairs revealed sequence similarities ranging from 89% – 94%, while the parts of the genes that were analysed for MLST purposes displayed 88% – 100% sequence identity (Table 4.3).

The current consensus MLST scheme for *C. albicans* examines 2,883 nucleotides from 7 loci, namely *AAT1a, ACC1, ADP1, PMIb, ALA1, VPS13,* and *ZWF1b*. For the purposes of this study, this scheme is referred to as Scheme A (Table 4.4). In *C. albicans*, SNPs have been detected in 172 of these 2,883 (6%) nucleotides (Bain *et al.*, 2007). The corresponding Scheme A loci were examined in the initial 50 *C. dubliniensis* isolates included in this study. The Scheme A loci resulted in 23 SNPs from the equivalent 2,883 nucleotides (0.8%), and identified 33 genotypes. In Scheme A, only one site (i.e., position 186 in *ALA1*) showed a polymorphism at the same position in both *C. albicans* and *C. dubliniensis*, whereas all other variable sites were at different locations in the two species. Tavanti *et al.* (2003) analysed eight loci (*AAT1a, AAT1b, ACC1, ADP1, PMIb, ALA1, VPS13,* and *ZWF1b*) for sites of heterozygosity amongst 50 isolates of *C. albicans*. The group identified 87 sites that displayed SNPs, 71 of which (81.6%) also displayed heterozygosity. The same loci were studied for sites of heterozygosity in the initial 50 isolates of *C. dubliniensis* (Table 4.1), and of the 30 sites displaying variability, nine (30%) displayed heterozygosity.

Due to the low levels of polymorphism observed in the *C. dubliniensis* loci examined, increasing the length of the sequences examined was investigated to determine whether this might increase the discriminatory power of the method. To

achieve this, additional nucleotide sequence data (range; 100 bp to 280 bp) at each of the loci in both 5′ and 3′ directions of the original sequence fragment were also analysed for the potential presence of polymorphic sites. Sequence analysis of the extended fragments revealed the presence of additional SNPs in *VPS13*, and *ZWF1b* (Table 4.4). The inclusion of these two extended sequences together with the sequences of the five other loci generated a second MLST scheme, termed Scheme B (Table 4.4), which was made up of the sequences of the *AAT1a, ACC1, ADP1, PMIb, ALA1,* ex*VPS13,* and ex*ZWF1b* loci. Examination of the extended sequence from *VPS13* (ex*VPS13*) revealed one extra SNP which gave rise to two extra genotypes, and the extended sequence of *ZWF1b* (ex*ZWF1b*) yielded two further SNPs, which in turn gave rise to one extra genotype. Scheme B resulted in a new total of 3,285 nucleotides, 26 (0.8%) of which displayed SNPs, resulting in four additional genotypes (Table 4.4).

In a further attempt to improve the discriminatory power of this method for *C. dubliniensis*, three other sequence segments were also analysed in *C. dubliniensis* that were previously investigated for possible use in MLST analysis of *C. albicans*, although these were not included in the final consensus *C. albicans* MLST scheme. Sequences from these additional loci, *AAT1b, GLN4* and *RPN2* (Table 4.3), comprised an additional 1,051 nucleotides, 11 of which displayed SNPs (1.05%) amongst the 50 isolates of *C. dubliniensis* examined. The 10 sequence fragments (the Scheme B loci and the three additional loci) were all analysed together in a third scheme, termed Scheme C (Table 4.4), which was based on the sequences of the *AAT1a, AAT1b, ACC1, ADP1, GLN4, PMIb, RPN2, ALA1,* ex*VPS13,* and ex*ZWF1b* loci. The *GLN4* locus displayed four of the further 11 SNPs. However, it only gave rise to two genotypes amongst the 50 isolates investigated. Similarly, *AAT1b* displayed four SNPs, giving rise to five genotypes, and the *RPN2* locus contained three polymorphic sites which gave rise to three genotypes. Scheme C analysed a total of 4,336 nucleotides, 37 of which (0.85%) displayed SNPs, resulting in a total of 48 genotypes from 50 *C. dubliniensis* isolates. The minimum number of MLST loci for maximum discrimination amongst isolates of *C. dubliniensis* was determined according to the highest number of genotypes per variable base in each locus (Table 4.4) and is referred to as Scheme D, consisting of the 8 loci *AAT1b, ACC1, ADP1, PMIb, RPN2, ALA1,* ex*VPS13*, and ex*ZWF1b*. Polymorphic sites and resulting genotypes are summarised in Tables 4.4 and 4.5.

Table 4.4. Summary of loci used in individual MLST schemes

	MLST scheme	Loci												No. polymorphic sites	No. of DSTs	DP[a]
		AAT1a	*AAT1b*	*ACC1*	*ADP1*	*GLN4*	*PMIb*	*RPN2*	*ALA1*	*VPS13*	ex*VPS13*	*ZWF1b*	ex*ZWF1b*			
Length of analysed sequence		373	341	407	443	404	375	306	391	403	675	491	621			
No. (%) of polymorphic nucleotide sites		1 (0.27)	4 (1.17)	3 (0.74)	5 (1.13)	4 (0.99)	2 (0.5)	3 (0.98)	5 (1.28)	2 (0.5)	3 (0.44)	5 (1.02)	7 (1.13)			
No. of polymorphic amino acid sites		0	2	2	1	2	1	0	2	1	2	1	1			
No. of resulting genotypes		2	5	4	6	2	7	3	5	2	4	4	6			
	A	+	-	+	+	-	+	-	+	+	-	+	-	23	20	0.899
	B	+	-	+	+	-	+	-	+	-	+	-	+	26	22	0.901
	C	+	+	+	+	+	+	+	+	-	+	-	+	37	26	0.910
	D	-	+	+	+	-	+	+	+	-	+	-	+	32	25	0.909
	E	+	-	+	+	-	+	+	+	+	-	+	-	26	22	0.906

[a] DP; Discriminatory power, determined using Hunter's formula (Hunter, 1990).

Symbols; +/- denotes the presence/absence of individual loci in each MLST scheme.

Table 4.5. Summary of polymorphic sites and resulting genotypes in the loci used in *C. dubliniensis* MLST analysis

Locus	Length of analysed sequence	No. (%) of polymorphic nucleotide sites	No. of polymorphic amino acid sites	No. of resulting genotypes
AAT1a	373	1 (0.27)	0	2
AAT1b	341	4 (1.17)	2	5
ACC1	407	3 (0.74)	2	4
ADP1	443	5 (1.13)	1	6
GLN4	404	4 (0.99)	2	2
PMIb	375	2 (0.5)	1	7
RPN2	306	3 (0.98)	0	3
ALA1	391	5 (1.28)	2	5
VPS13	403	2 (0.5)	1	2
ex*VPS13*	675	3 (0.44)	2	4
ZWF1b	491	5 (1.02)	1	4
ex*ZWF1b*	621	7 (1.13)	1	6

4.3.2. Examination of stabilising selection on housekeeping MLST loci

The SNPs that occur in the MLST loci examined were used to generate a ratio of nonsynonymous to synonymous (dN/dS) amino acid changes in their encoded peptides. This ratio should be less than 1.0 if the genes investigated are under stabilising selection. The effect of each SNP on the resulting amino acid sequence was investigated by mapping the triplet codons for each gene fragment and examining the effect the SNP had on each codon in question. Nucleotide polymorphisms and amino acid substitutions are summarised in Table 4.6. Of the 37 SNPs identified in the 10 *C. dubliniensis* loci examined, substituting *VPS13* for ex*VPS13* and *ZWF1b* for ex*ZWF1b*, 13 (35%) resulted in nonsynonymous amino acid changes, while all of the remaining nucleotide changes resulted in synonymous polymorphisms. The dN/dS ratio of the MLST loci was therefore found to be 0.54, indicating that the loci investigated for use in *C. dubliniensis* MLST were under stabilising selection. Seven of the 13 nonsynonymous polymorphisms affected the resulting amino acid substantially, i.e, affected the charge (4/7) or polarity (3/7) of the amino acid.

4.3.3. Stability and reproducibility of *C. dubliniensis* MLST

Sequence analysis was performed in duplicate for each of two randomly selected isolates (Table 4.1) per MLST locus, using independently prepared DNA, PCR and sequencing reactions. For each MLST locus, duplicate sequences for each isolate showed 100% sequence identity, displaying full conservation of both SNPs and sites of heterozygosity at each locus.

4.3.4. Discriminatory power of *C. dubliniensis* MLST schemes

Examination of the *C. albicans* MLST locus set (Scheme A) in the 50 *C. dubliniensis* isolates investigated identified 20 unique DSTs based on the unique combinations of the genotype numbers for the seven loci examined. Application of Hunter's formula (Hunter, 1990) to this data set infers that MLST using Scheme A has a discriminatory power (D) of 0.899 when applied to *C. dubliniensis*, compared with a value of 0.996 when applied to *C. albicans* (Odds *et al.*, 2007). Extension of the *VPS13* and *ZWF1b* loci (Scheme B) resulted in a further 2 DSTs, and adding the *AAT1b, GLN4*, and *RPN2* loci (Scheme C; Table 4.4) resulted in a further 4 DSTs, giving a total of 26 DSTs from the 10 loci. Extending the *ZWF1b* and *VPS13* gene fragments and including the three other loci in Scheme C increased the discriminatory power to 0.9102

(Table 4.4). DST 4 was the most common DST identified using the Scheme A set of loci in *C. dubliniensis*, corresponding to 14 of the 50 isolates examined. The set of seven loci identified 11 DSTs that were unique to single *C. dubliniensis* isolates. When the larger set of loci was used in Scheme C, the same 14 isolates from DST 4 in Scheme A referred to above also gave an identical sequence type, now termed DST 6, which correlated with the previously identified DST 4, making this the most common DST in the scheme examined for *C. dubliniensis*. Using this larger set of loci, 19 of the 26 DSTs were unique to single isolates.

4.3.5. Population analysis of *C. dubliniensis* using MLST

A UPGMA dendrogram was constructed based on the sequence data from all 10 loci (scheme C) examined in *C. dubliniensis* using the Bionumerics version 5.1. software program (Fig. 4.1). At a cut off node of 99.7% sequence identity, the dendrogram revealed the presence of three major clades of isolates termed C1–C3, which showed a significant degree of correlation with the major clades previously identified by fingerprinting using the Cd25 fingerprint probe (Fig. 4.1 and Table 4.1). Clades C1 and C2 corresponded to the previously identified Cd25 fingerprint groups I and II, respectively, whereas clade C3 included strains previously identified as belonging to Cd25 fingerprint groups II and III, respectively (Fig. 4.1 and Table 4.1). Furthermore, clade C1 consisted solely of ITS genotype 1 isolates, clade C2 consisted solely of ITS genotype 2 isolates, whereas clade C3 consisted of ITS genotypes 3 and 4 isolates (Gee *et al.*, 2002).

4.3.6. Linkage disequilibrium and clonality

Genetic diversity and linkage disequilibria were assessed by using statistics implemented in the Multilocus 1.3 software (Agapow & Burt, 2001) and genotypes obtained for all loci investigated in this study (Scheme C). Each of these statistics tested the null hypothesis of a freely recombining population. Highly significant linkage disequilibria ($p < 10^{-5}$ with 100,000 randomisations) were found for both the total collection of 50 *C. dubliniensis* isolates and a reduced collection of 26 isolates that represented each of the 26 DSTs defined by Scheme C (i.e., that did not contain repeated genotypes) (data not shown). These results provide evidence that the sample of *C. dubliniensis* isolates analysed in this study represents a clonal population.

Table 4.6. Amino acid substitutions in *C. dubliniensis* MLST loci

Locus	Polymorphic nucleotide position [a]	Triplet polymorphism[b]	Amino Acids[c]	Corresponding codon in *C. albicans*	Corresponding amino acid in *C. albicans*
AAT1a	2	(T/C)TG	L/L	TTG	L
AAT1b	105	A(T/A)T	I/N	ATC	I
	114	TC(A/T)	S/S	TCA	S
	121	(G/A)AA	E/K	GAA	E
	129	TC(T/G)	S/S	TCT	S
	198	**TC(T/K)**	**S/S**	**TCT**	**S**
ACC1	128	TT(T/C)	F/F	TTC	F
	313	T(T/C)G	L/S	CTG	L
	389	AA(A/T)	K/N	AAA	K
ADP1	31	AT(C/T)	I/I	ATT	I
	214	AG(C/T)	S/S	AGT	S
	234	G(A/G)G	E/G	GAG	E
	331	TG(T/C)	C/C	TGT	C
	421	TC(G/T)	S/S	TCA	S
GLN4	104	GA(A/T)	E/D	GTA	V
	141	AA(T/C)	N/N	AAA	K
	147	GA(T/C)	D/D	GAC	D
	244	(A/G)TT	I/V	ATT	I
PMIb	101	A(G/A)A	R/K	AGA	R
	375	GC(A/G)	A/A	GCA	A
RPN2	127	TC(C/A)	S/S	TCC	S
	136	GC(C/T)	A/A	GCC	A
	298	GC(T/C)	A/A	GCA	A
ALA1	82	AA(C/T)	N/N	AAT	N
	148	TT(A/G)	L/L	TTA	L
	<u>186</u>	**<u>AA(T/C)</u>**	**<u>N/N</u>**	**<u>AAT</u>**	**<u>N</u>**
	203	G(T/C)T	V/A	GTT	V
	383	(T/G)CT	S/A	GCT	A
VPS13	397[d]	AA(A/G)	K/K	AAG	K
	402[e]	T(G/C)G	W/S	TGG	W
ex*VPS13*	45	GA(G/C)	E/D	GAC	D
	211	**(G/T)GT**	**G/C**	**GGT**	**G**
	654[d]	AA(A/G)	K/K	AAG	K
	659[e]	T(G/C)G	W/S	TGG	W
ZWF1b	1[f]	AC(C/T)	T/T	AAA	K
	28[g]	AC(G/A)	T/T	ACC	T
	265[h]	AA(G/A)	K/K	AAA	K
	299[i]	(G/T)AT	D/Y	GAT	D
	334[j]	AA(A/G)	K/K	AAA	K
ex*ZWF1b*	78[f]	AC(C/T)	T/T	AAA	K
	105[g]	AC(G/A)	T/T	ACC	T
	342[h]	AA(G/A)	K/K	AAA	K
	376[i]	(G/T)AT	D/Y	GAT	D
	411[j]	AA(A/G)	K/K	AAA	K
	557	**A(A/R) A**	**K/R**	**AAA**	**K**
	573	TA(T/C)	Y/Y	TAT	Y
	588	TC(A/T)	S/S	TCA	S

[a] Nucleotide positions refer to the positions in the analysed fragment only.

[b] Position of the polymorphism relative to its codon.

[c] Amino acid resulting from codon both before and after polymorphism.

[d-j] Identical polymorphic sites in the both the extended (ex) gene segments analysed and in the shorter sequences that were based on the *C. albicans* scheme lengths.

The nucleotide site shown in bold and underlined typeface denotes the only common nucleotide site that displays polymorphism in both *C. albicans* and *C. dubliniensis.*

Nucleotide sites shown in bold typeface denotes polymorphisms that were observed in the loci of the recommended *C. dubliniensis* MLST scheme after its establishment, and using different isolates to those used in the development of the scheme.

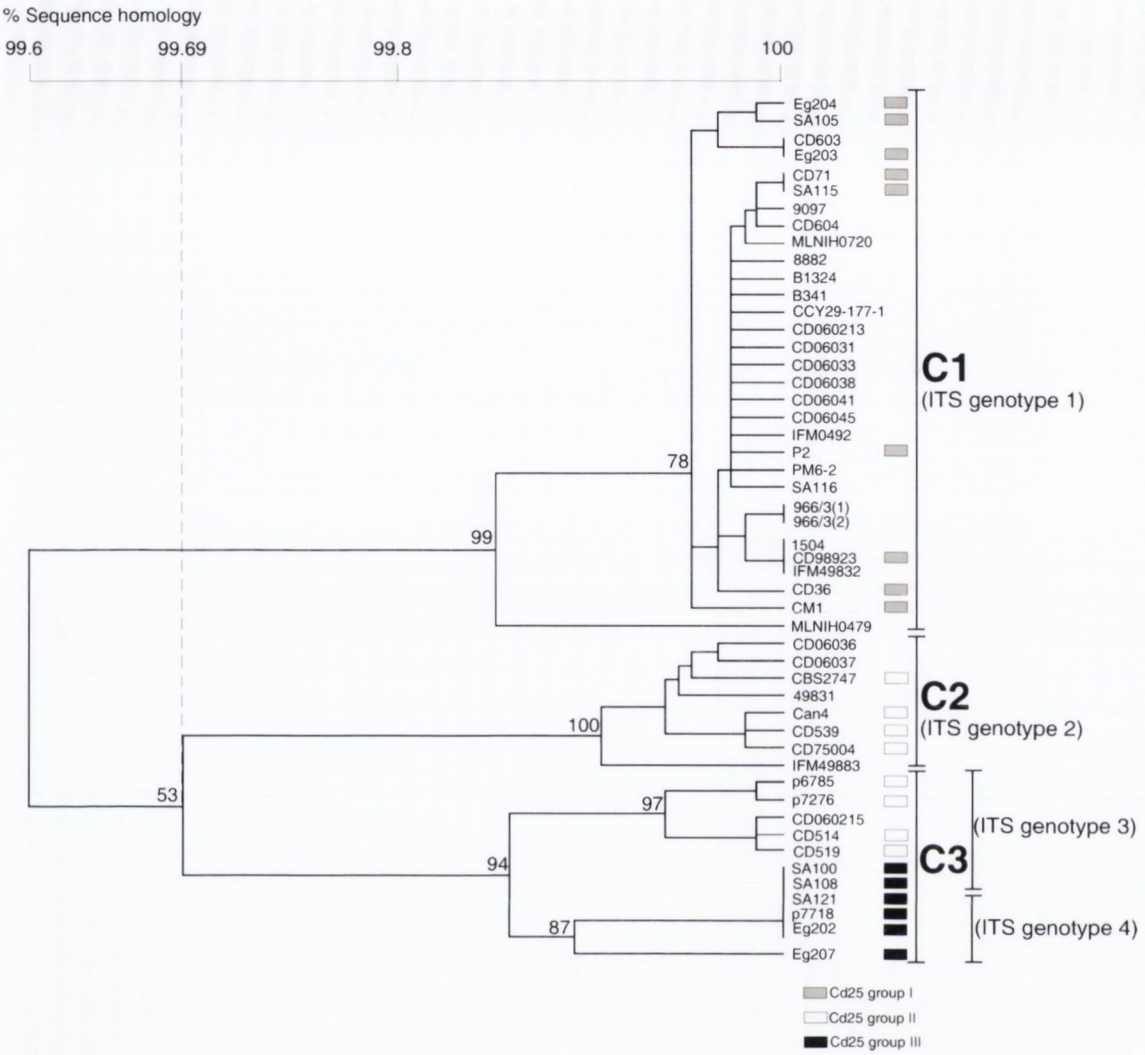


Figure 4.1. Population structure of *C. dubliniensis* defined by MLST.

UPGMA dendrogram based on concatenated sequences from the MLST Scheme C loci; *AAT1a, ACC1, ADP1, PMIb, ALA1,* ex*VPS13,* ex*ZWF1b, RPN2, GLN4* and *AAT1b* (Table 4.4) showing percentage sequence identity for 50 *C. dubliniensis* isolates from a broad range of geographical locations, including isolates from all known ITS genotypes (Table 4.1). Three distinct major clades termed clades C1–C3 are evident at 99.69 % sequence homology. C1 consists of ITS genotype 1 isolates exclusively. C2 consists of ITS genotype 2 isolates exclusively, and C3 consists of ITS genotype 3 and 4 isolates. Twenty-three of the 50 isolates included in the study were previously fingerprinted with the *C. dubliniensis*-specific probe Cd25 (Gee *et al.*, 2002; Al Mosaid *et al.,* 2005) and the Cd25 fingerprint groups of these isolates are indicated by shaded rectangular boxes.

4.3.7. Comparative population structure analysis of *C. albicans* and *C. dubliniensis*

Candida albicans MLST sequences were available for the seven consensus MLST loci (i.e., *AAT1a, ACC1, ADP1, PMIb, ALA1, VPS13, ZWF1b*) and an extra locus *RPN2*. The resulting 3,189 nucleotides from the eight loci resulted in a fifth scheme, termed Scheme E (Table 4.4). Sequences for all loci were concatenated and treated as one sequence for each of 50 *C. albicans* isolates and the corresponding sequences in 50 *C. dubliniensis* isolates in order to allow comparison between the two species using MLST. These concatenated sequences were used to construct a maximum parsimony tree in order to assess comparative phylogenies between the species (Fig. 4.2). The maximum parsimony tree displays the comparative divergence within the two species, as well as the level of relatedness between the two species. Previous Ca3 fingerprinting and MLST studies have shown that *C. albicans* consists of many major and minor clades (Blignaut *et al.*, 2002; Pujol *et al.*, 2002; Soll & Pujol, 2003; Odds *et al.*, 2007), whereas MLST data obtained in the present study showed that the *C. dubliniensis* population is significantly less diverse (Fig. 4.2). The maximum parsimony tree demonstrated that *C. albicans* isolates belonging to different MLST clades can differ at as many as 90 nucleotide sites, whereas isolates belonging to the same MLST clade can differ at as many as 31 nucleotide sites (data not shown). In contrast, the present data showed that *C. dubliniensis* consisted of three closely related clades, termed clades C1, C2 and C3 (Fig. 4.2), that show complete agreement with the clades described by the UPGMA dendrogram (Fig. 4.1). *Candida dubliniensis* isolates from the three MLST clades differ from each other at a minimum of 10 nucleotide sites (range 10–23 nucleotides). Isolates belonging to clade C1 differ at a maximum of 10 nucleotide sites, isolates belonging to clade C2 differ at a maximum of six nucleotide sites and isolates belonging to clade C3 differ at a maximum of eight nucleotide sites (data not shown). Overall, 257 nucleotides of the 3,189 nucleotides analysed (8%) were identified as being different between the two species, correlating with the level of sequence identity typically exhibited between the two species (Moran *et al.*, 2004).

A UPGMA dendrogram was constructed on the basis of the consensus *C. albicans* MLST scheme loci (Scheme A, Table 4.4), for the same set of *C. dubliniensis* and *C. albicans* isolates. This was based on allelic profiles and corresponding DSTs rather than concatenated sequence data (Fig. 4.3). One representative for each DST obtained from these isolates was included in the analysis, totalling 20 *C. dubliniensis*

DSTs and 50 *C. albicans* DSTs. This UPGMA dendrogram also displays the comparative divergence amongst the two species (Fig. 4.3), and correlates with the findings of the maximum parsimony analysis (Fig. 4.2), that the population structure of *C. dubliniensis* is significantly less divergent than that of *C. albicans.*

4.3.8. The recommended MLST scheme for use in *C. dubliniensis*

The minimum number of MLST loci necessary for maximum discrimination amongst isolates of *C. dubliniensis* was determined according to the highest number of genotypes per variable base in all loci investigated for use in *C. dubliniensis*. The recommended scheme is therefore *AAT1b, ACC1, ADP1, PMIb, RPN2, ALA1,* ex*VPS13,* and ex*ZWF1b* (Scheme D, Table 4.4). This scheme identified 32 SNPs amongst the 50 isolates, 21 of which encoded synonymous amino acid substitutions, and 11 of which encoded nonsynonymous substitutions. The dN/dS ratio was therefore 0.524, indicating that the loci in the scheme were under stabilising selection. The discriminatory power of this scheme was found to be 0.909 amongst the 50 *C. dubliniensis* isolates (Table 4.4).

4.3. Results II: Use of the established *C. dubliniensis* MLST scheme

4.3.9. Genetic relationship between avian-excrement-associated and human *C. dubliniensis* isolates

Three new isolates of *C. dubliniensis* were obtained from 134 faecal samples as described (Chapter 3, Section 3.2.2). Like the Great Saltee isolates (Nunn *et al.*, 2007), the three new isolates were shown to belong to ITS genotype 1 (Gee *et al.*, 2002). The 14 avian-excrement-associated isolates were compared with 31 human *C. dubliniensis* strains belonging to MLST clade C1 (Fig. 4.1), and five additional *C. dubliniensis* clade C1 human isolates from Ireland – CD505, CD06032, CD06027, CD0512 and CD524 (Table 4.1). Isolates were assigned DSTs based on the genotype numbers for the 8 loci in the recommended *C. dubliniensis* MLST scheme, Scheme D, (Table 4.4). Six new DSTs were identified in 13/14 avian-excrement-associated isolates due to the identification of two new *ZWF1b* genotypes that occurred exclusively in the avian-excrement-associated isolates. The only previously identified DST found amongst the avian-excrement-associated isolates was DST 2 (isolate AV7, Table 4.7). The most frequently found DST in avian-excrement-associated *C. dubliniensis* isolates was DST

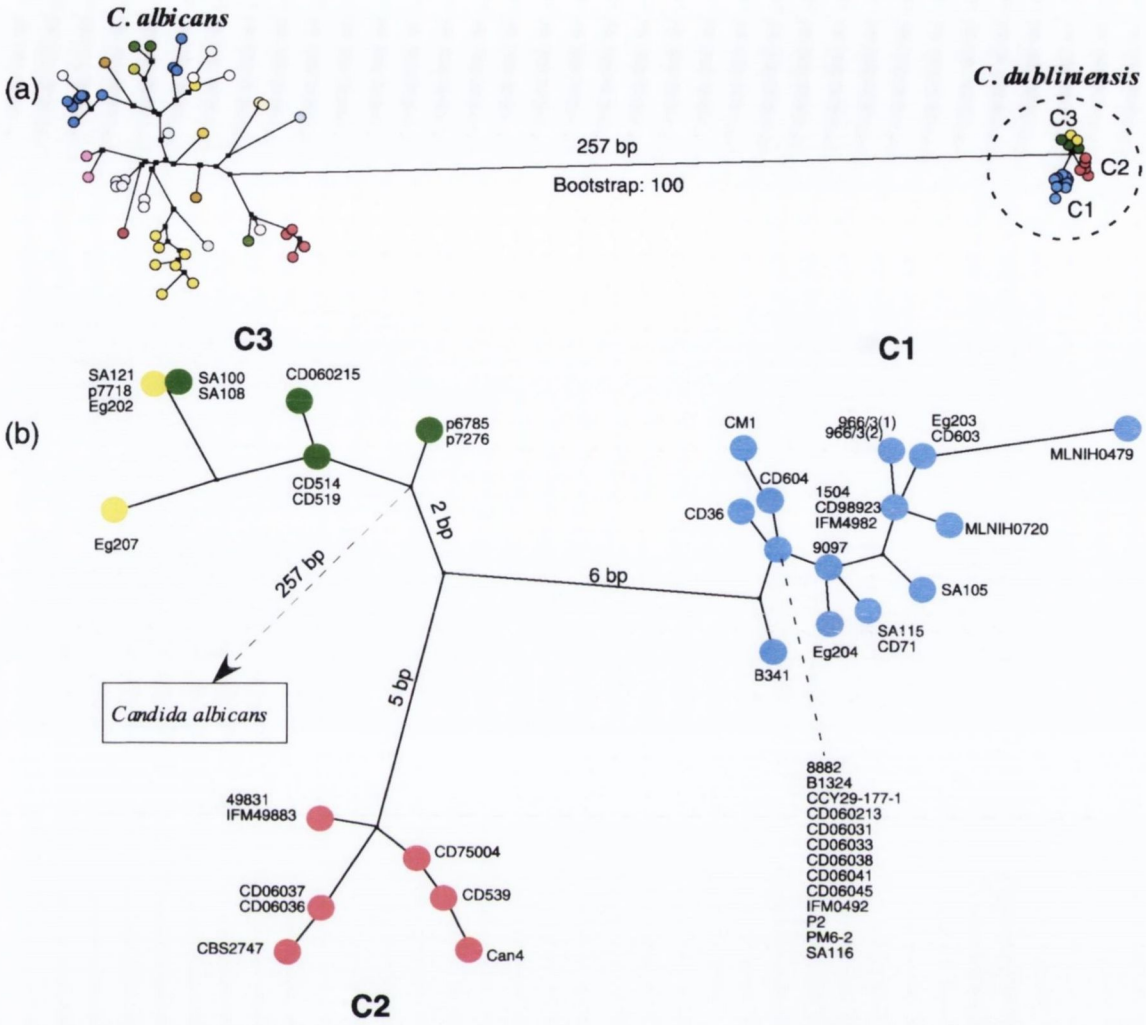


Figure 4.2. Comparative population structure analyses of *C. dubliniensis* and *C. albicans* based on concatenated MLST sequences.

Maximum parsimony tree showing the comparative divergence between 50 isolates each of *C. albicans* and *C. dubliniensis* based on concatenated sequences from the MLST loci *AAT1a, ACC1, ADP1, PMIb, ALA1, VPS13, ZWF1b* and *RPN2* in MLST scheme E. *Candida dubliniensis* isolates were selected from a diverse range of geographic locations (Fig. 3.1) and from all four ITS genotypes (Table 4.1). *Candida albicans* isolates were selected as representatives of the MLST clades recently described by Odds *et al.* (2007) (Table 4.2). Panel (a) Comparative divergence between the *C. albicans* and *C. dubliniensis* isolates tested showing that the two species are separated by 257 bp differences. The *C. dubliniensis* isolates formed three closely related groups of isolates (C1–C3), which correspond to those identified in the UPGMA dendrogram shown in Fig. 4.1. Isolates from distinct clades are highlighted by specific colours for each species. Panel (b) shows an enlarged view of the three *C. dubliniensis* major clades encircled in panel (a). Clade C1 consists exclusively of ITS genotype 1 isolates (blue), clade C2 consists exclusively of ITS genotype 2 isolates (red), and clade C3 consists of ITS genotype 3 isolates (green) and ITS genotype 4 isolates (yellow).

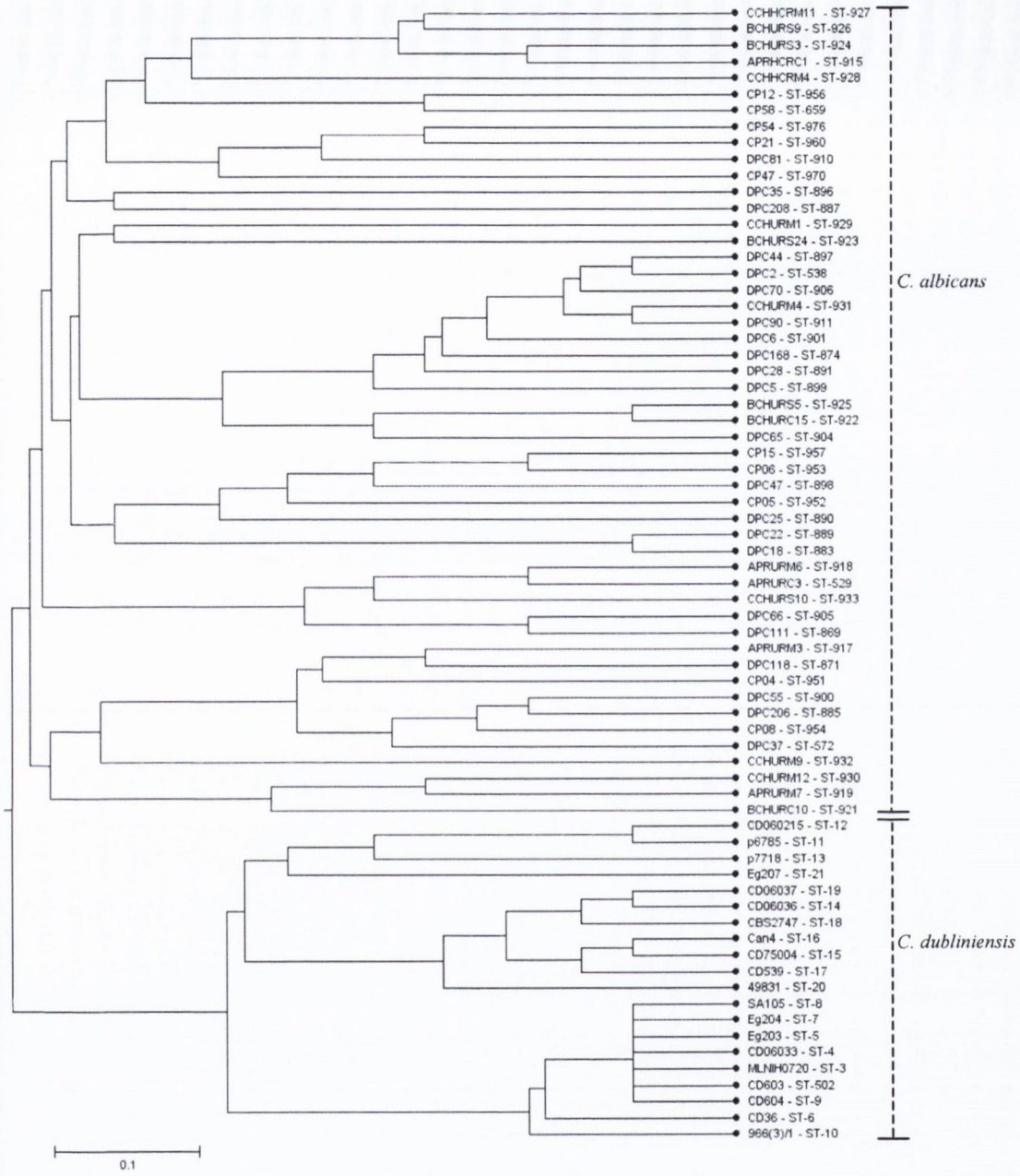


Figure 4.3. Comparative population structure analyses of *C. dubliniensis* and *C. albicans* based on MLST allelic profiles and DST data.

UPGMA dendrogram based on the allelic profiles and corresponding DSTs for 50 isolates of *C. albicans* and 50 isolates of *C. dubliniensis* analysed using the consensus *C. albicans* MLST scheme (Scheme A, Table 4.4) This dendrogram includes only one representative isolate for each DST in both *C. albicans* and *C. dubliniensis* populations. The scale bar denotes p-distance, isolates and associated DSTs are indicated at each branch. The UPGMA tree correlates with the maximum parsimony tree (Fig. 4.2), in that, the population structure of *C. dubliniensis* is significantly less divergent than *C. albicans*.

Table 4.7. Summary of MLST, *MTL* type and *CDR1* polymorphism analyses in *C. dubliniensis* isolates

Human isolate	Origin[a]	ITS Genotype	Cd25 group	MLST clade[b]	MLST DST[c] (scheme D)	MTL type	*CDR1* polymorphism[d]
CD36	Ireland	1	I	1	8	aα	Pos
CD06033	Ireland	1	ND	1	7	aα	Pos
CD06031	Ireland	1	ND	1	7	aα	Pos
CD060213	Ireland	1	ND	1	7	aα	Het
CD06041	Ireland	1	ND	1	7	aα	Het
CD06038	Ireland	1	ND	1	7	aα	Pos
CD06045	Ireland	1	ND	1	7	aα	Het
CD06037	Ireland	2	ND	2	19	aα	Neg
CD06036	Ireland	2	ND	2	19	aα	Neg
CD060215	Ireland	3	ND	3	24	aα	Neg
CD604	France	1	ND	1	12	aα	Neg
CD603	France	1	ND	1	3	aa	Neg
CM1	Australia	1	I	1	2	aα	Het
SA105	S. Arabia	1	I	1	11	αα	Neg
SA115	S. Arabia	1	I	1	2	aα	Neg
SA116	S. Arabia	1	I	1	7	aa	Neg
SA108	S. Arabia	3	III	3	14	aα	Het
SA100	S. Arabia	3	III	3	14	aα	Neg
SA121	S. Arabia	4	III	3	14	αα	Neg
Eg202	Egypt	4	III	3	14	aα	Neg
Eg203	Egypt	1	I	1	6	aα	Het
Eg204	Egypt	1	I	1	10	aα	Pos
Eg207	Egypt	4	III	3	21	aα	Neg
p6785	Israel	3	II	3	22	aα	Neg
p7276	Israel	3	II	3	22	aa	Neg
p7718	Israel	4	III	3	14	αα	Neg
CD71	Argentina	1	I	1	2	aα	Neg
CD98923	India	1	I	1	5	aα	Neg
B1324	USA	1	ND	1	7	aα	Neg
B341	USA	1	ND	1	7	aα	Pos
PM6-2	Chile	1	ND	1	7	aα	Pos
P2	Switzerland	1	I	1	7	αα	Het
1504	Slovakia	1	ND	1	5	aα	Het
8882	Slovakia	1	ND	1	7	aα	Neg
9097	Slovakia	1	ND	1	9	aa	Neg
966/3(1)	Slovakia	1	ND	1	13	aα	Pos
966/3(2)	Slovakia	1	ND	1	13	aα	Pos
CCY29-177-1	Slovakia	1	ND	1	7	aα	Pos
MLNIH0479	Thailand	1	ND	1	6	aa	Neg

Continued overleaf

Table 4.7. ***Continued***

Human isolate	Origin[a]	ITS Genotype	Cd25 group	MLST clade[b]	MLST DST[c] (scheme D)	MTL type	*CDR1* polymorphism[d]
MLNIH0720	Thailand	1	ND	1	4	aα	Neg
49831	Japan	2	ND	2	20	aα	Neg
IFM49883	Japan	2	ND	2	26	aα	Neg
IFM0492	Thailand	1	ND	1	7	aα	Neg
IFM49832	Japan	1	ND	1	5	aα	Neg
Can4	Canada	2	II	2	16	aα	Neg
CD539	U.K	2	II	2	17	aα	Neg
CBS2747	Netherlands	2	II	2	18	aα	Neg
CD514	Ireland	3	II	3	25	aα	Neg
CD519	Ireland	3	II	3	23	aα	Neg
CD75004	UK	2	II	2	15	αα	Neg
98-277	Spain	1	ND	1	2	αα	Pos
95-677	Spain	1	ND	1	7	aα	Neg
94-234	Spain	1	ND	1	32	αα	Pos
CD505	Ireland	1	I	1	28	αα	Pos
CD06032	Ireland	1	ND	1	36	αα	Neg
CD06027	Ireland	1	ND	1	1	aα	Pos
CD0512	Ireland	1	ND	1	37	aα	Neg
CD524	Ireland	1	I	1	35	aα	Neg
17P	Spain	2	ND	2	34	aα	Neg
18P	Spain	2	ND	2	34	aα	Neg
22BP	Spain	2	ND	2	34	aα	Neg
25P	Spain	2	ND	2	34	aα	Neg
110P	Spain	2	ND	2	34	aα	Neg
140P	Spain	2	ND	2	34	aα	Neg
Avian *C. dubliniensis* isolates							
SL411	*I. uriae*	1	ND	1	27	aa	Pos
SL422	*I. uriae*	1	ND	1	27	aa	Pos
SL370	*I. uriae*	1	ND	1	27	aa	Pos
SL410	*I. uriae*	1	ND	1	29	aa	Pos
SL375-I	*I. uriae*	1	ND	1	31	aa	Pos
SL375-II	*I. uriae*	1	ND	1	31	aa	Pos
SL397	*I. uriae*	1	ND	1	31	aa	Pos
SL414	*I. uriae*	1	ND	1	31	aa	Pos
SL495	*I. uriae*	1	ND	1	33	aa	Pos
SL509	*I. uriae*	1	ND	1	30	aa	Pos
SL522	*I. uriae*	1	ND	1	31	aa	Pos
AV5	*L. argentatus*	1	ND	1	29	aa	Pos
AV6	*L. argentatus*	1	ND	1	27	aa	Pos
AV7	*L. argentatus*	1	ND	1	2	aα	Pos

[a] Origin referring to the countries from which human *C. dubliniensis* isolates were recovered. For avian-associated isolates, the origin refers to the avian species from which the isolate was recovered.

[b] MLST clades assigned in the present study.

[c] DSTs assigned to each isolate according to the recommended MLST scheme (Scheme D) for *C. dubliniensis*

[d] *CDR1* polymorphism refers to the presence (Pos), absence (Neg) or heterozygous presence (het) of the TAG polymorphism in the *CDR1* gene that occurs in 58 % of human ITS genotype 1 isolates, as determined according to Moran *et al.* (2002).

Abbreviations; S. Arabia; Saudi Arabia; *I. uriae; Ixodes uriae* (tick); *L. argentatus; Larus argentatus* (herring gull); ND; Not determined.

31 (i.e., 5/14 isolates), all 5 of which were originally recovered from *I. uriae* ticks by Nunn *et al.* (2007). Four isolates belonged to DST 27, three of which were originally recovered from *I. uriae* and one of which was recovered from herring gull excrement in Dublin (Table 4.7).

Polymorphism sites (n=36) for the 8 MLST loci were concatenated and used to construct a neighbour joining tree (Mega software program version 3.1 [Kumar *et al.*, 2004]) including all 50 clade C1 human and avian-excrement-associated *C. dubliniensis* isolates. Nearly all (13/14) avian-excrement-associated *C. dubliniensis* isolates, including 11 isolates recovered from *I. uriae* (Nunn *et al.*, 2007) and 2/3 avian-excrement-associated isolates recovered in Dublin, formed a distinct subgroup within the C1 clade (Fig. 4.4). This same subgroup was also identified in trees generated using the UPGMA and maximum parsimony clustering algorithms. The remaining avian-excrement-associated isolate, AV7, was indistinguishable from three other human isolates, CD71, SA115 and CM1 by the MLST method (Fig. 4.4). To test for genetic separation between the human and avian-excrement-associated isolates recovered from the same country, a neighbour joining tree was constructed using 13 Irish avian and human clade C1 isolates, each of which represented unique DSTs. The tree displays the robustness of the avian-excrement-associated subgroup of isolates within a population of human isolates recovered from the same country (Fig. 4.5). Fisher's exact testing (http://www.exactoid.com/fisher/index.php) based on nearest neighbour analysis found the distribution of the avian-excrement-associated and the human isolates to differ significantly ($p<0.05$).

4.3.10. Analysis of clonal complexes and prediction of founder genotypes

In order to further determine the genetic relationship between human and avian-excrement-associated isolates of *C. dubliniensis*, a BURST analysis was undertaken analysing the DSTs of 13 Irish avian-excrement-associated and human clade C1 isolates, each of which represented unique DSTs (Fig. 4.6). The group definition set for the analysis ensured that each DST matched at least one other DST in the same group at six or more of the eight consensus *C. dubliniensis* loci analysed. The analysis predicted three clonal complexes, of which one was formed by avian-excrement-associated DSTs exclusively (Fig. 4.6). The predicted founder of this avian-excrement-associated clonal complex was DST 31. The other two clonal complexes consisted of human isolate DSTs with the predicted founder DSTs identified as 7 and 28,

respectively. One of the avian-excrement-associated isolates AV7 (DST 2) grouped into a clonal complex with human isolates, which originated from the predicted founder DST 7 (Fig. 4.6).

4.3.11. Prevalence of the TAG polymorphism in the *CdCDR1* gene

The prevalence of a common point mutation previously identified in the *CDR1* gene of some ITS genotype 1 isolates was determined in the avian-excrement-associated and human isolates of *C. dubliniensis*. All avian-excrement-associated *C. dubliniensis* were found to contain the TAG polymorphism (Table 4.7), while 19/36 (52.7%) of the human clade C1 (ITS genotype 1) isolates were found to contain the polymorphism, which correlated with previous studies (Moran *et al.*, 2002). Of these 19 human isolates, 7 were heterozygous for the TAG polymorphism, the remaining 12 were homozygous. The polymorphism was not found in any *C. dubliniensis* isolates that belonged to clades C2 or C3 which were made up of ITS genotype 2, 3 and 4 isolates (Table 4.7). All 14 of the avian-excrement-associated *C. dubliniensis* isolates were found to contain the TAG polymorphism, and in each of these isolates the polymorphism was homozygous. Fisher's exact testing showed that the distribution of the TAG polymorphism differed significantly ($p<0.005$) between isolates that were recovered from either humans or avian-excrement-associated sources.

4.3.12. Analysis of mating types in the population

Of the 14 avian *C. dubliniensis* isolates studied, 13 were found to be *MTL*a homozygous (a/a), and the remaining avian-excrement-associated *C. dubliniensis* isolate (AV7) was *MTL* heterozygous (a/α) (Table 4.7). Of the 36 other human isolates belonging to the same MLST clade (C1), only 4/36 (11.1%) isolates were found to be *MTL*a homozygous, whereas 28/36 (77.7%) were *MTL* heterozygous (a/α) (Table 4.7). Fisher's exact testing showed that the distribution of these *MTL* types differed significantly ($p<0.0005$) between isolates that were recovered from either humans or avian-excrement-associated sources. Of the 64 human isolates that were included in the current MLST study from all three MLST clades (C1–C3), 5/64 (7.8%) were *MTL*a homozygous, (a/a) and 9/64 (14.1%) were *MTL*α homozygous (α/α). The remaining 50/64 (78.1%) were *MTL* heterozygous (Table 4.7).

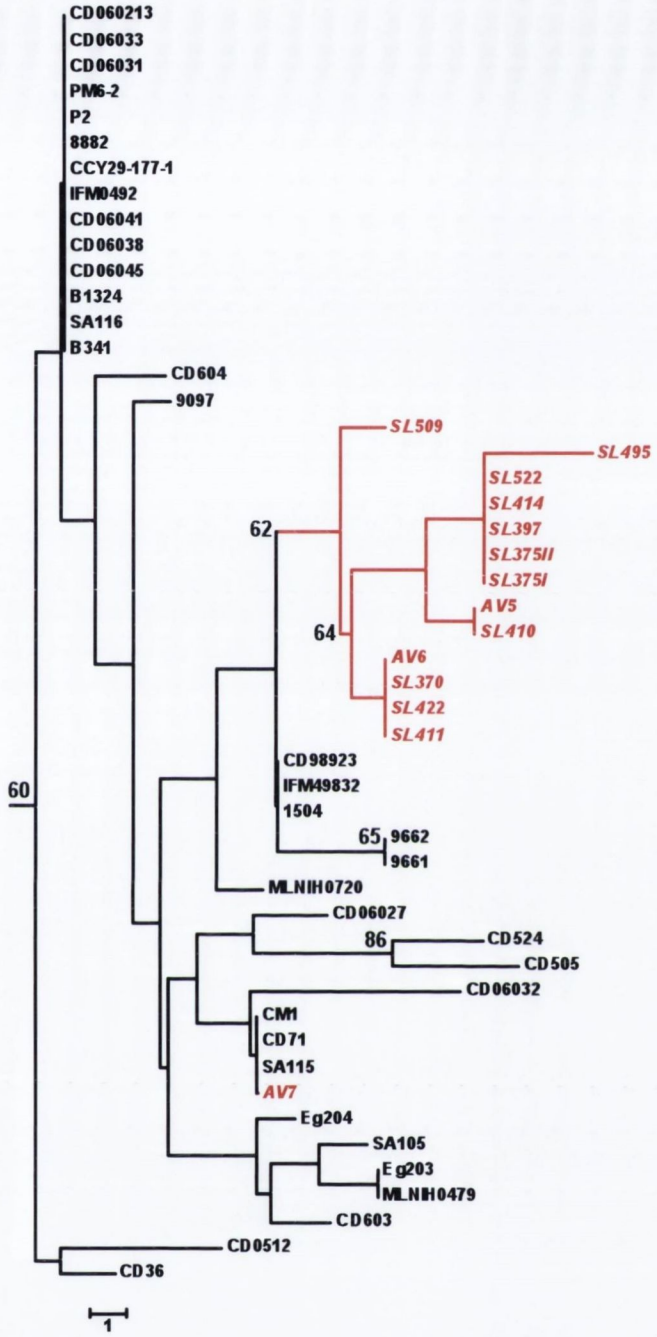


Figure 4.4. Genetic relationships of avian-excrement-associated and human *C. dubliniensis* isolates.

Neighbour joining tree based on the polymorphic sites in *C. dubliniensis* MLST sequences. Bootstrap values greater than 60 % are indicated at cluster nodes. Avian-excrement-associated isolates are highlighted with red italicised typeface. The numbers of polymorphic sites in isolates are indicated by the scale bar. The neighbour-joining tree depicts the isolates of MLST clade C1 as previously defined (Section 4.3.5) and displays the localisation of the avian-excrement-associated isolates in relation to isolates in the same clade.

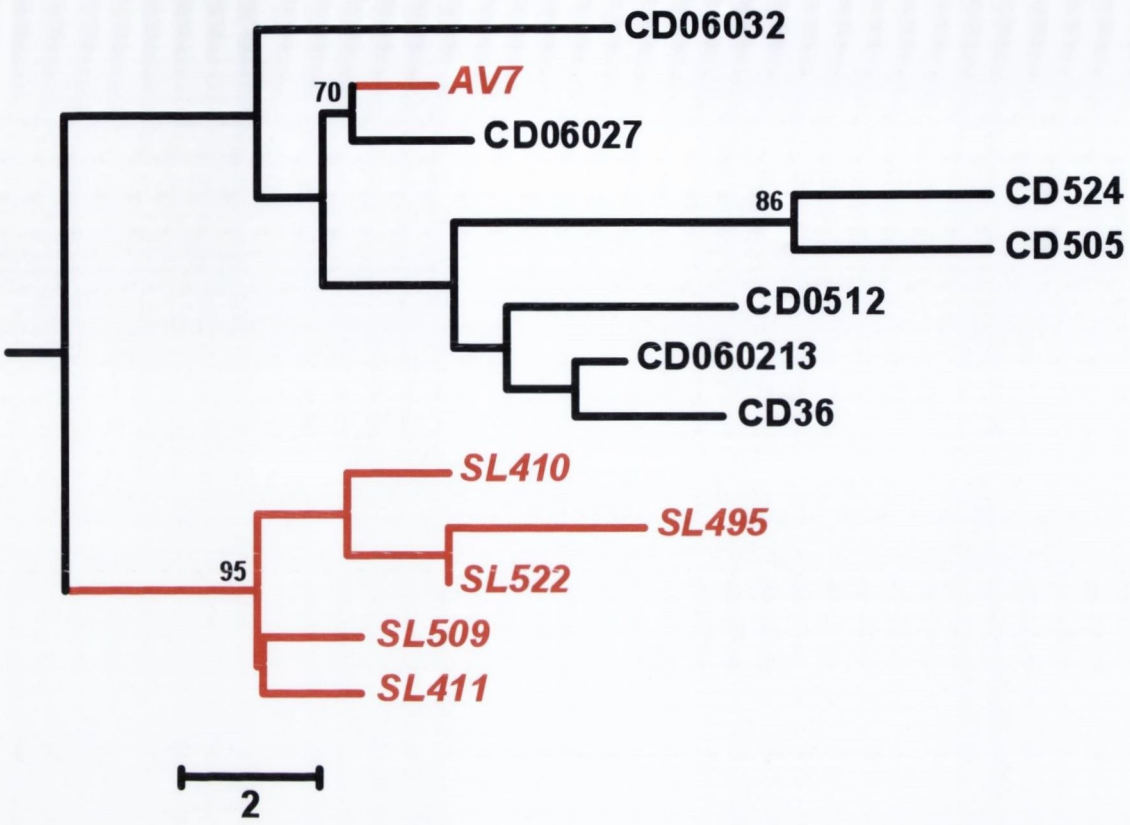


Figure 4.5. Genetic relationship between avian-excrement-associated and human *C. dubliniensis* isolates recovered in Ireland only.

Neighbour joining tree based on the polymorphic sites in MLST sequences for each of 13 ITS genotype 1 *C. dubliniensis* isolates, seven of which were recovered from humans in Ireland and six of which were recovered from seabird excrement in Ireland. Avian-excrement-associated isolates are highlighted with red italicised typeface. Bootstrap values greater than 60 % are indicated at cluster nodes. The numbers of polymorphic sites in isolates are indicated by the scale bar. Isolates which had identical DSTs were not included in the tree so that only one of each DST is included in order to reduce bias. This tree displays the robustness of the avian-excrement-associated subgroup of isolates within a population of similar human *C. dubliniensis* isolates recovered from the same region.

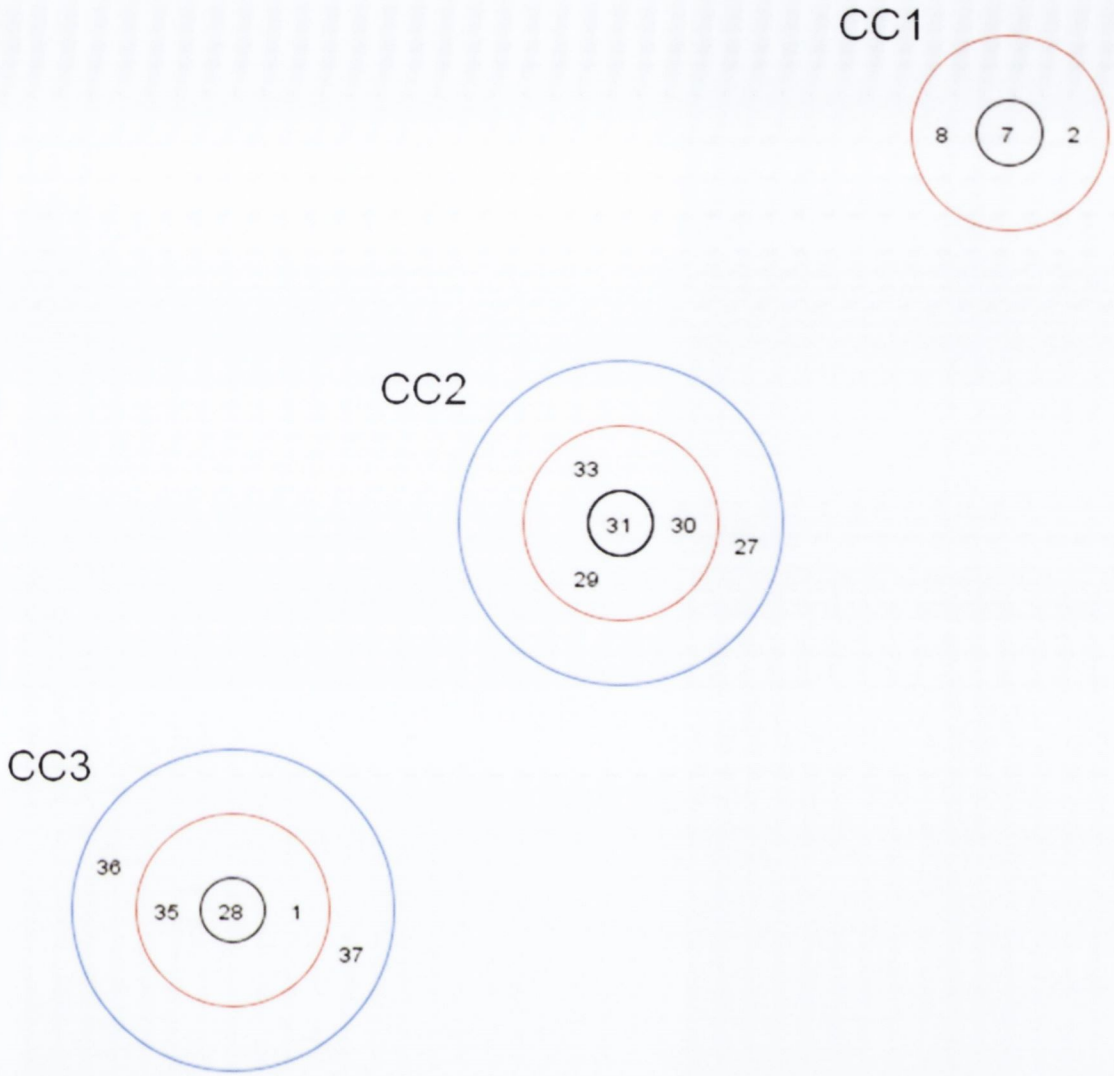


Figure 4.6. BURST analysis of avian-excrement-associated and human *C. dubliniensis* isolates recovered in Ireland only.

BURST analysis based on the allelic profiles and resulting DSTs for both avian and human-associated ITS genotype 1 isolates of *C. dubliniensis*, all of which were recovered in various locations in Ireland. Only one of each individual DST was included in the analysis. Predicted founder DSTs are displayed as the inmost black circles, single locus variants in the inner red circles, and double locus variants in outer blue circles for each of three clonal complexes (CC1–CC3) observed within this group of isolates. In this panel, the avian-associated isolates are observed to form a clonal complex (CC2) that is distinct from the other two human-associated clonal complexes CC1 and CC3, with the exception of AV7 (DST 2), which also clusters in CC1. BURST analysis was carried out using START software version 2 (Jolley *et al.*, 2001).

4.3.13. Analysis of *C. albicans* recovered from avian-excrement-associated sources

Of the 24 isolates of *C. albicans* that were recovered from seabirds as previously described (Chapter 3, Section 3.3.2), 17 were analysed using the *C. albicans* consensus MLST loci (Bougnoux *et al.*, 2003). Nine new DSTs were identified from these 17 isolates according to the standard *C. albicans* MLST scheme (Table 4.2) (Bougnoux *et al.*, 2003). The most common DST recovered from five of the avian *C. albicans* isolates was the newly identified DST 1243, a single locus variant of three previously defined human DSTs. Interestingly, one of the isolates AV115 (DST 1248) is a SLV of DST 1069 at the *AAT1a* locus. DST 1069 is shared by two *C. albicans* isolates recovered from starlings in France in 1993 and 2000, respectively. Only four of the newly recovered avian-excrement-associated *C. albicans* isolates were assigned previously designated DSTs, three of these were identified as DST 52 (Table 4.2) and the fourth was identified as DST 538 (Table 4.2).

4.3.14. Summary of the established *C. dubliniensis* MLST scheme

Since the establishment of the MLST scheme (Scheme D, Table 4.4) a further 28 *C. dubliniensis* isolates have been studied, resulting in a total of 78 isolates (Table 4.1). This number includes 14 ITS genotype 1 isolates recovered from seabird excrement, as well as nine *C. dubliniensis* isolates with unusual phenotypes (as previously mentioned in Section 4.2.1) which were sent to this laboratory by Dr. José Pontón. Three of these nine isolates were of ITS genotype 1 and localised to MLST clade C1 using the recommended *C. dubliniensis* MLST scheme. The remaining six isolates were ITS genotype 2, had identical DSTs (DST 34), and localised to MLST clade C2 (Table 4.7).

Upon the inclusion of the further 28 isolates, a further five SNPs have been identified amongst the Scheme D loci, giving a total of 37 SNPs identified to date. A further six genotypes have been identified since establishment of the MLST scheme, giving a new total of 46 genotypes from the eight loci. These additional genotypes gave rise to a further 12 DSTs, resulting in a new total of 37. No additional MLST clades have been identified. The MLST clades and DSTs assigned to each isolate using MLST Scheme D are listed in Table 4.7 and are displayed in Fig. 4.7.

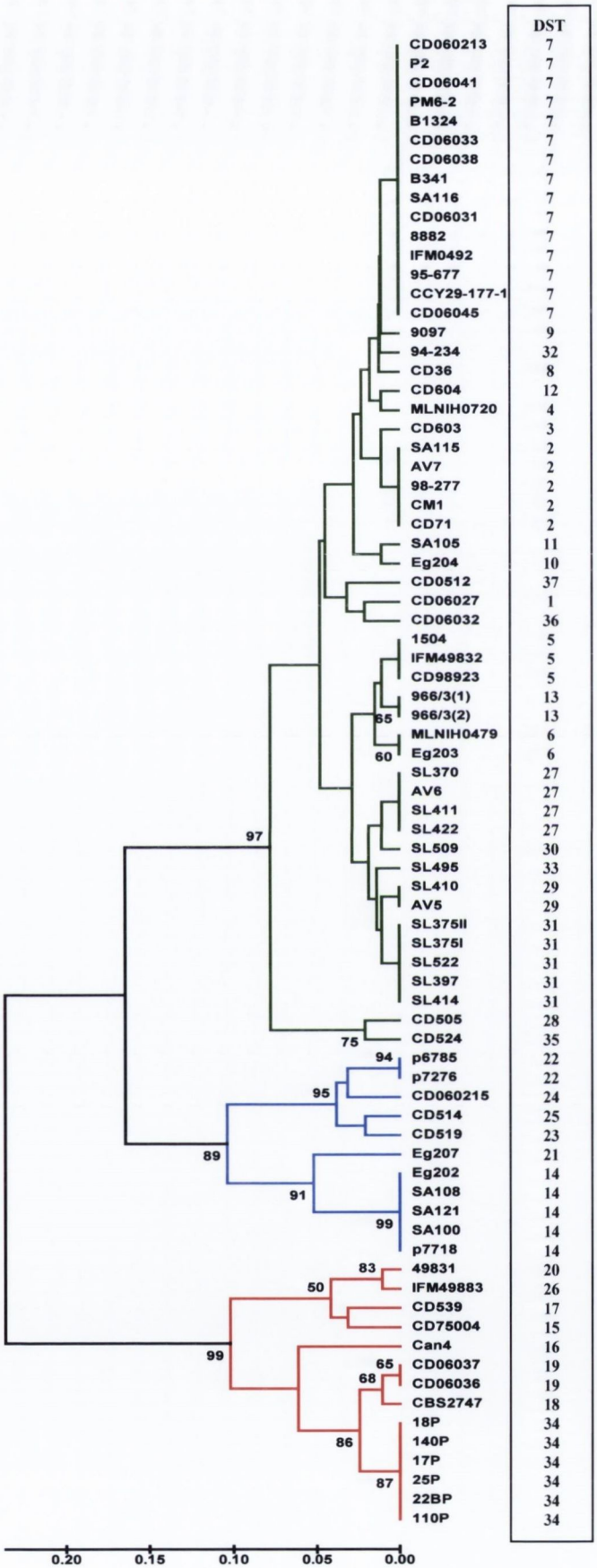
4.4. Discussion

Multilocus sequence typing has previously been shown to be a useful tool in the analysis of the epidemiology and population structure of *C. albicans* (Bougnoux *et al.*, 2002, 2003, 2006; Tavanti *et al.*, 2003, 2005a; Odds *et al.*, 2007). The purpose of the present study was to determine the usefulness of MLST in the analysis of *C. dubliniensis*. The same set of MLST loci currently used for *C. albicans* was investigated to determine whether it could be applied to the analysis of *C. dubliniensis* due to the high levels of nucleotide sequence identity (~ 90%) shared by the two species, thus allowing comparative sequence data to be used to provide valuable information concerning the evolutionary relatedness of the two species.

The current MLST scheme in use for isolates for *C. albicans* examines seven loci, namely *AAT1a, ACC1, ADP1, PMIb, ALA1, VPS13*, and *ZWF1b* (Scheme A; Table 4.4) and was also applied in the analysis of *C. dubliniensis* isolates. It demonstrated poor levels of discrimination amongst *C. dubliniensis* isolates, identifying only 20 DSTs from 50 isolates. Extension of the *VPS13* and *ZWF1b* loci to include flanking sequences increased the discrimination slightly, raising the number of DSTs by two in the second scheme examined, Scheme B (Table 4.4). The relatively poor levels of discrimination achieved using both Schemes A and B led to the examination of an additional three loci, *AAT1b, GLN4,* and *RPN2,* which had also previously been examined for use in the MLST analysis of *C. albicans*. This scheme, Scheme C (Table 4.4), identified 26 DSTs from the 50 *C. dubliniensis* isolates. The ten loci (Scheme C) were concatenated and used in the construction of a UPGMA dendrogram (Fig. 4.1), which clustered the isolates into three distinct major clades at a cut off sequence identity of 99.7%. A maximum parsimony tree (Fig. 4.2), was also constructed using concatenated sequence data, this time using another scheme, Scheme E (Table 4.4), which was based on the seven loci from Scheme A, and an extra locus, *RPN2* (Table 4.4). Sequence data for these eight loci were available for both *C. albicans* and *C. dubliniensis*, thus enabling the comparative study of the population structure for both species. A UPGMA dendrogram based on allelic profile and DST data according to the consensus *C. albicans* MLST scheme (Scheme A, Table 4.4) was also used for the comparative population analysis of *C. albicans* and *C. dubliniensis* (Fig. 4.3). This study shows that the *C. albicans* clades are more divergent than those observed in *C. dubliniensis* (Figs.

Figure 4.7. Population structure of *C. dubliniensis* defined using the recommended *C. dubliniensis* MLST scheme.

UPGMA dendrogram based on the concatenated polymorphisms for each of the MLST loci included in Scheme D (Table 4.4), which are *AAT1b, ACC1, ADP1, PMIb, RPN2, ALA1,* ex*VPS13* and ex*ZWF1b.* This dendrogram is based on p-distance as indicated by the scale bar. The same three clades previously identified using other MLST schemes and correlating with Cd25 fingerprinting studies, are evident. These MLST clades are indicated by colour, MLST clade C1 is highlighted in green, clade C2 is highlighted in red, and clade C3 is highlighted in blue. Bootstrap values greater than 50 % are indicated at cluster nodes, and DSTs are indicated to the right of each isolate included in the tree.


DST
CD060213 7
P2 7
CD06041 7
PM6-2 7
B1324 7
CD06033 7
CD06038 7
B341 7
SA116 7
CD06031 7
8882 7
IFM0492 7
95-677 7
CCY29-177-1 7
CD06045 7
9097 9
94-234 32
CD36 8
CD604 12
MLNIH0720 4
CD603 3
SA115 2
AV7 2
98-277 2
CM1 2
CD71 2
SA105 11
Eg204 10
CD0512 37
CD06027 1
CD06032 36
1504 5
IFM49832 5
CD98923 5
966/3(1) 13
966/3(2) 13
MLNIH0479 6
Eg203 6
SL370 27
AV6 27
SL411 27
SL422 27
SL509 30
SL495 33
SL410 29
AV5 29
SL375II 31
SL375I 31
SL522 31
SL397 31
SL414 31
CD505 28
CD524 35
p6785 22
p7276 22
CD060215 24
CD514 25
CD519 23
Eg207 21
Eg202 14
SA108 14
SA121 14
SA100 14
p7718 14
49831 20
IFM49883 26
CD539 17
CD75004 15
Can4 16
CD06037 19
CD06036 19
CBS2747 18
18P 34
140P 34
17P 34
25P 34
22BP 34
110P 34
65
60
97
75
94
95
89
91
99
83
50
99
65
68
86
87
0.20
0.15
0.10
0.05
0.00

4.2 and 4.3). The maximum parsimony tree identified three distinct major clades (C1–C3) in the population of *C. dubliniensis*, which were identical to those of the UPGMA dendrogram (Fig. 4.1).

The population structure of *C. dubliniensis* as determined by MLST correlated with the population structure previously determined both using the complex fingerprinting probe Cd25 and on the basis of ITS genotypes (Joly *et al.*, 1999; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005). In previous studies, isolates assigned to ITS genotype 1 belonged exclusively to Cd25 group I (Joly *et al.*, 1999; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005). Similarly, using the MLST method, all of the ITS genotype 1 isolates tested belonged to the MLST clade C1 (Fig. 4.1 and Table 4.1). The Cd25 group II was previously shown to consist of isolates belonging to ITS genotypes 2, 3, and 4 (Gee *et al.*, 2002). The MLST method breaks the Cd25 group II into two, assigning all of the ITS genotype 2 isolates to MLST clade C2 exclusively, and assigning the ITS genotype 3 and 4 isolates to MLST clade C3 (Fig. 4.1 and Table 4.1). Interestingly, a number of ITS genotype 3 and 4 isolates in MLST clade C3 displayed an identical DST and localised to the same area on both the UPGMA and maximum parsimony trees (Figs. 4.1 and 4.2). These isolates were all previously associated with the Cd25 fingerprint group III (Table 4.1), and most displayed high levels of resistance to the antifungal drug 5FC (Al Mosaid *et al.*, 2005). The agreement between these three methods suggests that MLST may be applied as a reliable alternative method for studying the population structure of *C. dubliniensis*. MLST is less time consuming, more reproducible, and more conducive to comparison between laboratories than other methods previously used for this purpose.

The best set of eight loci (*AAT1b, ACC1, ADP1, PMIb, RPN2, ALA1,* ex*VPS13,* and ex*ZWF1b*) proposed for maximum discrimination amongst isolates of *C. dubliniensis* using the minimum number of MLST loci possible was determined according to the highest number of genotypes per variable base in each locus and is referred to as Scheme D (Table 4.4). The *AAT1b* locus is no longer recommended for use in *C. albicans* MLST studies, and therefore the corresponding *AAT1b* locus should be replaced with the *AAT1a* locus for the purpose of comparative population analysis between the two species. The eight loci recommended for the purposes of comparative population analysis between *C. albicans* and *C. dubliniensis* using MLST are therefore from Scheme E; *AAT1a, ACC1, ADP1, PMIb, RPN2, ALA1, VPS13,* and *ZWF1b* (Table 4.4).

The MLST data suggest a relatively low level of divergence in the population structure of *C. dubliniensis* relative to that of *C. albicans*. Interestingly, this is in contrast to the findings of electrophoretic karyotypic analysis, in which isolates of *C. dubliniensis* have significantly different major karyotypic patterns due to the processes of microevolution and chromosomal rearrangement (Gee *et al.*, 2002). Karyotype analysis was able to distinguish the Cd25 group I and II isolates to a large extent, but was unable to distinguish amongst the ITS genotypes. In comparison, the MLST method is able to distinguish readily amongst certain ITS genotypes, most notably the ITS genotype 1 and 2 isolates. ITS genotype 3 and 4 isolates could not be distinguished reliably by MLST, due to the lack of sequence variation among these isolates, the majority of which to date have been recovered from the Middle East. However, this may possibly be improved on by the inclusion of a larger number of ITS genotype 3 and 4 isolates from geographical locations outside of the Middle East.

The low level of sequence variation throughout the population of *C. dubliniensis* suggests that MLST may not be ideal for local epidemiological studies (e.g., an outbreak in a hospital). In this instance, karyotypic analysis may be more appropriate. A possible reason for the low level of discrimination is the relatively small collection of isolates studied. However, the studied isolates were recovered from a wide range of geographical locations (Fig. 3.1) and included isolates recovered from both carriage and systemic infection. Another possible reason for the low level of sequence variation and heterozygosity may be the lack of divergence within the population of *C. dubliniensis*. Results of genotypic diversity and linkage diversity analyses suggested that the sample of 50 *C. dubliniensis* isolates investigated in this study represent a clonal population. However, it is important to emphasise that the sample number was relatively small, even though many of the isolates were recovered from disparate geographic locations around the world (Fig. 3.1).

Candida dubliniensis is a poor pathogen in comparison to *C. albicans*, since it rarely causes infections in normal healthy individuals, and therefore may be under less pressure to adapt to different host environments. Furthermore, it is possible that a more diverse population of *C. dubliniensis* exists in non-human hosts and that this is not reflected in the present study (see below). *Candida* species have also been recovered from non-human hosts such as dogs, cats, birds, and chameleons (Buck, 1990; Pressler *et al.*, 2003; Tavanti *et al.*, 2005a; Cafarchia *et al.*, 2006) and MLST analysis of *C. albicans* isolates recovered from such sources has revealed the enrichment of MLST

clade 8 with non-human isolates, and the absence (with the exception of isolates recovered from primates) of such isolates in MLST clade 1, the most heavily populated *C. albicans* MLST clade (Jacobsen *et al.*, 2008; Wrobel *et al.*, 2008). Recent environmental studies have reported the recovery of ITS genotype I *C. dubliniensis* isolates from *I. uriae* ticks on the Great Saltee Island off South-Eastern Ireland (Nunn *et al.*, 2007), the supposed source being bird excrement. *Candida* species such as *C. albicans*, *C. guilliermondii*, and *C. tropicalis* have been recovered previously from the gastrointestinal tracts and cloacae of birds (Hasenclever & Kogan, 1975; Buck, 1990; Cafarchia *et al.*, 2006).

The avian-excrement-associated *C. dubliniensis* isolates described in this study are members of MLST clade C1, which includes the majority of isolates that have been recovered from humans. However, the majority (13/14) of the avian-excrement-associated isolates form a distinct subgroup within this clade using both neighbour-joining and BURST algorithims (Figs. 4.4 – 4.7), suggesting that, despite the low level of variation evident within *C. dubliniensis*, there may be a distinct avian subpopulation present in birds. This is supported by the observation that 2/3 isolates (AV5 and AV6) recovered from seabird excrement in Dublin fell into the same subpopulation (defined by MLST, *CDR1* and *MTL* loci) as the isolates recovered from the Great Saltee Island which is 150 km distant from Dublin. The TAG polymorphism data and the *MTL* analysis data from the avian-excrement-associated isolates suggest a highly clonal population. Genetic separation and differential clade distribution between human and animal populations of *C. albicans* have recently been reported using similar MLST data (Jacobsen *et al.*, 2008; Wrobel *et al.*, 2008). This may suggest that specific niches of *Candida* species are present in non-human hosts.

The presence of the avian-excrement-associated subgroup within the most predominant clade (C1) which had previously only been identified in isolates recovered from humans, and the close genetic relatedness between isolates, in particular gull isolate AV7, suggests that transmission between the two hosts can occur, although in this instance the most likely direction of transfer is from man to bird.

Chapter 5

Investigation of the molecular mechanism(s) of clade-specific 5-fluorocytosine resistance in *Candida dubliniensis*

Part I:

Examination of the genes involved in 5-fluorocytosine metabolism

5.1. Introduction

5.1.1. Clade-specific 5FC resistance in *C. dubliniensis*

The population structure of the opportunistic yeast pathogen *C. dubliniensis* has been investigated previously using both the species-specific complex DNA fingerprinting probe Cd25 and by MLST (Joly *et al.*, 1999; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005; Chapter 4, this thesis). Early Cd25 fingerprinting analyses demonstrated that *C. dubliniensis* consists of two fingerprinting groups, termed Cd25 group I and Cd25 group II (Joly *et al.*, 1999). Group I isolates comprised the majority of isolates investigated which were recovered from many countries around the world and were very closely related, with average S_{AB} values of 0.8. Group II isolates were less closely related with an average S_{AB} value of 0.47. These results were later confirmed with a larger collection of isolates from diverse geographical sources by Gee *et al.* (2002), who also showed that Cd25 group I isolates comprised a single ITS genotype (genotype 1). Furthermore, Cd25 group II isolates were found to belong to three ITS genotypes (genotypes 2–4). Recently, a study by Al Mosaid *et al.* (2005) identified a third Cd25 fingerprinting group, termed Cd25 group III, which exhibited an average S_{AB} value of 0.35. Isolates belonging to this clade were recovered exclusively in Egypt, Saudi Arabia and Israel, and belonged to ITS genotypes 3 or 4. All the examined isolates belonging to Cd25 group III examined exhibited high level intrinsic resistance to 5FC. This phenotype did not occur in isolates belonging to either Cd25 groups I or II, including Group I or II isolates recovered from Egypt, Saudi Arabia and Israel. (Al Mosaid *et al.*, 2005). In *C. dubliniensis*, resistance patterns are clearly defined, with 5FC-susceptible isolates exhibiting MICs of ≤ 0.125 μg/ml, and 5FC-resistant isolates exhibiting MICs of ≥ 128 μg/ml (Al Mosaid *et al.*, 2005). In this thesis MLST analysis was used to investigate the population structure of *C. dubliniensis* and revealed the presence of three distinct clades, termed C1–C3 (see Chapter 4, Section 4.3.5). All the examined 5FC-resistant isolates belonging to Cd25 fingerprint group III were found to cluster exclusively in MLST clade C3 (see Chapter 4, this thesis).

5.1.2. Clade-specific 5FC resistance in *C. albicans*

The closest relative of *C. dubliniensis, C. albicans,* also exhibits clade-specific resistance to 5FC, with 72.7% of isolates in Ca3 fingerprinting clade I and MLST clade

1 exhibiting reduced susceptibility to this antifungal agent and exhibiting MICs of ≥ 0.5 μg/ml (Odds *et al.*, 2007; Pujol *et al.*, 2004). In other *C. albicans* Ca3 fingerprint clades, the prevalence of isolates with MIC values similar to this was 2% (Pujol *et al.*, 2004; Odds *et al.*, 2007). In *C. albicans,* 5FC resistance patterns vary amongst isolates, ranging from those exhibiting reduced susceptibility (0.5 – 2 μg/ml), to intermediate (2 – 8 μg/ml) or high-level resistance (≥ 8 μg/ml) and a wide range of MICs for this drug have been reported amongst isolates, ranging from 0.06 μg/ml to ≥ 128 μg/ml (Dodgson *et al.*, 2004; Hope *et al.*, 2004). Throughout Ca3 fingerprinting clade I, 5FC-resistant strains were reported to be distributed randomly, and it has been suggested that recombination occurs throughout the clade in order to homogenise the resistance phenotype (Pujol *et al.*, 2004).

5.1.3. Permeases associated with 5FC uptake

Early studies using *S. cerevisiae* demonstrated that cytosine uptake is mediated by the *FCY2*-encoded polytopic permease which is 533 amino acids in length and has a molecular mass of 58.2 kDa (Weber *et al.*, 1990; Ferreira *et al.*, 1997, 1999b). This permease is a proton symporter that also actively transports purines such as adenine, guanine, and hypoxanthine using the Δ pH component of the proton motive force as an energy source (Ferreira *et al.*, 1997, 1999b; Paluszynski *et al.*, 2006). A hydrophilic portion (I371–N377) of the permease was deduced as being important in the three-dimensional structure and ligand-binding of the active carrier (Ferreira *et al.*, 1997, 1999a, 1999b). More recent studies in the same species have demonstrated that there is more than one method of cytosine uptake into yeast cells, as *FCY2* mutants demonstrate a low-level but dose-dependent 5FC susceptibility (Paluszynski *et al.*, 2006). A genome-wide search of *S. cerevisiae* found many *FCY2* homologues, of which the previously uncharacterised genes *FCY21* and *FCY22* displayed the highest percentage identity (85% and 89%, respectively). Disruption of both of these genes in *S. cerevisiae* resulted in increased 5FC MICs, implicating both genes in 5FC-transport (Paluszynski *et al.*, 2006). Further to these findings, strains containing mutated transporter genes *TPN1* or *FUR4*, or the yet uncharacterised *yor071c* gene also demonstrated reduced 5FC susceptibility. The *TPN1* and *FUR4* genes encode vitamin B and uracil permeases, respectively (Seron *et al.*, 1999; Stolz & Vielreicher, 2003) for both of which the substrates display structural similarities with cytosine. In the haploid *Candida lusitaniae,* the purine-cytosine permease encoded by the *FCY2* gene is closely

associated with influx of cytosine in a manner that is distinct to uracil influx (Noël *et al.*, 2003), and 5FC resistance correlates with undetectable intracellular levels of the drug, suggesting a pivotal role of the *FCY2*-encoded permease in 5FC metabolism (Noël *et al.*, 2003). This has been confirmed using a null mutant for the *FCY2* gene (Chapeland-Leclerc *et al.*, 2005; Papon *et al.*, 2007).

Many individual SNPs have been detected in the *FCY21* and *FCY22* genes of *C. albicans* of which eight resulted in radical amino acid replacements. However, none of these has been associated with 5FC resistance (Hope *et al.*, 2004).

5.1.4. Metabolic pathway of 5FC in yeasts

The antifungal action of 5FC relies on the intracellular conversion of 5FC to 5-fluorouracil (5FU) by cytosine deaminases upon entry into fungal cells (Fig. 5.1). The cytosine deaminase is encoded by the *FCA1* gene in *C. albicans* and by the *CdFCA1* gene in *C. dubliniensis* (Erbs *et al.*, 1997; Al Mosaid *et al.*, 2005), which are homologues of the *FCY1* gene in *S. cerevisiae* (Erbs *et al.*, 1997) and in other *Candida* species, such as *C. lusitaniae* (Papon *et al.*, 2007). The absence of cytosine deaminases in mammalian cells prevents 5FC toxicity in humans, as the 5FC prodrug itself is non-toxic. After conversion of 5FC to 5FU, the *FUR1*-encoded UPRT catalyses the phosphorylation of 5FU to 5-fluorouridine monophosphate (5FUMP) (Fig. 5.1). Two specific kinases catalyze the further phosphorylation of 5FUMP, eventually converting it to 5-fluorouridine triphosphate (5FUTP), which in turn gets incorporated into RNA causing miscoding and leading to an inhibition of fungal protein synthesis (Fig. 5.1). As a secondary method of inhibition, the monophosphorylated form of 5FU (5FUMP) inhibits thymidylate synthetase (Fig. 5.1) leading to a depletion of dTTP and misincorporation of dUTP into newly synthesised DNA, causing irreversible DNA damage and cell cycle arrest (Polak & Scholer, 1975; Waldorf & Polak, 1983; Hoskins & Butler, 2008).

5.1.5. Mechanisms of 5FC resistance in *Candida* spp.

A number of studies have previously suggested that decreased UPRT or cytosine deaminase activity is associated with 5FC resistance in *C. albicans* (Normark & Schonebeck, 1972; Polak & Scholer, 1975; Whelan & Kerridge, 1984). Dodgson *et al.* (2004) examined the DNA sequences of the *FCA1* and *FUR1* genes in 5FC-resistant *C. albicans* isolates in an attempt to elucidate the mechanism of resistance. These

researchers identified a SNP at nucleotide position 301 (C → T transition) in the *FUR1* gene which was consistently present in 5FC-resistant isolates, and resulted in an Arg101Cys amino acid substitution in the encoded UPRT protein (Dodgson *et al.*, 2004). Strains that were homozygous and non-polymorphic were susceptible to 5FC, strains that were heterozygously polymorphic displayed reduced susceptibility, and strains that were homozygous for the polymorphism at this locus were reported as resistant to 5FC, exhibiting MIC values ≥ 16 μg/ml (Dodgson *et al.*, 2004).

The polymorphism occurred in a region of the *FUR1* gene that was highly conserved amongst eukaryotic species such as *C. albicans, S. cerevisiae, Toxoplasma gondii, Neurospora crassa* and *Schizosaccharomyces pombe* (Dodgson *et al.*, 2004). The corresponding Arg126 of *T. gondii* has been shown to play a role in the dimerisation of UPRT monomers, which is required for enzymatic activation (Schumacher *et al.*, 2002) and Dodgson *et al.* (2004) hypothesised that the Arg101Cys substitution may interfere with dimerisation and thus activation of the UPRT enzyme, leading to decreased 5-FC metabolism. In the same year, Hope *et al.* (2004) examined the DNA sequences of the *C. albicans FUR1* gene, the *FCA1* gene, and the two cytosine permease genes *FCY21* and *FCY22*. The latter study confirmed the earlier findings of Dodgson *et al.* (2004), and implicated a polymorphism in the *FCA1* gene as a cause of 5FC resistance in one isolate included in the study. This polymorphism resulted in a Gly28Asp substitution in the cytosine deaminase protein. A Ser29Leu substitution in the same protein was also noted in one isolate displaying an intermediate level of 5FC resistance (MIC = 4 μg/ml) (Hope *et al.*, 2004).

In the haploid *C. lusitaniae*, 5FC resistance can occur via reduced uptake and metabolism of the drug (Noël *et al.*, 2003; Chapeland-Leclerc *et al.*, 2005; Papon *et al.*, 2007). Kinetic transport and transformation studies attributed resistance to defects in a purine-cytosine permease protein encoded by the *FCY22* gene, and also in the cytosine deaminase. However, 5FC MICs achieved by these mutations were observed to be lower (MICs 64 – 128 μg/ml) than the 5FC MICs resulting from mutations in the *FUR1* gene (MICs ≥ 512 μg/ml) (Noël *et al.*, 2003; Chapeland-Leclerc *et al.*, 2005; Papon *et al.*, 2007).

Al Mosaid *et al.* (2005) examined the cytosine deaminase and UPRT-encoding gene homologues in *C. dubliniensis* in an attempt to elucidate the primary causes of 5FC resistance in the Middle Eastern Cd25 group III isolates. The latter study observed a SNP at position 258 in the *CdFCA1* gene of all resistant isolates. However, this did

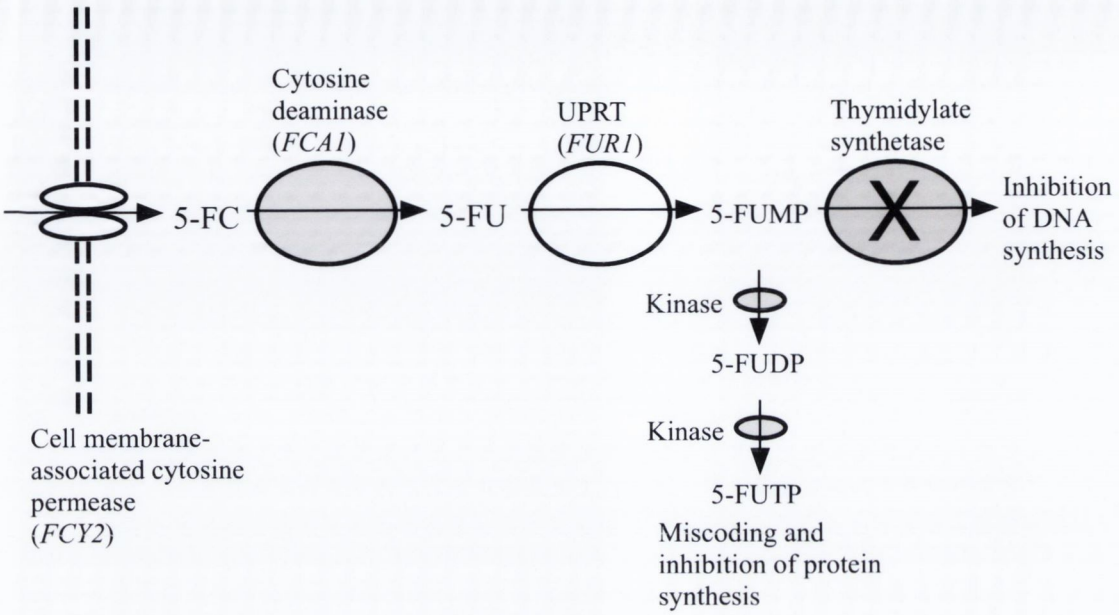


Figure 5.1. Metabolic pathway and mode of action of 5FC in yeasts.

Both 5FC and 5FU are transported into the cell by cell membrane associated cytosine-purine permeases. In *Candida* spp. these are encoded by two genes that display amino acid sequence homology with the *FCY2* gene of *S. cerevisiae* (Hope *et al.*, 2004). Upon entry to the cell, 5FC is then deaminated to 5FU by Fca1p encoded by *FCA1*. 5FU is then phosphorylated by uracil phosphoribosyltransferase (UPRT) encoded by *FUR1* yielding 5-fluorouridine monophosphate (5FUMP). 5FUMP inhibits thymidylate synthetase leading to thymidine depletion in the cell and ultimately interrupting DNA synthesis. 5FUMP is also metabolised by two kinases, yielding 5-fluorouridine diphosphate (5FUDP) and subsequently fluorouridine triphosphate (5FUTP), the latter of which is incorporated into RNA in the place of UTP, leading to miscoding and an inhibition of protein synthesis.

not result in an amino acid substitution, and therefore was unlikely to be a cause of 5FC resistance (Al Mosaid *et al.*, 2005). A number of SNPs were also observed in the *CdFUR1* gene, but no amino acid substitutions resulted. The study concluded that an alternative mechanism(s) of resistance must be responsible for the clade-specific 5FC resistance in *C. dubliniensis* (Al Mosaid *et al.*, 2005).

5.1.6. Objectives

The purpose of the present study was to identify any blocks/lesions in the genes involved in 5FC uptake and metabolism in order to investigate the molecular mechanism(s) of *C. dubliniensis* clade-specific 5FC resistance.

- The nucleotide sequences of the *CdFCY21* and *CdFCY22* genes encoding cytosine permeases and the amino acid sequences of their gene products were compared between 5FC-resistant and 5FC-susceptible isolates in order to identify any changes that could potentially affect drug uptake.
- The nucleotide sequences of the *CdFCY23* and *CdFCY24* genes encoding permeases and the amino acid sequences of their gene products which display similarity to the *TPN1* vitamin B transporter of *S. cerevisiae* were also compared between 5FC-resistant and 5FC-susceptible isolates in order to identify any changes that could affect the uptake of 5FC into cells.
- The nucleotide sequences of the *CdFUR1* gene encoding UPRT and the amino acid sequence of its gene product were also re-analysed in 5FC-resistant and 5FC-susceptible isolates in order to identify any changes that could affect phosphorylation of 5FU in *C. dubliniensis* cells.
- An accumulation assay was carried out to compare 5FC uptake between three 5FC-resistant isolates and three 5FC-susceptible isolates.

5.2. Materials and methods

5.2.1. Isolates used in the study

Twenty-one epidemiologically unrelated human *C. dubliniensis* isolates were included in the present study, including nine 5FC-susceptible isolates and 12 5FC-resistant isolates (Table 5.1) as previously reported (Sullivan *et al.*, 1995; Polacheck *et al.*, 2000; Al Mosaid *et al.*, 2001, 2005; Gee *et al.*, 2002). Previously, 5FC resistance in *C. dubliniensis* has only ever been reported in isolates from Saudi Arabia, Egypt and

Kuwait (Pfaller *et al.*, 1999; Ahmad *et al.*, 2004; Al Mosaid *et al.*, 2005). All but two of the studied isolates had been investigated previously by MLST analysis and/or Cd25 fingerprint analysis and were shown to belong to *C. dubliniensis* MLST clade C3 and Cd25 fingerprint group III (Al Mosaid *et al.*, 2005; Chapter 4, this thesis). For these reasons, 20 of the 21 isolates chosen for study (12 5FC-resistant and eight 5FC-susceptible) were originally recovered in Egypt, Saudi Arabia or Israel (Table 5.1). The 5FC-susceptible isolates belonged to MLST clades C1 or C2 and to Cd25 fingerprint groups I or II, with the exception of isolate p7718 (Table 5.1) which belongs to Cd25 group III/MLST clade C3 (Al Mosaid *et al.*, 2005; Chapter 4, this thesis). The *C. dubliniensis* type strain CD36, originally isolated from the oral cavity of a HIV-infected individual in Ireland was also included as a reference isolate because the complete genome sequence of this organism has been determined (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/).

5.2.2. Determination of *C. dubliniensis* 5FC and 5FU MICs

Broth microdilution assays were used to test susceptibility of the *C. dubliniensis* isolates to 5FC as well as 5FU and were carried out as described in Chapter 2, Section 2.7.

5.2.3. Radiolabelled 5FC accumulation assay

The uptake of 5-fluorocytosine-6-^{3}H ([^{3}H] 5FC) with a specific activity of 37 MBq/ml (Sigma-Aldrich Ltd.) into *C. dubliniensis* cells was monitored according to the method described by Sanglard *et al.* (1995). Cells were grown to mid-log phase in YPD medium at 30°C with constant shaking at 200 rpm (Gallenkamp) and harvested by centrifugation at 500 × *g* for 5 min. Cells were then washed in fresh YNB medium and resuspended in YNB to a cell density of 5×10^{8} cfu /ml in 500 µl YNB. The assay was started by the addition of 12.7 µl (corresponding to 1nmol or 470 MBq) [^{3}H] 5FC to the cells followed by immediate incubation at 30°C. Aliquots (100 µl) were removed at 0, 10, 20, and 30 min time points and cells were washed in 400 µl YNB containing 20 µM unlabelled FC using Spin-X microcentrifuge tubes (pore size, 0.45 µm) with a nylon membrane (Sigma-Aldrich Ltd.). Centrifugation was carried out at 9000 × *g* for 1 min. The aqueous medium was discarded between washes. Filter washing in unlabelled 5FC was repeated three times for each time point as described above. The cells were finally resuspended on the filter in 1.85 M NaOH/7.5% mercaptoethanol and transferred to

Table 5.1. *Candida dubliniensis* isolates used in the study of 5FC metabolism

Isolate	Origin	Yr of isolation	Cd25 group[d]	ITS[e]	5FC MIC (μg/ml)	5FU MIC (μg/ml)	Reference
CD36[abc]	Ireland	1988	I	1	≤ 0.25	32	(Sullivan *et al.*, 1995; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005)
SA101	S.A	2002	I	1	≤ 0.25	16	(Al Mosaid *et al.*, 2005)
SA105[c]	S.A	2002	I	1	≤ 0.25	32	(Al Mosaid *et al.*, 2005)
SA115	S.A	2002	I	1	≤ 0.25	32	(Al Mosaid *et al.*, 2005)
Eg203	Egypt	2002	I	1	≤ 0.25	16	(Al Mosaid *et al.*, 2005)
Eg206	Egypt	2002	I	1	≤ 0.25	32	(Al Mosaid *et al.*, 2005)
Eg200	Egypt	2002	III	4	≥ 128	8	(Al Mosaid *et al.*, 2005)
Eg201[b]	Egypt	2002	III	4	≥ 128	32	(Al Mosaid *et al.*, 2005)
Eg202[b,c]	Egypt	2002	III	4	≥ 128	32	(Al Mosaid *et al.*, 2005)
Eg207	Egypt	2002	III	4	≥ 128	32	(Al Mosaid *et al.*, 2005)
SA100	S.A	2002	III	3	≥ 128	32	(Al Mosaid *et al.*, 2005)
SA103	S.A	2002	III	3	≥ 128	32	(Al Mosaid *et al.*, 2005)
SA107	S.A	2002	III	3	≥ 128	32	(Al Mosaid *et al.*, 2005)
SA108	S.A	2002	III	3	≥ 128	32	(Al Mosaid *et al.*, 2005)
SA109	S.A	2002	III	3	≥ 128	8	(Al Mosaid *et al.*, 2005)
SA113[b,c]	S.A	2002	III	4	≥ 128	32	(Al Mosaid *et al.*, 2005)
SA118	S.A	2002	III	3	≥ 128	32	(Al Mosaid *et al.*, 2005)
SA121[b,c]	S.A	2002	III	4	≥ 128	32	(Al Mosaid *et al.*, 2005)
p7276[b,c]	Israel	1999	II	3	≤ 0.25	8	(Gee *et al.*, 2002; Al Mosaid *et al.*, 2005)
p6785	Israel	1999	II	3	≤ 0.25	16	(Polacheck *et al.*, 2000, Gee *et al.*, 2002; Al Mosaid *et al.*, 2005)
p7718	Israel	1999	III	4	≤ 0.25	16	Gee *et al.*, 2002; Al Mosaid *et al.*, 2005)

Abbreviations: S.A, Saudi Arabia.

[a] *Candida dubliniensis* type strain.

[b] Isolates from which the DNA sequences of the *CdFCY21, CdFCY22, CdFCY23, CdFCY24,* and *CdFUR1* genes were determined and analysed.

[c] Isolates used in radiolabelled [^{3}H] 5FC accumulation assays.

[d] Fingerprint group assigned according to studies using the *C. dubliniensis*-specific DNA fingerprinting probe, Cd25.

[e] Genotypes assigned according to sequence of the ITS region of the rDNA operon.

scintillation vials, (Sigma-Aldrich Ltd.) to which 5 ml of scintillation fluid (Ultima Gold; Packard, Groningen, the Netherlands) was added. The concentration of [^{3}H] 5FC in the cells was determined by counting in a TRI CARB 2100TR liquid scintillation analyser (Canberra Packard, Schwadorf, Austria). Accumulation assays were carried out in duplicate (with the exception of *C. dubliniensis* isolate p7276 which was only assayed once) and included a 5FC-susceptible isolate and a 5FC-resistant isolate on each occasion (Table 5.1).

5.2.4. Amplification of genes associated with metabolism of 5FC

5.2.4.1. Amplification of CdFCY21–CdFCY24 *permease genes*

The complete ORFs of genes encoding putative cytosine permeases were amplified using the proofreading DNA polymerase enzyme Expand High Fidelity Polymerase (Roche) and the corresponding forward and reverse primers listed in Table 5.2. Amplification reactions were carried out in 50 µl volumes containing 3 mM $MgCl_2$, 2.5 U of Expand High Fidelity polymerase, 1 × reaction buffer (all supplied by Roche) 12 pmol of the forward and reverse primers, 200 µM of each dNTP (Promega) and 10 ng of DNA template. Reaction cycling conditions for these genes incorporated an initial denaturation step of 94°C for 2 min, followed by 30 cycles of 30 s at 94°C, 30 s at 52°C, and 2 min at 72°C, followed by a final 7 min extension step at 72°C. Amplification products were visualised on a UV transilluminator following electrophoresis through 1% (w/v) agarose gels containing 0.5 µg ethidium bromide/ml.

5.2.4.2. Amplification of the CdFUR1 *gene*

The complete ORF of the *C. dubliniensis* UPRT-encoding gene *CdFUR1* was amplified using the same conditions as described above, except that 2.5 mM $MgCl_2$ was used instead of 3 mM, as described previously (Al Mosaid *et al.*, 2005). Primers used are listed in Table 5.2. PCR cycling conditions for the *CdFUR1* gene were 7 min at 94°C, followed by 30 cycles of 94°C for 1 min, 53°C for 1 min and 72°C for 1 min. A final elongation step at 72°C was carried out for 10 min after the 30 cycles. Amplification products were visualised on a UV transilluminator following electrophoresis through 2% (w/v) agarose gels containing 0.5 µg ethidium bromide/ml.

5.2.5. Purification of PCR products

All PCR products were purified using the Sigma GenElute PCR Clean-up kit (Sigma-Aldrich Ltd.) or were purified from agarose gels using a Qiaex Gel Extraction Kit (Qiagen) according to the manufacturers' instructions as outlined in Chapter 2, Section 2.5.2.

5.2.6. Cloning and sequencing of genes associated with metabolism of 5FC

Purified PCR products were ligated to the P-GEM T-Easy vector in separate ligation reactions as outlined in Chapter 2, Section 2.5.4. Ligation reactions were transformed into *E. coli* strain DH5α as outlined in Chapter 2, Section 2.5.5. Plasmid DNA was recovered from the transformed cells as outlined in Chapter 2 Section 2.5.3, and sequencing reactions were carried out as described in Chapter 2, Section 2.6.

5.2.7. Sequence analysis and alignments

Chromatograms were analysed using the 373A Data Analysis program version 1.2.0 (Applied Biosystems). Sequence analysis was carried out using the DNA Strider™ version 1.3f11 software packages for DNA and protein analysis (CEA/Saclay, Gif-sur-Yvette, France). Searches of the EMBL and GenBank databases for nucleotide and amino acid sequence similarities were performed using the BLAST series of computer programmes (Altschul *et al.*, 1990). Alignments of nucleotide and amino acid sequences were carried out using the CLUSTAL W sequence alignment computer programme (Thompson *et al.*, 1994). The consensus *C. dubliniensis* sequences were obtained for each gene investigated by BLAST analysis of *C. albicans* gene homologues (http://www.candidagenome.org/) using the *C. dubliniensis* genome sequence database (http://www.sanger.ac.uk.sequencing/Candida/dubliniensis/) based on the type strain CD36.

5.3. Results

5.3.1. Sequence analyses of the *CdFCY21* genes

Comparative analyses of the sequences of the *CdFCY21* permease genes from five 5FC-resistant isolates (SA121, SA108, SA103, Eg201, Eg202) and the 5FC-susceptible isolate p7276, which was recovered from a similar geographical location and belonged to the same MLST clade C3 as the 5FC-resistant isolates (Table 4.7) with

Table 5.2. Oligonucleotide primers used in the amplification of *C. dubliniensis* genes associated with 5FC metabolism

Primer	Primer Sequence 5′ - 3′	Primer Coordinates[a]	Amplicon length (bp)	Gene length (bp)	Sequence identity[b] (%)
FUR1 F	GAAATAGTGGGAGCATCT	-140 → +772	932	658	93
FUR1 R	TAAAGTGTATGTAAATGGTGAGG				
FCY21 F	ATGCCAGAAAAGACATCAGTC	-140 → +1,630	1,786	1,555	89
FCY21 R	TTTTTTCGAAATATCTAGCCAC				
FCY22 F	ATGGCAGAGAATTACGATTTAG	-201 → +1,773	1,991	1,578	90
FCY22 R	CAAATGGTCTAACCAAGTTG				
FCY23 F	TGAAATCAGTTTTTCTTCACC	-25 → +1,805	1,853	1,776	92
FCY23 R	CAGGTTATATGTCTCTATCTATGA				
FCY24 F	GCTTTAAGCTTCTTTTCTTTAATTTGG	-64 → +1,704	1,785	1,593	89
FCY24 R	AAAAAGGAAACATTTTTGTCAGC				

[a] Nucleotide coordinates are indicated from the first base of the forward primer and the first base of the reverse primer, relative to the first base of the ATG start codon being designated +1.

[b] The nucleotide sequence of the *C. dubliniensis* genes were obtained using BLAST analysis of corresponding genes in *C. albicans* against the *C. dubliniensis* genome sequence database (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_dubliniensis).

the reference 5FC-susceptible type strain CD36 were undertaken. Fourteen SNPs were found amongst the seven isolates, seven of which were found amongst the 5FC-resistant isolates exclusively, and one which occurred in the 5FC-susceptible isolate p7276 exclusively (Table 5.3). The remaining six SNPs occurred in both 5FC-resistant and 5FC-susceptible isolates, but were not present in every isolate sequenced (Table 5.3).

Six of the 14 SNPs gave rise to amino acid substitutions, five of which occurred in the 5FC-resistant isolates, and one of which (Glu239Lys), was observed in both the 5FC-susceptible isolate p7276 and 5FC-resistant isolates (Table 5.3 and Fig. 5.2) in comparison to the consensus *CdFCY21* sequence from strain CD36. Of the five amino acid substitutions that occurred amongst the 5FC-resistant isolates, only one (Phe502Leu) occurred in more than one isolate. This was present in the two 5FC-resistant Egyptian isolates Eg201 and Eg202 (Table 5.3 and Fig. 5.2). No amino acid substitutions occurred in the Val356–Asn361 region of the *CdFCY21*-encoded permease which displayed similarity to the Ile371–Asn377 region of the *S. cerevisiae FCY2*-encoded permease. This hydrophilic region was previously shown to be important for the 3D structure and ligand-binding activity of the encoded permease (Ferreira *et al.*, 1997).

5.3.2. Sequence analyses of the *CdFCY22* genes

Comparative analyses of the sequences of the *CdFCY22* genes from five 5FC-resistant isolates (SA121, SA108, SA103, Eg201, Eg202) and the 5FC-susceptible isolate p7276 (Table 4.7) with the reference 5FC-susceptible type strain CD36 were undertaken. The *CdFCY22* sequences contained a total of 16 SNPs amongst these seven isolates (Table 5.4). Two of these 16 SNPs were unique to the 5FC-susceptible isolate p7276, 9/16 SNPs occurred amongst the 5FC-resistant isolates exclusively, and 5/16 SNPs occurred in all isolates examined in comparison to the consensus *CdFCY22* sequence from the 5FC-susceptible type strain CD36.

Seven of these SNPs gave rise to amino acid substitutions in the encoded permease. One of these amino acid substitutions (Cys203Gly) occurred in all 5FC-resistant and the 5FC-susceptible isolate p7276 examined (Table 5.4 and Fig. 5.3) in comparison to the consensus *CdFCY22* sequence from the 5FC-susceptible type strain CD36. The other amino acid substitutions occurred amongst the 5FC-resistant isolates exclusively. However, none of these amino acid substitutions was common to more than one of the 5FC-resistant isolates examined (Table 5.4 and Fig. 5.3). No amino acid

substitutions occurred in the Ile363–Gly369 region of the *CdFCY22*-encoded protein which displays homology to the Ile371–Asn377 region of the *FCY2*-encoded permease of *S. cerevisiae* (Ferreira *et al.*, 1997).

5.3.3. Sequence analyses of the *CdFCY23* genes

A phylogenetic relationship based on amino acid sequence identity of the *TPN1* gene from *S. cerevisiae* and the *FCY23* and *FCY24* genes of *C. albicans* has previously been observed (Hope *et al.*, 2004). Since the *TPN1* gene product has shown the ability to transport cytosine (Paluszynski *et al.*, 2006), these genes were also investigated in the present study in order to determine if they play a role in 5FC resistance in *C. dubliniensis*.

Comparative analyses of the sequences of the *CdFCY23* genes from five 5FC-resistant isolates (SA121, SA108, SA113, Eg201, Eg202) and the 5FC-susceptible isolate p7276 (Table 4.7) with the reference 5FC-susceptible type strain CD36 were undertaken. The *CdFCY23* sequences contained a total of 24 SNPs amongst these seven isolates (Table 5.5). Three of these SNPs occurred in the 5FC-susceptible isolate p7276 only. The remaining 21 SNPs occurred in the 5FC-resistant isolates exclusively. Of these 21 SNPs, three were common to all of the 5FC-resistant isolates examined. One of these three SNPs was non-synonymous, resulting in a Ser280Leu substitution (Table 5.5 and Fig. 5.4). Of the 18 remaining SNPs that occurred amongst the 5FC-resistant isolates, eight resulted in amino acid substitutions. However, none of these eight substitutions were common to more than one isolate (Table 5.5 and Fig. 5.4).

5.3.4. Sequence analyses of the *CdFCY24* genes

Comparative analyses of the sequences of the *CdFCY24* genes from five 5FC-resistant isolates (SA121, SA108, SA113, Eg201, Eg202) and the 5FC-susceptible isolate p7276 (Table 4.7) with the reference 5FC-susceptible type strain CD36 were undertaken. The *CdFCY24* sequences contained a total of 38 SNPs amongst these seven isolates (Table 5.6). Of these 38 SNPs, two occurred in the 5FC-susceptible isolate p7276 and all 5FC-resistant isolates in comparison to the consensus *CdFCY24* gene from the 5FC-susceptible type strain CD36. Four occurred in all five of the 5FC-resistant isolates examined exclusively, 12 occurred in the 5FC-susceptible isolate p7276 exclusively, and the remaining 20 SNPs occurred amongst the five 5FC-resistant

Table 5.3. Polymorphic nucleotides and amino acid substitutions in *CdFCY21*

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
p7276	327	ATA	ATT	109	I	I
SA121	327	ATA	ATT	109	I	I
SA108	327	ATA	ATT	109	I	I
SA103	327	ATA	ATT	109	I	I
Eg201	327	ATA	ATT	109	I	I
Eg202	327	ATA	ATT	109	I	I
SA103	382	TGT	CGT	128	C	R
p7276	450	ACC	ACT	150	T	T
SA121	450	ACC	ACT	150	T	T
SA108	450	ACC	ACT	150	T	T
SA103	514	TCA	CCA	172	S	P
SA108	538	AAT	GAT	180	N	D
p7276	564	GCC	GCT	188	A	A
SA103	564	GCC	GCT	188	A	A
Eg201	564	GCC	GCT	188	A	A
Eg202	564	GCC	GCT	188	A	A
p7276	715	GAA	AAA	239	E	K
SA121	715	GAA	AAA	239	E	K
SA108	715	GAA	AAA	239	E	K
SA103	715	GAA	AAA	239	E	K
Eg201	715	GAA	AAA	239	E	K
Eg202	715	GAA	AAA	239	E	K
Eg202	759	AGT	AGC	253	S	S
p7276	867	ATA	ATT	289	I	I
SA121	867	ATA	ATT	289	I	I
SA108	867	ATA	ATT	289	I	I
SA103	867	ATA	ATT	289	I	I
Eg201	867	ATA	ATT	289	I	I
Eg202	867	ATA	ATT	289	I	I
SA121	891	TTA	TTG	297	L	L
SA108	891	TTA	TTG	297	L	L
p7276	1173	GCA	GCC	391	A	A
SA121	1173	GCA	GCC	391	A	A
SA108	1173	GCA	GCC	391	A	A

Table 5.3. ***Continued***

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
Eg202	1219	ATG	GTG	407	M	V
p7276	1476	ATT	ATC	492	I	I
Eg201	1504	TTT	CTT	502	F	L
Eg202	1504	TTT	CTT	502	F	L

[a] Nucleotide position that displays polymorphism. Positions are indicated from the first base of the ATG start codon being designated 1.

[b] Codon present at this position in the consensus *CdFCY21* sequence from the *C. dubliniensis* type strain CD36, for which the complete genome has been sequenced.

[c] Codon containing polymorphic nucleotide.

[d] Position of amino acid encoded by the codon that contains SNP. Positions are indicated from the first encoded methionine residue being designated 1.

[e] Amino acid at this position in the consensus *CdFCY21* sequence, which was obtained by translation of *CdFCY21* consensus sequence of the CD36 type strain.

[f] Amino acid residue resulting from nucleotide polymorphism.

```
Consensus CdFCY21 MSSDPEKNLSMPEKTSVNLYLYDTNISTTMEPSSSGGEIESTKLNFIDKWAHKLNAETKG 60
p7276 (5FCS)      MSSDPEKNLSMPEKTSVNLYLYDTNISTTMEPSSSGGEIESTKLNFIDKWAHKLNAETKG 60
SA121 (5FCR)      MSSDPEKNLSMPEKTSVNLYLYDTNISTTMEPSSSGGEIESTKLNFIDKWAHKLNAETKG 60
SA108 (5FCR)      MSSDPEKNLSMPEKTSVNLYLYDTNISTTMEPSSSGGEIESTKLNFIDKWAHKLNAETKG 60
SA103 (5FCR)      MSSDPEKNLSMPEKTSVNLYLYDTNISTTMEPSSSGGEIESTKLNFIDKWAHKLNAETKG 60
Eg201 (5FCR)      MSSDPEKNLSMPEKTSVNLYLYDTNISTTMEPSSSGGEIESTKLNFIDKWAHKLNAETKG 60
Eg202 (5FCR)      MSSDPEKNLSMPEKTSVNLYLYDTNISTTMEPSSSGGEIESTKLNFIDKWAHKLNAETKG 60
                  ************************************************************

Consensus CdFCY21 IELVTDEEKTDNSFWNLATMWLSANLVIATFSLGALGITVFNLAFGQAILVIIFFSILGG 120
p7276 (5FCS)      IELVTDEEKTDNSFWNLATMWLSANLVIATFSLGALGITVFNLAFGQAILVIIFFSILGG 120
SA121 (5FCR)      IELVTDEEKTDNSFWNLATMWLSANLVIATFSLGALGITVFNLAFGQAILVIIFFSILGG 120
SA108 (5FCR)      IELVTDEEKTDNSFWNLATMWLSANLVIATFSLGALGITVFNLAFGQAILVIIFFSILGG 120
SA103 (5FCR)      IELVTDEEKTDNSFWNLATMWLSANLVIATFSLGALGITVFNLAFGQAILVIIFFSILGG 120
Eg201 (5FCR)      IELVTDEEKTDNSFWNLATMWLSANLVIATFSLGALGITVFNLAFGQAILVIIFFSILGG 120
Eg202 (5FCR)      IELVTDEEKTDNSFWNLATMWLSANLVIATFSLGALGITVFNLAFGQAILVIIFFSILGG 120
                  ************************************************************

Consensus CdFCY21 FPVAFFSCFGSALGLRQMLLSKFLIGDLTTRLFAAINVIACVGWGAVNTMSSAQLLHIVN 180
p7276 (5FCS)      FPVAFFSCFGSALGLRQMLLSKFLIGDLTTRLFAAINVIACVGWGAVNTMSSAQLLHIVN 180
SA121 (5FCR)      FPVAFFSCFGSALGLRQMLLSKFLIGDLTTRLFAAINVIACVGWGAVNTMSSAQLLHIVN 180
SA108 (5FCR)      FPVAFFSCFGSALGLRQMLLSKFLIGDLTTRLFAAINVIACVGWGAVNTMSSAQLLHIVD 180
SA103 (5FCR)      FPVAFFSRFGSALGLRQMLLSKFLIGDLTTRLFAAINVIACVGWGAVNTMSPAQLLHIVN 180
Eg201 (5FCR)      FPVAFFSCFGSALGLRQMLLSKFLIGDLTTRLFAAINVIACVGWGAVNTMSSAQLLHIVN 180
Eg202 (5FCR)      FPVAFFSCFGSALGLRQMLLSKFLIGDLTTRLFAAINVIACVGWGAVNTMSSAQLLHIVN 180
                  ******* ****************************************** *******

Consensus CdFCY21 NGALPPWAGCLIIVVCTVLVTFFGYHVIHIYEKWAWIPNLIIFIIIIVRFAMTNKFTNES 240
p7276 (5FCS)      NGALPPWAGCLIIVVCTVLVTFFGYHVIHIYEKWAWIPNLIIFIIIIVRFAMTNKFTNKS 240
SA121 (5FCR)      NGALPPWAGCLIIVVCTVLVTFFGYHVIHIYEKWAWIPNLIIFIIIIVRFAMTNKFTNKS 240
SA108 (5FCR)      NGALPPWAGCLIIVVCTVLVTFFGYHVIHIYEKWAWIPNLIIFIIIIVRFAMTNKFTNKS 240
SA103 (5FCR)      NGALPPWAGCLIIVVCTVLVTFFGYHVIHIYEKWAWIPNLIIFIIIIVRFAMTNKFTNKS 240
Eg201 (5FCR)      NGALPPWAGCLIIVVCTVLVTFFGYHVIHIYEKWAWIPNLIIFIIIIVRFAMTNKFTNKS 240
Eg202 (5FCR)      NGALPPWAGCLIIVVCTVLVTFFGYHVIHIYEKWAWIPNLIIFIIIIVRFAMTNKFTNKS 240
                  ************************************************************ *

Consensus CdFCY21 FQQGETTAGNVLSFGGTVFGFATGWTTYLSDYVVYHPRNTNPWKIFFSIFFGLLTPLMFT 300
p7276 (5FCS)      FQQGETTAGNVLSFGGTVFGFATGWTTYLSDYVVYHPRNTNPWKIFFSIFFGLLTPLMFT 300
SA121 (5FCR)      FQQGETTAGNVLSFGGTVFGFATGWTTYLSDYVVYHPRNTNPWKIFFSIFFGLLTPLMFT 300
SA108 (5FCR)      FQQGETTAGNVLSFGGTVFGFATGWTTYLSDYVVYHPRNTNPWKIFFSIFFGLLTPLMFT 300
SA103 (5FCR)      FQQGETTAGNVLSFGGTVFGFATGWTTYLSDYVVYHPRNTNPWKIFFSIFFGLLTPLMFT 300
Eg201 (5FCR)      FQQGETTAGNVLSFGGTVFGFATGWTTYLSDYVVYHPRNTNPWKIFFSIFFGLLTPLMFT 300
Eg202 (5FCR)      FQQGETTAGNVLSFGGTVFGFATGWTTYLSDYVVYHPRNTNPWKIFFSIFFGLLTPLMFT 300
                  ************************************************************

Consensus CdFCY21 LILGAACATGIANDPQWTKLYNEDSVGGLVYAILVQDSLHGFGQFCCVILALSTVANNVP 360
p7276 (5FCS)      LILGAACATGIANDPQWTKLYNEDSVGGLVYAILVQDSLHGFGQFCCVILALSTVANNVP 360
SA121 (5FCR)      LILGAACATGIANDPQWTKLYNEDSVGGLVYAILVQDSLHGFGQFCCVILALSTVANNVP 360
SA108 (5FCR)      LILGAACATGIANDPQWTKLYNEDSVGGLVYAILVQDSLHGFGQFCCVILALSTVANNVP 360
SA103 (5FCR)      LILGAACATGIANDPQWTKLYNEDSVGGLVYAILVQDSLHGFGQFCCVILALSTVANNVP 360
Eg201 (5FCR)      LILGAACATGIANDPQWTKLYNEDSVGGLVYAILVQDSLHGFGQFCCVILALSTVANNVP 360
Eg202 (5FCR)      LILGAACATGIANDPQWTKLYNEDSVGGLVYAILVQDSLHGFGQFCCVILALSTVANNVP 360
                  ************************************************************

Consensus CdFCY21 NMYSMALSAQTVWSGFRKIPRVAWTIIGNGATLAICIPAYYKFEAVMENFMNLISYYLSI 420
p7276 (5FCS)      NMYSMALSAQTVWSGFRKIPRVAWTIIGNGATLAICIPAYYKFEAVMENFMNLISYYLSI 420
SA121 (5FCR)      NMYSMALSAQTVWSGFRKIPRVAWTIIGNGATLAICIPAYYKFEAVMENFMNLISYYLSI 420
SA108 (5FCR)      NMYSMALSAQTVWSGFRKIPRVAWTIIGNGATLAICIPAYYKFEAVMENFMNLISYYLSI 420
SA103 (5FCR)      NMYSMALSAQTVWSGFRKIPRVAWTIIGNGATLAICIPAYYKFEAVMENFMNLISYYLSI 420
Eg201 (5FCR)      NMYSMALSAQTVWSGFRKIPRVAWTIIGNGATLAICIPAYYKFEAVMENFMNLISYYLSI 420
Eg202 (5FCR)      NMYSMALSAQTVWSGFRKIPRVAWTIIGNGATLAICIPAYYKFEAVVENFMNLISYYLSI 420
                  ********************************************* **************

Consensus CdFCY21 YESIMFASHFIWNNGRFDGYDYERWNDKQAYPIGYAGVFGFACGVAGVVLGMNQTWYSGV 480
p7276 (5FCS)      YESIMFASHFIWNNGRFDGYDYERWNDKQAYPIGYAGVFGFACGVAGVVLGMNQTWYSGV 480
SA121 (5FCR)      YESIMFASHFIWNNGRFDGYDYERWNDKQAYPIGYAGVFGFACGVAGVVLGMNQTWYSGV 480
SA108 (5FCR)      YESIMFASHFIWNNGRFDGYDYERWNDKQAYPIGYAGVFGFACGVAGVVLGMNQTWYSGV 480
SA103 (5FCR)      YESIMFASHFIWNNGRFDGYDYERWNDKQAYPIGYAGVFGFACGVAGVVLGMNQTWYSGV 480
Eg201 (5FCR)      YESIMFASHFIWNNGRFDGYDYERWNDKQAYPIGYAGVFGFACGVAGVVLGMNQTWYSGV 480
Eg202 (5FCR)      YESIMFASHFIWNNGRFDGYDYERWNDKQAYPIGYAGVFGFACGVAGVVLGMNQTWYSGV 480
                  ************************************************************
```

```
Consensus CdFCY21 IGRQIGEFGGDIGFELAIGFAFIGFNVARYFEKKIY 516
p7276 (5FCS)      IGRQIGEFGGDIGFELAIGFAFIGFNVARYFEKKIY 516
SA121 (5FCR)      IGRQIGEFGGDIGFELAIGFAFIGFNVARYFEKKIY 516
SA108 (5FCR)      IGRQIGEFGGDIGFELAIGFAFIGFNVARYFEKKIY 516
SA103 (5FCR)      IGRQIGEFGGDIGFELAIGFAFIGFNVARYFEKKIY 516
Eg201 (5FCR)      IGRQIGEFGGDIGFELAIGFALIGFNVARYFEKKIY 516
Eg202 (5FCR)      IGRQIGEFGGDIGFELAIGFALIGFNVARYFEKKIY 516
                  ********************* **************
```

Figure 5.2. Alignment of amino acid sequences for the *CdFCY21* gene in 5FC-susceptible and 5FC-resistant isolates.

Sequences were compared with that of the consensus *CdFCY21* sequence from the *C. dubliniensis* type strain CD36, for which the entire genome sequence has been determined (http://www.sanger.ac.uk.sequencing/Candida/dubliniensis/). Amino acid transitions that occurred in 5FC-resistant isolates only are highlighted in red, bold, underlined typeface. Amino acid transitions that occurred in both the 5FC-susceptible isolate p7276 and the five 5FC-resistant isolates in contrast to the consensus *CdFCY21* sequence from the 5FC-susceptible *C. dubliniensis* type strain CD36 are highlighted in green, bold, underlined typeface. A region showing homology to the Ile371–Asn377 region of the *FCY2* gene product in *S. cerevisiae*, which was previously determined to be important to the three-dimensional structure of the permease (Ferreira *et al.*, 1997) occurs from Val356–Asn361 of the *CdFCY21*-encoded protein. This region is indicated in black, bold, underlined typeface, and contains no amino acid transitions. Six amino acid transitions (Table 5.3) were observed amongst the seven isolates, five of which occurred amongst the 5FC-resistant isolates, and one of which (Glu239Lys) occurred in the 5FC-susceptible isolate p7276 and all 5FC-resistant isolates examined, in comparison to the consensus *CdFCY21* sequence from CD36. Of the five amino acid transitions that occurred amongst the 5FC-resistant isolates, only one (Phe502Leu) occurred in more than one isolate (Table 5.3). This was present in the two 5FC-resistant Egyptian isolates, Eg201 and Eg202. Abbreviations: 5FCS, 5FC-susceptible; 5FCR, 5FC-resistant.

Table 5.4. Polymorphic nucleotides and amino acid substitutions in *CdFCY22*

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
p7276	57	TTT	TTC	19	F	F
SA121	57	TTT	TTC	19	F	F
SA108	57	TTT	TTC	19	F	F
SA103	57	TTT	TTC	19	F	F
Eg201	57	TTT	TTC	19	F	F
Eg202	57	TTT	TTC	19	F	F
p7276	198	GAA	GAG	66	E	E
SA113	546	GCT	GCC	182	A	A
p7276	607	TGT	GGT	203	C	G
SA121	607	TGT	GGT	203	C	G
SA108	607	TGT	GGT	203	C	G
SA103	607	TGT	GGT	203	C	G
Eg201	607	TGT	GGT	203	C	G
Eg202	607	TGT	GGT	203	C	G
Eg202	640	ACT	GCT	214	T	A
Eg201	710	TTC	TCC	237	F	S
p7276	780	AGT	AGC	260	S	S
Eg201	816	TGG	TGC	272	W	C
SA121	837	TAT	TAC	279	Y	Y
p7276	1032	GGT	GGC	344	G	G
SA121	1032	GGT	GGC	344	G	G
SA108	1032	GGT	GGC	344	G	G
SA103	1032	GGT	GGC	344	G	G
Eg201	1032	GGT	GGC	344	G	G
Eg202	1032	GGT	GGC	344	G	G
SA108	1124	CTT	CCT	375	L	P
Eg201	1173	GCT	GCC	391	A	A
SA113	1297	TCC	CCC	433	S	P
p7276	1413	GGG	GGA	471	G	G
SA121	1413	GGG	GGA	471	G	G
SA108	1413	GGG	GGA	471	G	G
SA103	1413	GGG	GGA	471	G	G
Eg201	1413	GGG	GGA	471	G	G
Eg202	1413	GGG	GGA	471	G	G

Table 5.4. ***Continued***

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
p7276	1446	ACT	ACG	482	T	T
SA121	1446	ACT	ACG	482	T	T
SA108	1446	ACT	ACG	482	T	T
SA103	1446	ACT	ACG	482	T	T
Eg201	1446	ACT	ACG	482	T	T
Eg202	1446	ACT	ACG	482	T	T
SA121	1463	ATC	ACC	488	I	T

[a] Nucleotide position that displays polymorphism. Positions are indicated from the first base of the ATG start codon being designated 1.

[b] Codon present at this position in the consensus *CdFCY22* sequence from the *C. dubliniensis* type strain CD36, for which the complete genome has been sequenced.

[c] Codon containing polymorphic nucleotide.

[d] Position of amino acid encoded by the codon that contains SNP. Positions are indicated from the first encoded methionine residue being designated 1.

[e] Amino acid at this position in the consensus *CdFCY22* sequence, which was obtained by translation of *CdFCY22* consensus sequence of the CD36 type strain.

[f] Amino acid residue resulting from nucleotide polymorphism.

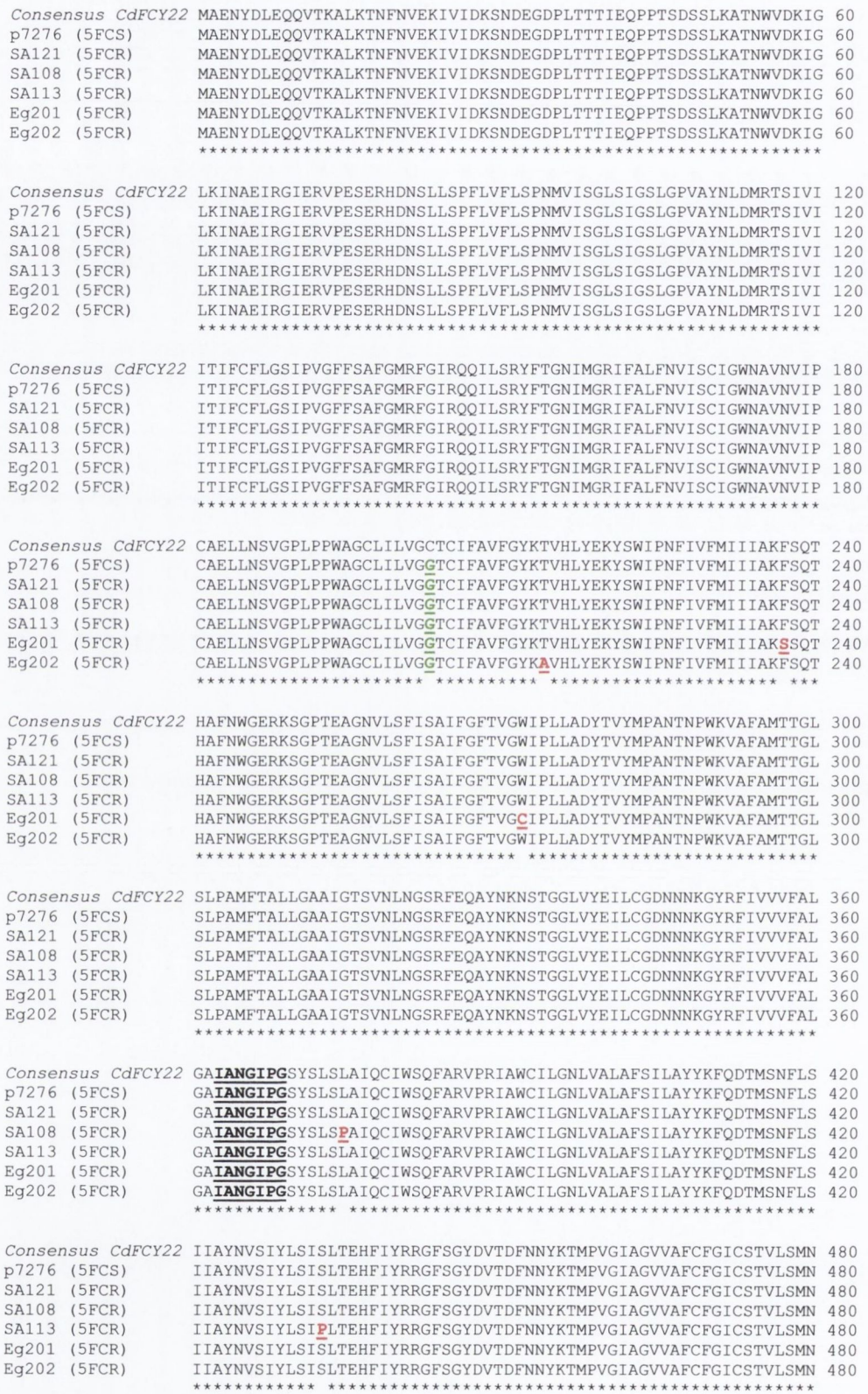

Consensus CdFCY22 MAENYDLEQQVTKALKTNFNVEKIVIDKSNDEGDPLTTTIEQPPTSDSSLKATNWVDKIG 60
p7276 (5FCS) MAENYDLEQQVTKALKTNFNVEKIVIDKSNDEGDPLTTTIEQPPTSDSSLKATNWVDKIG 60
SA121 (5FCR) MAENYDLEQQVTKALKTNFNVEKIVIDKSNDEGDPLTTTIEQPPTSDSSLKATNWVDKIG 60
SA108 (5FCR) MAENYDLEQQVTKALKTNFNVEKIVIDKSNDEGDPLTTTIEQPPTSDSSLKATNWVDKIG 60
SA113 (5FCR) MAENYDLEQQVTKALKTNFNVEKIVIDKSNDEGDPLTTTIEQPPTSDSSLKATNWVDKIG 60
Eg201 (5FCR) MAENYDLEQQVTKALKTNFNVEKIVIDKSNDEGDPLTTTIEQPPTSDSSLKATNWVDKIG 60
Eg202 (5FCR) MAENYDLEQQVTKALKTNFNVEKIVIDKSNDEGDPLTTTIEQPPTSDSSLKATNWVDKIG 60
Consensus CdFCY22 LKINAEIRGIERVPESERHDNSLLSPFLVFLSPNMVISGLSIGSLGPVAYNLDMRTSIVI 120
p7276 (5FCS) LKINAEIRGIERVPESERHDNSLLSPFLVFLSPNMVISGLSIGSLGPVAYNLDMRTSIVI 120
SA121 (5FCR) LKINAEIRGIERVPESERHDNSLLSPFLVFLSPNMVISGLSIGSLGPVAYNLDMRTSIVI 120
SA108 (5FCR) LKINAEIRGIERVPESERHDNSLLSPFLVFLSPNMVISGLSIGSLGPVAYNLDMRTSIVI 120
SA113 (5FCR) LKINAEIRGIERVPESERHDNSLLSPFLVFLSPNMVISGLSIGSLGPVAYNLDMRTSIVI 120
Eg201 (5FCR) LKINAEIRGIERVPESERHDNSLLSPFLVFLSPNMVISGLSIGSLGPVAYNLDMRTSIVI 120
Eg202 (5FCR) LKINAEIRGIERVPESERHDNSLLSPFLVFLSPNMVISGLSIGSLGPVAYNLDMRTSIVI 120
Consensus CdFCY22 ITIFCFLGSIPVGFFSAFGMRFGIRQQILSRYFTGNIMGRIFALFNVISCIGWNAVNVIP 180
p7276 (5FCS) ITIFCFLGSIPVGFFSAFGMRFGIRQQILSRYFTGNIMGRIFALFNVISCIGWNAVNVIP 180
SA121 (5FCR) ITIFCFLGSIPVGFFSAFGMRFGIRQQILSRYFTGNIMGRIFALFNVISCIGWNAVNVIP 180
SA108 (5FCR) ITIFCFLGSIPVGFFSAFGMRFGIRQQILSRYFTGNIMGRIFALFNVISCIGWNAVNVIP 180
SA113 (5FCR) ITIFCFLGSIPVGFFSAFGMRFGIRQQILSRYFTGNIMGRIFALFNVISCIGWNAVNVIP 180
Eg201 (5FCR) ITIFCFLGSIPVGFFSAFGMRFGIRQQILSRYFTGNIMGRIFALFNVISCIGWNAVNVIP 180
Eg202 (5FCR) ITIFCFLGSIPVGFFSAFGMRFGIRQQILSRYFTGNIMGRIFALFNVISCIGWNAVNVIP 180
Consensus CdFCY22 CAELLNSVGPLPPWAGCLILVGCTCIFAVFGYKTVHLYEKYSWIPNFIVFMIIIAKFSQT 240
p7276 (5FCS) CAELLNSVGPLPPWAGCLILVGGTCIFAVFGYKTVHLYEKYSWIPNFIVFMIIIAKFSQT 240
SA121 (5FCR) CAELLNSVGPLPPWAGCLILVGGTCIFAVFGYKTVHLYEKYSWIPNFIVFMIIIAKFSQT 240
SA108 (5FCR) CAELLNSVGPLPPWAGCLILVGGTCIFAVFGYKTVHLYEKYSWIPNFIVFMIIIAKFSQT 240
SA113 (5FCR) CAELLNSVGPLPPWAGCLILVGGTCIFAVFGYKTVHLYEKYSWIPNFIVFMIIIAKFSQT 240
Eg201 (5FCR) CAELLNSVGPLPPWAGCLILVGGTCIFAVFGYKTVHLYEKYSWIPNFIVFMIIIAKSSQT 240
Eg202 (5FCR) CAELLNSVGPLPPWAGCLILVGGTCIFAVFGYKAVHLYEKYSWIPNFIVFMIIIAKFSQT 240
Consensus CdFCY22 HAFNWGERKSGPTEAGNVLSFISAIFGFTVGWIPLLADYTVYMPANTNPWKVAFAMTTGL 300
p7276 (5FCS) HAFNWGERKSGPTEAGNVLSFISAIFGFTVGWIPLLADYTVYMPANTNPWKVAFAMTTGL 300
SA121 (5FCR) HAFNWGERKSGPTEAGNVLSFISAIFGFTVGWIPLLADYTVYMPANTNPWKVAFAMTTGL 300
SA108 (5FCR) HAFNWGERKSGPTEAGNVLSFISAIFGFTVGWIPLLADYTVYMPANTNPWKVAFAMTTGL 300
SA113 (5FCR) HAFNWGERKSGPTEAGNVLSFISAIFGFTVGWIPLLADYTVYMPANTNPWKVAFAMTTGL 300
Eg201 (5FCR) HAFNWGERKSGPTEAGNVLSFISAIFGFTVGCIPLLADYTVYMPANTNPWKVAFAMTTGL 300
Eg202 (5FCR) HAFNWGERKSGPTEAGNVLSFISAIFGFTVGWIPLLADYTVYMPANTNPWKVAFAMTTGL 300
Consensus CdFCY22 SLPAMFTALLGAAIGTSVNLNGSRFEQAYNKNSTGGLVYEILCGDNNNKGYRFIVVVFAL 360
p7276 (5FCS) SLPAMFTALLGAAIGTSVNLNGSRFEQAYNKNSTGGLVYEILCGDNNNKGYRFIVVVFAL 360
SA121 (5FCR) SLPAMFTALLGAAIGTSVNLNGSRFEQAYNKNSTGGLVYEILCGDNNNKGYRFIVVVFAL 360
SA108 (5FCR) SLPAMFTALLGAAIGTSVNLNGSRFEQAYNKNSTGGLVYEILCGDNNNKGYRFIVVVFAL 360
SA113 (5FCR) SLPAMFTALLGAAIGTSVNLNGSRFEQAYNKNSTGGLVYEILCGDNNNKGYRFIVVVFAL 360
Eg201 (5FCR) SLPAMFTALLGAAIGTSVNLNGSRFEQAYNKNSTGGLVYEILCGDNNNKGYRFIVVVFAL 360
Eg202 (5FCR) SLPAMFTALLGAAIGTSVNLNGSRFEQAYNKNSTGGLVYEILCGDNNNKGYRFIVVVFAL 360
Consensus CdFCY22 GAIANGIPGSYSLSLAIQCIWSQFARVPRIAWCILGNLVALAFSILAYYKFQDTMSNFLS 420
p7276 (5FCS) GAIANGIPGSYSLSLAIQCIWSQFARVPRIAWCILGNLVALAFSILAYYKFQDTMSNFLS 420
SA121 (5FCR) GAIANGIPGSYSLSLAIQCIWSQFARVPRIAWCILGNLVALAFSILAYYKFQDTMSNFLS 420
SA108 (5FCR) GAIANGIPGSYSLSPAIQCIWSQFARVPRIAWCILGNLVALAFSILAYYKFQDTMSNFLS 420
SA113 (5FCR) GAIANGIPGSYSLSLAIQCIWSQFARVPRIAWCILGNLVALAFSILAYYKFQDTMSNFLS 420
Eg201 (5FCR) GAIANGIPGSYSLSLAIQCIWSQFARVPRIAWCILGNLVALAFSILAYYKFQDTMSNFLS 420
Eg202 (5FCR) GAIANGIPGSYSLSLAIQCIWSQFARVPRIAWCILGNLVALAFSILAYYKFQDTMSNFLS 420
Consensus CdFCY22 IIAYNVSIYLSISLTEHFIYRRGFSGYDVTDFNNYKTMPVGIAGVVAFCFGICSTVLSMN 480
p7276 (5FCS) IIAYNVSIYLSISLTEHFIYRRGFSGYDVTDFNNYKTMPVGIAGVVAFCFGICSTVLSMN 480
SA121 (5FCR) IIAYNVSIYLSISLTEHFIYRRGFSGYDVTDFNNYKTMPVGIAGVVAFCFGICSTVLSMN 480
SA108 (5FCR) IIAYNVSIYLSISLTEHFIYRRGFSGYDVTDFNNYKTMPVGIAGVVAFCFGICSTVLSMN 480
SA113 (5FCR) IIAYNVSIYLSIPLTEHFIYRRGFSGYDVTDFNNYKTMPVGIAGVVAFCFGICSTVLSMN 480
Eg201 (5FCR) IIAYNVSIYLSISLTEHFIYRRGFSGYDVTDFNNYKTMPVGIAGVVAFCFGICSTVLSMN 480
Eg202 (5FCR) IIAYNVSIYLSISLTEHFIYRRGFSGYDVTDFNNYKTMPVGIAGVVAFCFGICSTVLSMN 480

Figure 5.3. Alignment of amino acid sequences for the *CdFCY22* gene in 5FC-susceptible and 5FC-resistant isolates.

Sequences were compared with that of the consensus *CdFCY22* sequence from the *C. dubliniensis* type strain CD36, for which the entire genome sequence has been determined (http://www.sanger.ac.uk.sequencing/Candida/dubliniensis/). Amino acid transitions that occurred in 5FC-resistant isolates only are highlighted in red, bold, underlined typeface. Amino acid transitions that occurred in both the 5FC-susceptible isolate p7276 and the five 5FC-resistant isolates in contrast to the consensus *CdFCY22* sequence from the 5FC-susceptible *C. dubliniensis* type strain CD36 are highlighted in green, bold, underlined typeface. A region showing homology to the Ile371–Asn377 region of the *FCY2* gene product in *S. cerevisiae*, which previously observed to be important to the three-dimensional structure of the permease (Ferreira *et al.*, 1997) occurs from Ile363–Gly369 of the *CdFCY22*-encoded protein. This region is indicated in black, bold, underlined typeface, and contains no amino acid transitions. Seven amino acid transitions were observed amongst the seven isolates examined (Table 5.4). One of these amino acid transitions (Cys203Gly) occurred in all 5FC-resistant and 5FC-susceptible isolates examined (Table 5.4), the others occurred amongst the 5FC-resistant isolates exclusively; however none of these amino acid transitions were common to more than one of the 5FC-resistant isolates examined. Abbreviations: 5FCS, 5FC-susceptible; 5FCR, 5FC-resistant.

Table 5.5. Polymorphic nucleotides and amino acid substitutions in *CdFCY23*

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
SA108	158	GAA	GGA	53	E	G
Eg201	219	AAA	AAG	73	K	K
Eg202	220	AAA	GAA	74	K	E
SA113	361	TTT	CTT	121	F	L
SA121	459	TGT	TGC	153	C	C
SA108	459	TGT	TGC	153	C	C
SA113	459	TGT	TGC	153	C	C
Eg201	459	TGT	TGC	153	C	C
Eg202	459	TGT	TGC	153	C	C
SA108	537	TTA	TTG	179	L	L
SA121	751	TTA	CTA	251	L	L
SA108	751	TTA	CTA	251	L	L
SA113	751	TTA	CTA	251	L	L
Eg201	751	TTA	CTA	251	L	L
Eg202	751	TTA	CTA	251	L	L
Eg202	795	ACT	ACC	265	T	T
SA121	839	TCA	TTA	280	S	L
SA108	839	TCA	TTA	280	S	L
SA113	839	TCA	TTA	280	S	L
Eg201	839	TCA	TTA	280	S	L
Eg202	839	TCA	TTA	280	S	L
SA113	855	AAA	AAG	285	K	K
SA121	858	GGT	GGC	286	G	G
p7276	974	ATT	ACT	325	I	T
SA108	1092	GGT	GGC	364	G	G
Eg201	1107	GAA	GAG	369	E	E
p7276	1225	ACG	CCG	409	T	P
SA113	1242	ATT	ATC	414	I	I
SA113	1254	TTA	TTG	418	P	P
SA121	1273	GCA	ACA	425	A	T
SA113	1318	ACT	GCT	440	T	A

Table 5.5. ***Continued***

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
p7276	1453	AAA	GAA	485	K	E
Eg201	1466	TTT	TTC	489	F	S
SA121	1478	ATT	ACT	493	I	T
SA108	1497	CCA	CCG	496	P	P
Eg202	1699	ATG	GTG	567	M	V

[a] Nucleotide position that displays polymorphism. Positions are indicated from the first base of the ATG start codon being designated 1.

[b] Codon present at this position in the consensus *CdFCY23* sequence from the *C. dubliniensis* type strain CD36, for which the complete genome has been sequenced.

[c] Codon containing polymorphic nucleotide.

[d] Position of amino acid encoded by the codon that contains SNP. Positions are indicated from the first encoded methionine residue being designated 1.

[e] Amino acid at this position in the consensus *CdFCY23* sequence, which was obtained by translation of *CdFCY23* consensus sequence of the CD36 type strain.

[f] Amino acid residue resulting from nucleotide polymorphism.

```
Consensus CdFCY23 MSKIKNSITTINQQQTGATTNNNSSSSIEVVTTSAPRYDNNFNDETTHEITNETNSNVSL 60
p7276 (5FCS)      MSKIKNSITTINQQQTGATTNNNSSSSIEVVTTSAPRYDNNFNDETTHEITNETNSNVSL 60
SA121 (5FCR)      MSKIKNSITTINQQQTGATTNNNSSSSIEVVTTSAPRYDNNFNDETTHEITNETNSNVSL 60
SA108 (5FCR)      MSKIKNSITTINQQQTGATTNNNSSSSIEVVTTSAPRYDNNFNDETTHEITNGTNSNVSL 60
SA113 (5FCR)      MSKIKNSITTINQQQTGATTNNNSSSSIEVVTTSAPRYDNNFNDETTHEITNETNSNVSL 60
Eg201 (5FCR)      MSKIKNSITTINQQQTGATTNNNSSSSIEVVTTSAPRYDNNFNDETTHEITNETNSNVSL 60
Eg202 (5FCR)      MSKIKNSITTINQQQTGATTNNNSSSSIEVVTTSAPRYDNNFNDETTHEITNETNSNVSL 60
                  ************************************************** *******

Consensus CdFCY23 SSTLFQYLATISKKLDSLGVETRGIERIQPYERSTNRTKQFFSVMGLWLSACGGLSSMSS 120
p7276 (5FCS)      SSTLFQYLATISKKLDSLGVETRGIERIQPYERSTNRTKQFFSVMGLWLSACGGLSSMSS 120
SA121 (5FCR)      SSTLFQYLATISKKLDSLGVETRGIERIQPYERSTNRTKQFFSVMGLWLSACGGLSSMSS 120
SA108 (5FCR)      SSTLFQYLATISKKLDSLGVETRGIERIQPYERSTNRTKQFFSVMGLWLSACGGLSSMSS 120
SA113 (5FCR)      SSTLFQYLATISKKLDSLGVETRGIERIQPYERSTNRTKQFFSVMGLWLSACGGLSSMSS 120
Eg201 (5FCR)      SSTLFQYLATISKKLDSLGVETRGIERIQPYERSTNRTKQFFSVMGLWLSACGGLSSMSS 120
Eg202 (5FCR)      SSTLFQYLATISKELDSLGVETRGIERIQPYERSTNRTKQFFSVMGLWLSACGGLSSMSS 120
                  ************ ************************************************

Consensus CdFCY23 FYLGPLLFELGLRNTLLAGLLGEILGCFIAAYCSLMGPRSGCRQMVSGRFLFGWWFVKLV 180
p7276 (5FCS)      FYLGPLLFELGLRNTLLAGLLGEILGCFIAAYCSLMGPRSGCRQMVSGRFLFGWWFVKLV 180
SA121 (5FCR)      FYLGPLLFELGLRNTLLAGLLGEILGCFIAAYCSLMGPRSGCRQMVSGRFLFGWWFVKLV 180
SA108 (5FCR)      FYLGPLLFELGLRNTLLAGLLGEILGCFIAAYCSLMGPRSGCRQMVSGRFLFGWWFVKLV 180
SA113 (5FCR)      LYLGPLLFELGLRNTLLAGLLGEILGCFIAAYCSLMGPRSGCRQMVSGRFLFGWWFVKLV 180
Eg201 (5FCR)      FYLGPLLFELGLRNTLLAGLLGEILGCFIAAYCSLMGPRSGCRQMVSGRFLFGWWFVKLV 180
Eg202 (5FCR)      FYLGPLLFELGLRNTLLAGLLGEILGCFIAAYCSLMGPRSGCRQMVSGRFLFGWWFVKLV 180
                  ************************************************************

Consensus CdFCY23 ALVAIIGVMGWSVVNSVVGGQILSSVSNDKIPLWAGIIIIAAISLIVAIAGIKQLIRVEA 240
p7276 (5FCS)      ALVAIIGVMGWSVVNSVVGGQILSSVSNDKIPLWAGIIIIAAISLIVAIAGIKQLIRVEA 240
SA121 (5FCR)      ALVAIIGVMGWSVVNSVVGGQILSSVSNDKIPLWAGIIIIAAISLIVAIAGIKQLIRVEA 240
SA108 (5FCR)      ALVAIIGVMGWSVVNSVVGGQILSSVSNDKIPLWAGIIIIAAISLIVAIAGIKQLIRVEA 240
SA113 (5FCR)      ALVAIIGVMGWSVVNSVVGGQILSSVSNDKIPLWAGIIIIAAISLIVAIAGIKQLIRVEA 240
Eg201 (5FCR)      ALVAIIGVMGWSVVNSVVGGQILSSVSNDKIPLWAGIIIIAAISLIVAIAGIKQLIRVEA 240
Eg202 (5FCR)      ALVAIIGVMGWSVVNSVVGGQILSSVSNDKIPLWAGIIIIAAISLIVAIAGIKQLIRVEA 240
                  ************************************************************

Consensus CdFCY23 FLSIPVNCAFLLLYIVASQKFDYLTWKDAVVANESSEYSSAATVKGNWLSFFSLCYSITS 300
p7276 (5FCS)      FLSIPVNCAFLLLYIVASQKFDYLTWKDAVVANESSEYSSAATVKGNWLSFFSLCYSITS 300
SA121 (5FCR)      FLSIPVNCAFLLLYIVASQKFDYLTWKDAVVANESSEYSLAATVKGNWLSFFSLCYSITS 300
SA108 (5FCR)      FLSIPVNCAFLLLYIVASQKFDYLTWKDAVVANESSEYSLAATVKGNWLSFFSLCYSITS 300
SA113 (5FCR)      FLSIPVNCAFLLLYIVASQKFDYLTWKDAVVANESSEYSLAATVKGNWLSFFSLCYSITS 300
Eg201 (5FCR)      FLSIPVNCAFLLLYIVASQKFDYLTWKDAVVANESSEYSLAATVKGNWLSFFSLCYSITS 300
Eg202 (5FCR)      FLSIPVNCAFLLLYIVASQKFDYLTWKDAVVANESSEYSLAATVKGNWLSFFSLCYSITS 300
                  ************************************* ********************

Consensus CdFCY23 TWGTIASDYYILFPENTPDWEIFAITFFGISIPTTFVGVAGILIGNVALTYQPWGDAYKK 360
p7276 (5FCS)      TWGTIASDYYILFPENTPDWEIFATTFFGISIPTTFVGVAGILIGNVALTYQPWGDAYKK 360
SA121 (5FCR)      TWGTIASDYYILFPENTPDWEIFAITFFGISIPTTFVGVAGILIGNVALTYQPWGDAYKK 360
SA108 (5FCR)      TWGTIASDYYILFPENTPDWEIFAITFFGISIPTTFVGVAGILIGNVALTYQPWGDAYKK 360
SA113 (5FCR)      TWGTIASDYYILFPENTPDWEIFAITFFGISIPTTFVGVAGILIGNVALTYQPWGDAYKK 360
Eg201 (5FCR)      TWGTIASDYYILFPENTPDWEIFAITFFGISIPTTFVGVAGILIGNVALTYQPWGDAYKK 360
Eg202 (5FCR)      TWGTIASDYYILFPENTPDWEIFAITFFGISIPTTFVGVAGILIGNVALTYQPWGDAYKK 360
                  ************************ ***********************************

Consensus CdFCY23 LGMGGLLNEAFKPWGAGGKFLLMIIFLSLISNNIINTYSAAFGVQLAGTLFAKIPRWLWA 420
p7276 (5FCS)      LGMGGLLNEAFKPWGAGGKFLLMIIFLSLISNNIINTYSAAFGVQLAGPLFAKIPRWLWA 420
SA121 (5FCR)      LGMGGLLNEAFKPWGAGGKFLLMIIFLSLISNNIINTYSAAFGVQLAGTLFAKIPRWLWA 420
SA108 (5FCR)      LGMGGLLNEAFKPWGAGGKFLLMIIFLSLISNNIINTYSAAFGVQLAGTLFAKIPRWLWA 420
SA113 (5FCR)      LGMGGLLNEAFKPWGAGGKFLLMIIFLSLISNNIINTYSAAFGVQLAGTLFAKIPRWLWA 420
Eg201 (5FCR)      LGMGGLLNEAFKPWGAGGKFLLMIIFLSLISNNIINTYSAAFGVQLAGTLFAKIPRWLWA 420
Eg202 (5FCR)      LGMGGLLNEAFKPWGAGGKFLLMIIFLSLISNNIINTYSAAFGVQLAGTLFAKIPRWLWA 420
                  ********************************************** ***********

Consensus CdFCY23 ILLTAIYLICALVGRNEFSTILGNFLPMIGYWVSMYFIMLLEENTIFRTDKFKYLFTKEF 480
p7276 (5FCS)      ILLTAIYLICALVGRNEFSTILGNFLPMIGYWVSMYFIMLLEENTIFRTDKFKYLFTKEF 480
SA121 (5FCR)      ILLTTIYLICALVGRNEFSTILGNFLPMIGYWVSMYFIMLLEENTIFRTDKFKYLFTKEF 480
SA108 (5FCR)      ILLTAIYLICALVGRNEFSTILGNFLPMIGYWVSMYFIMLLEENTIFRTDKFKYLFTKEF 480
SA113 (5FCR)      ILLTAIYLICALVGRNEFSAILGNFLPMIGYWVSMYFIMLLEENTIFRTDKFKYLFTKEF 480
Eg201 (5FCR)      ILLTAIYLICALVGRNEFSTILGNFLPMIGYWVSMYFIMLLEENTIFRTDKFKYLFTKEF 480
Eg202 (5FCR)      ILLTAIYLICALVGRNEFSTILGNFLPMIGYWVSMYFIMLLEENTIFRTDKFKYLFTKEF 480
                  **** ************** ****************************************
```

```
Consensus CdFCY23 PYGNKEEDFDSTIAHPIDPRIIGTPLRQNQHYNFNIWNDYDKLTRGLAATTSFIIGATGA 540
p7276 (5FCS)      PYGNEEEDFDSTIAHPIDPRIIGTPLRQNQHYNFNIWNDYDKLTRGLAATTSFIIGATGA 540
SA121 (5FCR)      PYGNKEEDFDSTTAHPIDPRIIGTPLRQNQHYNFNIWNDYDKLTRGLAATTSFIIGATGA 540
SA108 (5FCR)      PYGNKEEDFDSTIAHPIDPRIIGTPLRQNQHYNFNIWNDYDKLTRGLAATTSFIIGATGA 540
SA113 (5FCR)      PYGNKEEDFDSTIAHPIDPRIIGTPLRQNQHYNFNIWNDYDKLTRGLAATTSFIIGATGA 540
Eg201 (5FCR)      PYGNKEEDSDSTIAHPIDPRIIGTPLRQNQHYNFNIWNDYDKLTRGLAATTSFIIGATGA 540
Eg202 (5FCR)      PYGNKEEDFDSTIAHPIDPRIIGTPLRQNQHYNFNIWNDYDKLTRGLAATTSFIIGATGA 540
                  **** *** *** ********************************************

Consensus CdFCY23 AVGMSQAYWIGPLARRIGGEFGGDIAMWLCMGFSGLVYPPLRYLELKKFGR 591
p7276 (5FCS)      AVGMSQAYWIGPLARRIGGEFGGDIAMWLCMGFSGLVYPPLRYLELKKFGR 591
SA121 (5FCR)      AVGMSQAYWIGPLARRIGGEFGGDIAMWLCMGFSGLVYPPLRYLELKKFGR 591
SA108 (5FCR)      AVGMSQAYWIGPLARRIGGEFGGDIAMWLCMGFSGLVYPPLRYLELKKFGR 591
SA113 (5FCR)      AVGMSQAYWIGPLARRIGGEFGGDIAMWLCMGFSGLVYPPLRYLELKKFGR 591
Eg201 (5FCR)      AVGMSQAYWIGPLARRIGGEFGGDIAMWLCMGFSGLVYPPLRYLELKKFGR 591
Eg202 (5FCR)      AVGMSQAYWIGPLARRIGGEFGGDIAVWLCMGFSGLVYPPLRYLELKKFGR 591
                  ************************* ************************
```

Figure 5.4. Alignment of amino acid sequences for the *CdFCY23* gene in 5FC-susceptible and 5FC-resistant isolates.

Sequences were compared with that of the consensus *CdFCY23* sequence from the *C. dubliniensis* type strain CD36, for which the entire genome sequence has been determined (http://www.sanger.ac.uk.sequencing/Candida/dubliniensis/). Amino acid transitions that occurred in 5FC-resistant isolates only are highlighted in red, bold, underlined typeface, and amino acid transitions that occurred in 5FC-susceptible isolates exclusively are highlighted in blue, bold, underlined typeface. Amino acid transitions that occurred in both the 5FC-susceptible isolate p7276 and the five 5FC-resistant isolates in contrast to the consensus *CdFCY23* sequence from the 5FC-susceptible *C. dubliniensis* type strain CD36 are highlighted in green, bold, underlined typeface. Twelve amino acid transitions were observed amongst the seven isolates examined, nine of which occurred in 5FC-resistant isolates exclusively (Table 5.5). One of these amino acid transitions occurred in all of the 5FC-resistant isolates sequenced and was not present in the 5FC-susceptible isolate p7276 or the 5FC-susceptible *C. dubliniensis* type strain CD36 (Table 5.5). Abbreviations: 5FCS, 5FC-susceptible; 5FCR, 5FC-resistant.

Table 5.6. Polymorphic nucleotides and amino acid substitutions in *CdFCY24*

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
Eg201	21	AAA	AAG	7	K	K
SA113	48	GTT	GTC	16	V	V
p7276	76	AAA	GAA	26	K	E
SA121	76	AAA	GAA	26	K	E
SA108	76	AAA	GAA	26	K	E
SA113	76	AAA	GAA	26	K	E
Eg201	76	AAA	GAA	26	K	E
Eg202	76	AAA	GAA	26	K	E
SA121	123	GGG	GGA	41	G	G
SA108	123	GGG	GGA	41	G	G
SA113	123	GGG	GGA	41	G	G
Eg201	123	GGG	GGA	41	G	G
Eg202	123	GGG	GGA	41	G	G
SA121	130	GTC	ATC	44	V	I
SA108	130	GTC	ATC	44	V	I
SA113	130	GTC	ATC	44	V	I
Eg201	130	GTC	ATC	44	V	I
Eg202	130	GTC	ATC	44	V	I
p7276	156	GAA	GAG	52	E	E
SA113	174	CTT	CTA	58	L	L
Eg202	211	ACT	GCT	71	T	A
SA121	261	TCA	TCT	87	S	S
SA108	261	TCA	TCT	87	S	S
SA113	261	TCA	TCT	87	S	S
Eg201	261	TCA	TCT	87	S	S
Eg202	261	TCA	TCT	87	S	S
p7276	267	TTT	TTC	89	F	F
p7276	354	GCT	GCC	118	A	A
p7276	411	TTT	TTC	137	F	F
SA121	411	TTT	TTC	137	F	F
SA108	411	TTT	TTC	137	F	F
SA113	411	TTT	TTC	137	F	F
Eg201	411	TTT	TTC	137	F	F
Eg202	411	TTT	TTC	137	F	F
SA121	438	TTT	TTC	146	F	F
SA121	448	ATT	GTT	150	I	V

Table 5.6. ***Continued***

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
Eg201	451	TGT	CGT	151	C	R
SA113	467	ATT	ACT	156	I	T
Eg201	475	TCA	CCA	159	S	P
p7276	639	TTC	TTT	213	F	F
SA121	706	ATG	GTG	236	M	V
p7276	730	AAA	GAA	244	E	K
Eg201	747	GGT	GGC	249	G	G
SA121	780	TCA	TCG	260	S	S
p7276	917	ATA	ACA	306	I	T
p7276	932	GCA	GTA	311	A	V
SA113	1003	TCA	ACA	335	S	T
SA121	1025	AAA	AGA	342	K	R
p7276	1038	GTG	GTT	346	V	V
Eg201	1070	ATT	ACT	357	I	T
SA113	1130	GTT	GCT	377	V	A
Eg201	1248	ATC	ATT	416	I	I
p7276	1299	TCT	TCA	433	S	S
p7276	1305	AAA	AAT	435	K	N
p7276	1306	GTT	ATT	436	V	I
Eg202	1390	ACA	GCA	464	T	A
SA121	1398	GGA	GGG	466	G	G
SA108	1398	GGA	GGG	466	G	G
SA113	1398	GGA	GGG	466	G	G
Eg201	1398	GGA	GGG	466	G	G
Eg202	1398	GGA	GGG	466	G	G
Eg201	1428	GGA	GGG	476	G	G
Eg202	1460	GTT	GCT	487	V	A
p7276	1543	ATA	GTA	515	I	V

[a] Nucleotide position that displays polymorphism. Positions are indicated from the first base of the ATG start codon being designated 1.

[b] Codon present at this position in the consensus *CdFCY24* sequence from the *C. dubliniensis* type strain CD36, for which the complete genome has been sequenced.

[c] Codon containing polymorphic nucleotide.

[d] Position of amino acid encoded by the codon that contains SNP. Positions are indicated from the first encoded methionine residue being designated 1.

[e] Amino acid at this position in the consensus *CdFCY24* sequence, which was obtained by translation of *CdFCY24* consensus sequence of the CD36 type strain.

[f] Amino acid residue resulting from nucleotide polymorphism.

isolates examined. None of these remaining 20 SNPs were common to more than one 5FC-resistant isolate (Table 5.6).

Of the 38 SNPs identified, 20 gave rise to amino acid substitutions in the gene product (Table 5.6 and Fig. 5.5). One of these (Lys26Glu) occurred in the 5FC-susceptible isolate p7276 and all 5FC-resistant isolates in comparison to the consensus *CdFCY24* sequence from the reference 5FC-susceptible types strain CD36. Six amino acid substitutions occurred in the 5FC-susceptible isolate p7276 exclusively, and 12 occurred amongst the five 5FC-resistant isolates, although none of these substitutions was common to more than one isolate. The remaining amino acid substitution (Val44Ile) occurred in all five of the 5FC-resistant isolates, but not in the 5FC-susceptible isolate p7276 (Table 5.6 and Fig. 5.5).

5.3.5. Sequence analyses of the *CdFUR1* genes

Comparative analyses of the sequences of the *CdFUR1* genes from five 5FC-resistant isolates (SA121, SA108, SA113, Eg201, Eg202) and the 5FC-susceptible isolate p7276 (Table 4.7) with the reference 5FC-susceptible type strain CD36 were undertaken. The *CdFUR1* sequences contained a total of five SNPs amongst these seven isolates (Table 5.7). Of these five SNPs, four occurred in the 5FC-susceptible isolate p7276 and all 5FC-resistant isolates examined, in comparison to the consensus *CdFUR1* sequence from the 5FC-susceptible type strain CD36. The remaining SNP occurred in the 5FC-resistant isolate SA121. No amino acid changes resulted from any of the five SNPs identified (Table 5.7 and Fig. 5.6).

5.3.6. Accumulation of [^{3}H] 5FC in *C. dubliniensis* isolates

Although amino acid substitutions were observed in the 5FC-resistant isolates throughout the predicted proteins encoded by the *CdFCY21* and *CdFCY22* permease genes, no common amino acid substitutions could be attributed to the 5FC-resistant phenotype. The putative vitamin B transporters, encoded by the *CdFCY23* and *CdFCY24* genes, displayed some amino acid substitutions that may be associated with the 5FC-resistant phenotype, although the exact role of these permeases in the uptake of cytosine is unknown. Uptake of 5FC by three 5FC-resistant isolates [SA113, SA121, and Eg202 (Table 5.1)] and three 5FC-susceptible isolates [CD36, p7276, and SA105 (Table 5.1)] was assessed in order to determine if 5FC was taken into cells of the 5FC-resistant isolates by incubating mid-log phase cells with [^{3}H] 5FC. The intracellular

radioactivity resulting from [^{3}H] 5FC uptake in the isolates was assessed over a 30 min time frame, at 10 min intervals. Each assay was performed in duplicate on the same occasion (with the exception of isolate p7276 which was only assayed once), comparing a 5FC-susceptible isolate to a 5FC-resistant isolate during each assay.

The intracellular radioactivity in 5FC-resistant SA113 cells increased from 109 cpm × 10^3 per 10^8 cells at the beginning of the assay to 773 cpm × 10^3 per 10^8 cells at the final time point of the assay. At the same time, the intracellular radioactivity in 5FC-susceptible CD36 cells decreased from 107 cpm × 10^3 per 10^8 cells at the beginning of the assay to 58 cpm × 10^3 per 10^8 cells at the final time point (Fig. 5.7A). In a separate assay, after 30 min, the intracellular radioactivity in 5FC-resistant Eg202 cells had increased by 321 cpm × 10^3 per 10^8 cells in comparison to the intracellular radioactivity at the start of the assay. At the same time, the intracellular radioactivity in the 5FC-susceptible SA105 cells decreased by 90 cpm × 10^3 per 10^8 cells (Fig. 5.7B). In a third assay, the 5FC-resistant isolate SA121 was observed to contain higher levels of intracellular radioactivity at the beginning of the assay than the 5FC-susceptible isolate p7276, and the 5FC-resistant isolate appeared to contain relatively constant levels of the labelled drug in comparison to the other 5FC-resistant isolates, with intracellular radioactivity levels falling by 163 cpm × 10^3 per 10^8 cells in the final 10 min of the assay. In comparison, the intracellular radioactivity of the 5FC-susceptible isolate p7276 cells fell by 110 cpm × 10^3 per 10^8 cells over the 30 min. However, the intracellular radioactivity in the cells were observed to be approximately 100 cpm × 10^3 per 10^8 cells higher at the beginning of the assay than observed in 5FC-susceptible isolates from previous assays (Fig. 5.7C). In each assay, 5FC-resistant isolates were observed to take up higher levels of the labelled drug in comparison to the 5FC-susceptible isolates (Fig. 5.7D).

```
Consensus CdFCY24 MEVATGKDSIVSQDIVSSNETDIHHKTPKYINWIYNLDKWGVEVRGIERVSEQERIELGQ 60
p7276 (5FCS)      MEVATGKDSIVSQDIVSSNETDIHHETPKYINWIYNLDKWGVEVRGIERVSEQERIELGQ 60
SA121 (5FCR)      MEVATGKDSIVSQDIVSSNETDIHHETPKYINWIYNLDKWGVEIRGIERVSEQERIELGQ 60
SA108 (5FCR)      MEVATGKDSIVSQDIVSSNETDIHHETPKYINWIYNLDKWGVEIRGIERVSEQERIELGQ 60
SA113 (5FCR)      MEVATGKDSIVSQDIVSSNETDIHHETPKYINWIYNLDKWGVEIRGIERVSEQERIELGQ 60
Eg201 (5FCR)      MEVATGKDSIVSQDIVSSNETDIHHETPKYINWIYNLDKWGVEIRGIERVSEQERIELGQ 60
Eg202 (5FCR)      MEVATGKDSIVSQDIVSSNETDIHHETPKYINWIYNLDKWGVEIRGIERVSEQERIELGQ 60
                  ************************* ***************** ****************

Consensus CdFCY24 KIPTWHLFIQTLGLWWSACGGLTTMSSFFLPTLLYGLNLRDAMISGFIGMIIGCLVPAYS 120
p7276 (5FCS)      KIPTWHLFIQTLGLWWSACGGLTTMSSFFLPTLLYGLNLRDAMISGFIGMIIGCLVPAYS 120
SA121 (5FCR)      KIPTWHLFIQTLGLWWSACGGLTTMSSFFLPTLLYGLNLRDAMISGFIGMIIGCLVPAYS 120
SA108 (5FCR)      KIPTWHLFIQTLGLWWSACGGLTTMSSFFLPTLLYGLNLRDAMISGFIGMIIGCLVPAYS 120
SA113 (5FCR)      KIPTWHLFIQTLGLWWSACGGLTTMSSFFLPTLLYGLNLRDAMISGFIGMIIGCLVPAYS 120
Eg201 (5FCR)      KIPTWHLFIQTLGLWWSACGGLTTMSSFFLPTLLYGLNLRDAMISGFIGMIIGCLVPAYS 120
Eg202 (5FCR)      KIPTWHLFIQALGLWWSACGGLTTMSSFFLPTLLYGLNLRDAMISGFIGMIIGCLVPAYS 120
                  ********** ************************************************

Consensus CdFCY24 STMGPKSGCRQMVTARFLFGQWGVKFVALICIVGGIGWSVVNCVLGGQMLLAINHNISLA 180
p7276 (5FCS)      STMGPKSGCRQMVTARFLFGQWGVKFVALICIVGGIGWSVVNCVLGGQMLLAINHNISLA 180
SA121 (5FCR)      STMGPKSGCRQMVTARFLFGQWGVKFVALVCIVGGIGWSVVNCVLGGQMLLAINHNISLA 180
SA108 (5FCR)      STMGPKSGCRQMVTARFLFGQWGVKFVALICIVGGIGWSVVNCVLGGQMLLAINHNISLA 180
SA113 (5FCR)      STMGPKSGCRQMVTARFLFGQWGVKFVALICIVGGTGWSVVNCVLGGQMLLAINHNISLA 180
Eg201 (5FCR)      STMGPKSGCRQMVTARFLFGQWGVKFVALIRIVGGIGWPVVNCVLGGQMLLAINHNISLA 180
Eg202 (5FCR)      STMGPKSGCRQMVTARFLFGQWGVKFVALICIVGGIGWSVVNCVLGGQMLLAINHNISLA 180
                  ****************************  **** ** ********************

Consensus CdFCY24 VGIVVIAIISLIVAIFGIKVLLKFQTIFSIPIFIASILFYVVVCQKANYIIESNKMINDA 240
p7276 (5FCS)      VGIVVIAIISLIVAIFGIKVLLKFQTIFSIPIFIASILFYVVVCQKANYIIESNKMINDA 240
SA121 (5FCR)      VGIVVIAIISLIVAIFGIKVLLKFQTIFSIPIFIASILFYVVVCQKANYIIESNKVINDA 240
SA108 (5FCR)      VGIVVIAIISLIVAIFGIKVLLKFQTIFSIPIFIASILFYVVVCQKANYIIESNKMINDA 240
SA113 (5FCR)      VGIVVIAIISLIVAIFGIKVLLKFQTIFSIPIFIASILFYVVVCQKANYIIESNKMINDA 240
Eg201 (5FCR)      VGIVVIAIISLIVAIFGIKVLLKFQTIFSIPIFIASILFYVVVCQKANYIIESNKMINDA 240
Eg202 (5FCR)      VGIVVIAIISLIVAIFGIKVLLKFQTIFSIPIFIASILFYVVVCQKANYIIESNKMINDA 240
                  ************************************************************ ****

Consensus CdFCY24 GYSKVTSRGNWLSYFSLCYSVTATWGSGAADYYILYPASTPSYQIFLITFLGIAVPSTFV 300
p7276 (5FCS)      GYSEVTSRGNWLSYFSLCYSVTATWGSGAADYYILYPASTPSYQIFLITFLGIAVPSTFV 300
SA121 (5FCR)      GYSKVTSRGNWLSYFSLCYSVTATWGSGAADYYILYPASTPSYQIFLITFLGIAVPSTFV 300
SA108 (5FCR)      GYSKVTSRGNWLSYFSLCYSVTATWGSGAADYYILYPASTPSYQIFLITFLGIAVPSTFV 300
SA113 (5FCR)      GYSKVTSRGNWLSYFSLCYSVTATWGSGAADYYILYPASTPSYQIFLITFLGIAVPSTFV 300
Eg201 (5FCR)      GYSKVTSRGNWLSYFSLCYSVTATWGSGAADYYILYPASTPSYQIFLITFLGIAVPSTFV 300
Eg202 (5FCR)      GYSKVTSRGNWLSYFSLCYSVTATWGSGAADYYILYPASTPSYQIFLITFLGIAVPSTFV 300
                  *** ************************************************************

Consensus CdFCY24 AIAGTICGNVALSYQPWNDAYNEYGVGGLIVGTFSHWGKFGKFVAVLLYISLICNNIMNT 360
p7276 (5FCS)      AIAGTTCGNVVLSYQPWNDAYNEYGVGGLIVGTFSHWGKFGKFVAVLLYISLICNNIMNT 360
SA121 (5FCR)      AIAGTICGNVALSYQPWNDAYNEYGVGGLIVGTFSHWGKFGRFVAVLLYISLICNNIMNT 360
SA108 (5FCR)      AIAGTICGNVALSYQPWNDAYNEYGVGGLIVGTFSHWGKFGKFVAVLLYISLICNNIMNT 360
SA113 (5FCR)      AIAGTICGNVALSYQPWNDAYNEYGVGGLIVGTFTHWGKFGKFVAVLLYISLICNNIMNT 360
Eg201 (5FCR)      AIAGTICGNVALSYQPWNDAYNEYGVGGLIVGTFSHWGKFGKFVAVLLYISLICNNTMNT 360
Eg202 (5FCR)      AIAGTICGNVALSYQPWNDAYNEYGVGGLIVGTFSHWGKFGKFVAVLLYISLICNNIMNT 360
                  ***** **** ********************** ****** ************** ***

Consensus CdFCY24 YSVAFEFQLIDLRLTYVPRWIWATIVTVIYLVLSVCGRYHFLTILSNFLPMLGYWITMYI 420
p7276 (5FCS)      YSVAFEFQLIDLRLTYVPRWIWATIVTVIYLVLSVCGRYHFLTILSNFLPMLGYWITMYI 420
SA121 (5FCR)      YSVAFEFQLIDLRLTYVPRWIWATIVTVIYLVLSVCGRYHFLTILSNFLPMLGYWITMYI 420
SA108 (5FCR)      YSVAFEFQLIDLRLTYVPRWIWATIVTVIYLVLSVCGRYHFLTILSNFLPMLGYWITMYI 420
SA113 (5FCR)      YSVAFEFQLIDLRLTYAPRWIWATIVTVIYLVLSVCGRYHFLTILSNFLPMLGYWITMYI 420
Eg201 (5FCR)      YSVAFEFQLIDLRLTYVPRWIWATIVTVIYLVLSVCGRYHFLTILSNFLPMLGYWITMYI 420
Eg202 (5FCR)      YSVAFEFQLIDLRLTYVPRWIWATIVTVIYLVLSVCGRYHFLTILSNFLPMLGYWITMYI 420
                  **************** ********************************************

Consensus CdFCY24 VLLLEENLIFRSSFKVRKLHEREFDGDYKQMYNWMNWNKPKGRTLGFAACLAFLCGCAGA 480
p7276 (5FCS)      VLLLEENLIFRSSFNIRKLHEREFDGDYKQMYNWMNWNKPKGRTLGFAACLAFLCGCAGA 480
SA121 (5FCR)      VLLLEENLIFRSSFKVRKLHEREFDGDYKQMYNWMNWNKPKGRTLGFAACLAFLCGCAGA 480
SA108 (5FCR)      VLLLEENLIFRSSFKVRKLHEREFDGDYKQMYNWMNWNKPKGRTLGFAACLAFLCGCAGA 480
SA113 (5FCR)      VLLLEENLIFRSSFKVRKLHEREFDGDYKQMYNWMNWNKPKGRTLGFAACLAFLCGCAGA 480
Eg201 (5FCR)      VLLLEENLIFRSSFKVRKLHEREFDGDYKQMYNWMNWNKPKGRTLGFAACLAFLCGCAGA 480
Eg202 (5FCR)      VLLLEENLIFRSSFKVRKLHEREFDGDYKQMYNWMNWNKPKGRALGFAACLAFLCGCAGA 480
                  **************  *************************** ****************
```

Figure 5.5. Alignment of amino acid sequences for the *CdFCY24* gene in 5FC-susceptible and 5FC-resistant isolates.

Sequences were compared with that of the consensus *CdFCY24* sequence from the *C. dubliniensis* type strain CD36, for which the entire genome sequence has been determined (http://www.sanger.ac.uk.sequencing/Candida/dubliniensis/). Amino acid transitions that occurred in 5FC-resistant isolates only are highlighted in red, bold, underlined typeface, amino acid transitions that occurred in 5FC-susceptible isolates exclusively are highlighted in blue, bold, underlined typeface. Amino acid transitions that occurred in both the 5FC-susceptible isolate p7276 and the five 5FC-resistant isolates in contrast to the consensus *CdFCY24* sequence from the 5FC-susceptible *C. dubliniensis* type strain CD36 are highlighted in green, bold, underlined typeface. Twenty amino acid transitions were observed amongst the seven isolates examined. Six amino acid transitions occurred in the 5FC-susceptible isolate exclusively, and 12 occurred amongst the five 5FC-resistant isolates, although none of these transitions were common to more than one isolate. One amino acid transition (Val44Ile) occurred in all of the 5FC-resistant isolates sequenced and was not present in the 5FC-susceptible isolate p7276 or the consensus *CdFCY24* sequence from the 5FC-susceptible *C. dubliniensis* type strain CD36 (Table 5.6). The remaining amino acid transition (Lys26Glu) occurred in the 5FC-susceptible isolate p7276 and all five 5FC-resistant examined in contrast to the 5FC-susceptible *C. dubliniensis* type strain CD36 (Table 5.6). Abbreviations: 5FCS, 5FC-susceptible; 5FCR, 5FC-resistant.

Table 5.7. Polymorphic nucleotides and amino acid changes in *CdFUR1*

Isolate	SNP position[a]	Reference codon[b]	Polymorphic codon[c]	Amino acid position[d]	Amino acid encoded by consensus codon[e]	Amino acid encoded by polymorphic codon[f]
p7276	156	GGT	GGC	52	G	G
SA121	156	GGT	GGC	52	G	G
SA108	156	GGT	GGC	52	G	G
SA113	156	GGT	GGC	52	G	G
Eg201	156	GGT	GGC	52	G	G
Eg202	156	GGT	GGC	52	G	G
p7276	222	GCA	GCC	74	A	A
SA121	222	GCA	GCC	74	A	A
SA108	222	GCA	GCC	74	A	A
SA113	222	GCA	GCC	74	A	A
Eg201	222	GCA	GCC	74	A	A
Eg202	222	GCA	GCC	74	A	A
p7276	318	GGG	GGA	106	G	G
SA121	318	GGG	GGA	106	G	G
SA108	318	GGG	GGA	106	G	G
SA113	318	GGG	GGA	106	G	G
Eg201	318	GGG	GGA	106	G	G
Eg202	318	GGG	GGA	106	G	G
p7276	579	GGA	GGG	193	G	G
SA121	579	GGA	GGG	193	G	G
SA108	579	GGA	GGG	193	G	G
SA113	579	GGA	GGG	193	G	G
Eg201	579	GGA	GGG	193	G	G
SA121	612	ATT	ATA	204	I	I

[a] Nucleotide position that displays polymorphism. Positions are indicated from the first base of the ATG start codon being designated 1.

[b] Codon present at this position in the consensus *CdFUR1* sequence from the *C. dubliniensis* type strain CD36, for which the complete genome has been sequenced.

[c] Codon containing polymorphic nucleotide.

[d] Position of amino acid encoded by the codon that contains SNP. Positions are indicated from the first encoded methionine residue being designated 1.

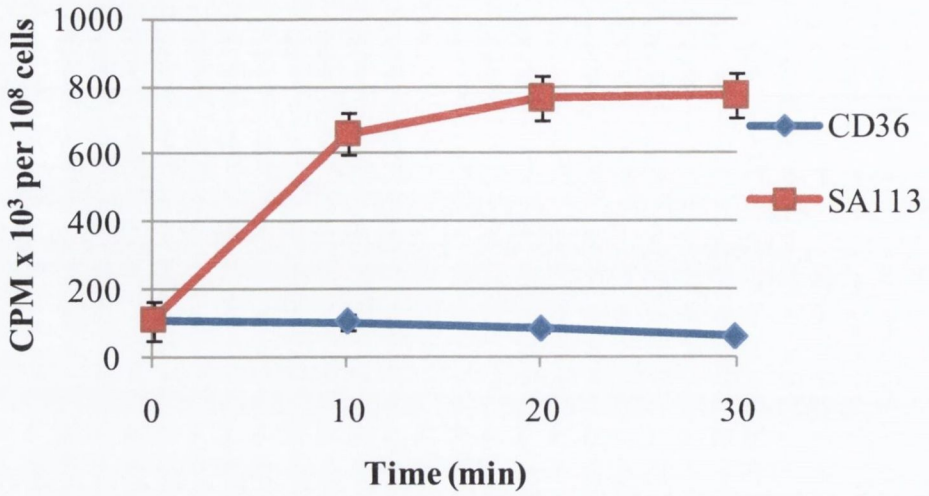
[e] Amino acid at this position in the consensus *CdFUR1* sequence, which was obtained by translation of *CdFUR1* consensus sequence of the CD36 type strain.

[f] Amino acid residue resulting from nucleotide polymorphism.

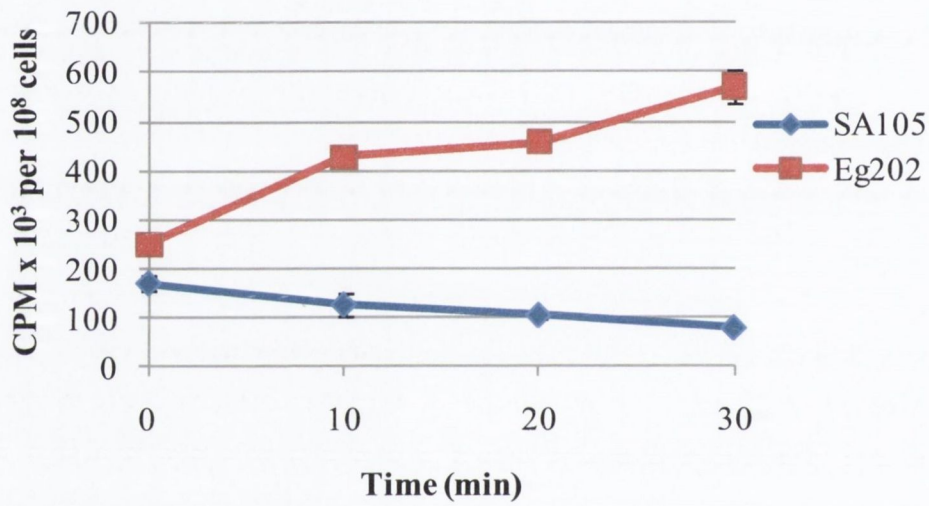
```
Consensus CdFUR1 MSVAKAVSKNVILLPQTNQLIGLYSIIRDQRTKRGDFVFYSDRIIRLLVEEGLNQLPVEE 60
p7276 (5FCS)     MSVAKAVSKNVILLPQTNQLIGLYSIIRDQRTKRGDFVFYSDRIIRLLVEEGLNQLPVEE 60
SA121 (5FCR)     MSVAKAVSKNVILLPQTNQLIGLYSIIRDQRTKRGDFVFYSDRIIRLLVEEGLNQLPVEE 60
SA108 (5FCR)     MSVAKAVSKNVILLPQTNQLIGLYSIIRDQRTKRGDFVFYSDRIIRLLVEEGLNQLPVEE 60
SA113 (5FCR)     MSVAKAVSKNVILLPQTNQLIGLYSIIRDQRTKRGDFVFYSDRIIRLLVEEGLNQLPVEE 60
Eg201 (5FCR)     MSVAKAVSKNVILLPQTNQLIGLYSIIRDQRTKRGDFVFYSDRIIRLLVEEGLNQLPVEE 60
Eg202 (5FCR)     MSVAKAVSKNVILLPQTNQLIGLYSIIRDQRTKRGDFVFYSDRIIRLLVEEGLNQLPVEE 60
                 ************************************************************

Consensus CdFUR1 AIIKCHGGYEYKGAKFLGKICGVSIVRAGESMEMGLRDCCRSVRIGKILIQRDEETALPK 120
p7276 (5FCS)     AIIKCHGGYEYKGAKFLGKICGVSIVRAGESMEMGLRDCCRSVRIGKILIQRDEETALPK 120
SA121 (5FCR)     AIIKCHGGYEYKGAKFLGKICGVSIVRAGESMEMGLRDCCRSVRIGKILIQRDEETALPK 120
SA108 (5FCR)     AIIKCHGGYEYKGAKFLGKICGVSIVRAGESMEMGLRDCCRSVRIGKILIQRDEETALPK 120
SA113 (5FCR)     AIIKCHGGYEYKGAKFLGKICGVSIVRAGESMEMGLRDCCRSVRIGKILIQRDEETALPK 120
Eg201 (5FCR)     AIIKCHGGYEYKGAKFLGKICGVSIVRAGESMEMGLRDCCRSVRIGKILIQRDEETALPK 120
Eg202 (5FCR)     AIIKCHGGYEYKGAKFLGKICGVSIVRAGESMEMGLRDCCRSVRIGKILIQRDEETALPK 120
                 ************************************************************

Consensus CdFUR1 LFYEKLPEDISERYVFLLDPMLATGGSAMMAVEVLLARGVKMDRILFLNLLAAPEGIKAF 180
p7276 (5FCS)     LFYEKLPEDISERYVFLLDPMLATGGSAMMAVEVLLARGVKMDRILFLNLLAAPEGIKAF 180
SA121 (5FCR)     LFYEKLPEDISERYVFLLDPMLATGGSAMMAVEVLLARGVKMDRILFLNLLAAPEGIKAF 180
SA108 (5FCR)     LFYEKLPEDISERYVFLLDPMLATGGSAMMAVEVLLARGVKMDRILFLNLLAAPEGIKAF 180
SA113 (5FCR)     LFYEKLPEDISERYVFLLDPMLATGGSAMMAVEVLLARGVKMDRILFLNLLAAPEGIKAF 180
Eg201 (5FCR)     LFYEKLPEDISERYVFLLDPMLATGGSAMMAVEVLLARGVKMDRILFLNLLAAPEGIKAF 180
Eg202 (5FCR)     LFYEKLPEDISERYVFLLDPMLATGGSAMMAVEVLLARGVKMDRILFLNLLAAPEGIKAF 180
                 ************************************************************

Consensus CdFUR1 HEKYPDVKIITGGIDEKLDEDKYIVPGLGDFGDRYYCI 218
p7276 (5FCS)     HEKYPDVKIITGGIDEKLDEDKYIVPGLGDFGDRYYCI 218
SA121 (5FCR)     HEKYPDVKIITGGIDEKLDEDKYIVPGLGDFGDRYYCI 218
SA108 (5FCR)     HEKYPDVKIITGGIDEKLDEDKYIVPGLGDFGDRYYCI 218
SA113 (5FCR)     HEKYPDVKIITGGIDEKLDEDKYIVPGLGDFGDRYYCI 218
Eg201 (5FCR)     HEKYPDVKIITGGIDEKLDEDKYIVPGLGDFGDRYYCI 218
Eg202 (5FCR)     HEKYPDVKIITGGIDEKLDEDKYIVPGLGDFGDRYYCI 218
                 ************************************
```

Figure 5.6. Alignment of amino acid sequences for the *CdFUR1* gene in 5FC-susceptible and 5FC-resistant isolates.

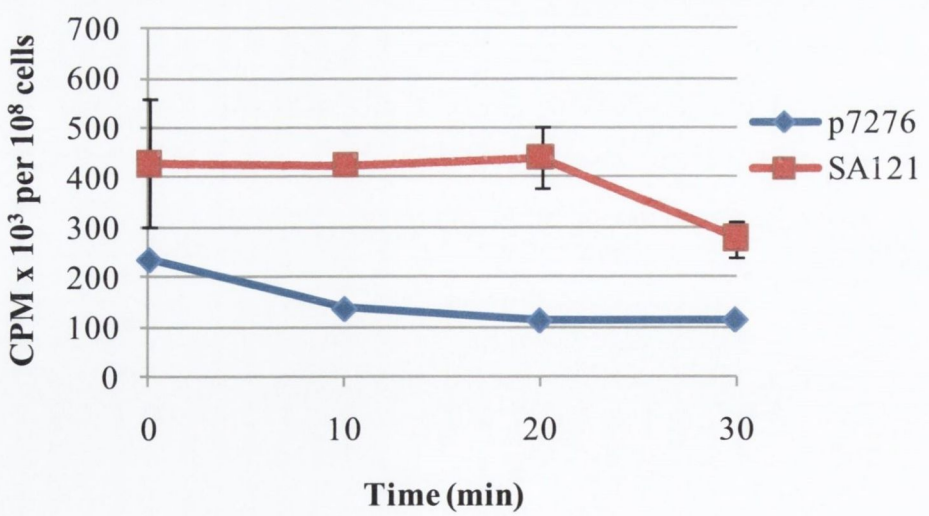
Sequences were compared with that of the consensus *CdFUR1* sequence from the *C. dubliniensis* type strain CD36, for which the entire genome sequence has been determined (http://www.sanger.ac.uk.sequencing/Candida/dubliniensis/). Although five SNPs were observed amongst the *CdFUR1* sequences for the seven isolates examined (Table 5.7), no amino acid transitions were observed. Abbreviations: 5FCS, 5FC-susceptible; 5FCR, 5FC-resistant.


1000
800
600
400
200
0
CPM x 10³ per 10⁸ cells
0
10
20
30
Time (min)
CD36
SA113

B


700
600
500
400
300
200
100
0
CPM x 10³ per 10⁸ cells
0
10
20
30
Time (min)
SA105
Eg202

C


700
600
500
400
300
200
100
0
CPM x 10³ per 10⁸ cells
0
10
20
30
Time (min)
p7276
SA121

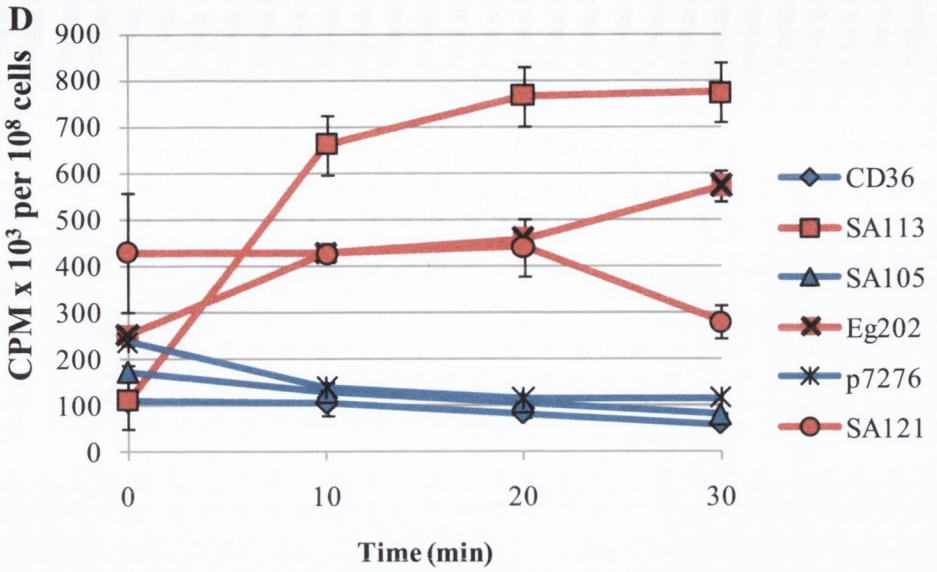


Figure 5.7. Uptake of radiolabelled [^{3}H] 5FC in 5FC-resistant and 5FC-susceptible isolates.

5FC-resistant isolates are denoted in red, and 5FC-susceptible isolates are denoted in blue. Each experiment was carried out in duplicate with the exception of p7276, which was only assayed on one occasion. Mean values are displayed for each time point, and SEM values are denoted by error bars. For each assay, intracellular radioactivity levels were at least three fold higher in the 5FC-resistant isolates than in the 5FC-susceptible isolates after 10 min incubation with the radiolabelled drug.

5.4. Discussion

All *C. dublinienisis* isolates exhibiting 5FC resistance to date have been recovered in the Middle East, in Egypt, Saudi Arabia or Kuwait, and belong to Cd25 fingerprint group III or MLST clade C3 (Pfaller *et al.*, 1999; Ahmad *et al.*, 2004; Al Mosaid *et al.*, 2005). The purpose of the present study was to investigate the genes encoding proteins that are involved or may be involved in the uptake and metabolism of 5FC, in order to identify any blocks or lesions that may result in *C. dubliniensis* clade-specific 5FC resistance. Mechanisms of antifungal resistance that have previously been identified in *Candida* species include; over-expression/mutation of the drug target, inactivation or degradation of the drug, reduced permeability to the drug, or increased drug efflux (Ghannoum & Rice, 1999; Yang & Lo, 2001; Sanglard, 2002; Sullivan *et al.*, 2004). The possibility of reduced uptake or mutation of drug targets in *C. dubliniensis* isolates demonstrating high levels of intrinsic 5FC resistance were investigated in this study by carrying out sequence analysis of the genes involved in the uptake and metabolism of 5FC. An accumulation assay was also carried out to determine 5FC uptake in the 5FC-resistant and 5FC-susceptible cells.

In contrast to isolates of *C. albicans* which show great variation in 5FC susceptibility patterns (Dodgson *et al.*, 2004; Hope *et al.*, 2004), isolates of *C. dubliniensis* exhibit either a high level of intrinsic 5FC resistance (MIC $\geq$ 128 μg/ml) or 5FC susceptibility (MIC $\leq$ 0.25 μg/ml) with little variation (Pfaller *et al.*, 1999; Al Mosaid *et al.*, 2005). All of the 5FC-resistant isolates were recovered in the Middle East, and all of these isolates tested were found to belong to MLST clade C3 or Cd25 fingerprint group III. Due to the highly clonal nature of the species, we hypothesised that clade-specific 5FC resistance was likely to be mediated by a common mechanism(s) amongst these isolates.

Genetic sequence analysis of the two genes encoding the principle permeases associated with cytosine transport in yeasts, *CdFCY21* and *CdFCY22,* was undertaken as these have previously been associated with 5FC resistance in the haploid *C. lusitaniae* (Noël *et al.*, 2003; Chapeland-Leclerc *et al.*, 2005; Papon *et al.*, 2007). The DNA and amino acid sequence of both genes were compared between one 5FC-susceptible and five 5FC-resistant *C. dubliniensis* isolates from the Middle East and from the same MLST clade, clade C3. All sequences were compared to the consensus

gene sequence of the 5FC-susceptible *C. dubliniensis* type strain CD36 (Figs. 5.2 and 5.3). Although several amino acid substitutions resulted from SNPs detected amongst resistant isolates in these genes (Tables 5.3 and 5.4), only one common amino acid substitution was conserved amongst the *CdFCY21* gene product of two 5FC-resistant isolates (Eg201 and Eg202; Fig. 5.2 and Table 5.3).

In the absence of the *FCY21*- and *FCY22*-encoded permeases in *S. cerevisiae*, several other permeases have been shown to substitute for their function (Paluszynski *et al.*, 2006), including the *TPN1*, *FUR4* and *yOR071c* gene products (Paluszynski *et al.*, 2006). The *TPN1* gene has been shown to be closely related to the *FCY23* and *FCY24* genes of *C. albicans*, which may also be able to transport cytosine (Hope *et al.*, 2004). The genetic sequence of the *CdFCY23* and *CdFCY24* genes were therefore compared between a 5FC-susceptible isolate and five 5FC-susceptible isolates of *C. dubliniensis* recovered in the Middle East. All sequences were compared to the consensus gene sequence of the 5FC-susceptible *C. dubliniensis* type strain CD36 (Figs. 5.4 and 5.5). These DNA and amino acid sequence analyses identified amino acid substitutions in both genes that were specific to the 5FC-resistant isolates (Tables 5.4 and 5.5). A radical Ser280Leu substitution was observed exclusively in the *CdFCY23* gene of all 5FC-resistant isolates examined (Fig. 5.4). Furthermore, a Val44Ile amino acid substitution was observed exclusively in the *CdFCY24* gene of all 5FC-resistant isolates examined, resulting in replacement of one aliphatic non-polar residue with another residue of similar polarity (Fig. 5.5). It is perhaps unlikely that an amino acid substitution in these permeases could play a role in 5FC resistance by preventing/reducing drug uptake, as they appear to play a role in cytosine transport only in the absence of the *FCY21*- and *FCY22*-encoded permeases in *S. cerevisiae* (Paluszynski *et al.*, 2006). In support of this suggestion, data from [^{3}H] 5FC accumulation assays suggested that purine-cytosine permeases are functional in the 5FC-resistant isolates of *C. dubliniensis*, as these isolates were shown to be capable of 5FC-uptake, and appeared to accumulate more of the radiolabelled drug than their 5FC-susceptible counterparts. However, this latter finding could be due to the metabolism of the radiolabelled drug in the 5FC-susceptible isolates; the radioactive component of the ^{3}H [5FC] may be lost from the cell upon conversion of the ^{3}H [5FC] to FU. The accumulation of intracellular radioactivity in the 5FC-resistant cells may be due to a block/lesion in the metabolic pathway of 5FC, possibly occurring at the point of cytosine deaminase activity. Such a lesion would block the intracellular deamination of

5FC to the toxic 5FU, resulting in the intracellular accumulation of the nontoxic 5FC prodrug in the 5FC-resistant cells. These findings parallel with the DNA sequence analyses of *CdFCY21* and *CdFCY22*, which identified only one common SNP amongst the *CdFCY21* gene of two 5FC-resistant isolates (Eg201 and Eg202; Table 5.3) resulting in a Phe502Leu amino acid substitution.

The most commonly identified cause of 5FC resistance in the closest relative of *C. dubliniensis; C. albicans*, occurs as a result of an Arg101Cys amino acid substitution in the *FUR1* gene product (Dodgson *et al.*, 2004; Hope *et al.*, 2004). According to crystallographic studies in *T. gondii*, the tetrameric UPRT encoded by *FUR1* is composed of two tightly packed dimers (Schumacher *et al.*, 1998), which are held together at the interface by an arginine and glutamic acid salt bridge, thus stabilising the quaternary structures of the tetrameric UPRT protein. The arginine cognate of the *T. gondii* salt bridge in the *C. albicans* UPRT occurs at position 101 (Dodgson *et al.*, 2004; Hope *et al.*, 2004). It has therefore been suggested that the Arg101Cys amino acid substitution disrupts the dimer interface of tetramer, leading to suboptimal activity of the enzyme (Dodgson *et al.*, 2004; Hope *et al.*, 2004). The genetic sequence of the *FUR1* homologue in *C. dubliniensis* has been investigated previously (Al Mosaid *et al.*, 2005), and also in the current study. Whilst SNPs were observed in the *CdFUR1* gene in both studies, none of these resulted in any amino acid substitutions, suggesting that the SNPs in *CdFUR1* are an unlikely cause of clade-specific 5FC resistance in *C. dubliniensis*.

Data from the present study suggests that mutations in the *FCY2* homologues *CdFCY21* and *CdFCY22*, or the *TPN1* homologues *CdFCY23* and *CdFCY24*, are unlikely to be causes of intrinsic resistance to 5FC. This suggestion is supported by the results of the ^{3}H [5FC] accumulation assay, as well as DNA and amino acid sequence analyses of the individual genes and gene products. Likewise, the *CdFUR1*-encoded UPRT is unlikely to contribute to resistance, as no amino acid substitutions were noted in this gene amongst 5FC-resistant isolates in either the present study or in previous studies (Al Mosaid *et al.*, 2005).

The present study identified SNPs in the genes encoding proteins involved in 5FC uptake or metabolism, although it did not elucidate any blocks/lesions that could be associated with *C. dubliniensis* clade-specific resistance. The remaining enzyme involved in 5FC metabolism to be investigated was the cytosine deaminase, encoded by *FCY1* in *S. cerevisiae* and *C. lusitaniae*, and by the *FCA1* gene homologue in *C.*

albicans. Amino acid substitutions in the *FCA1* gene have previously been associated with both high level and intermediate levels of 5FC resistance in individual isolates of *C. albicans* lacking the Arg101Cys substitution in the UPRT (Hope *et al.*, 2004), however the homologous *FCA1* gene in *C. dubliniensis* (*CdFCA1*) has previously been examined in 5FC-resistant isolates and identified no amino acid substitutions that could account for the resistant phenotype (Al Mosaid *et al.*, 2005). Despite this previous finding, the DNA sequence of the *CdFCA1*-encoded cytosine deaminase was further investigated using additional oligonucleotide primers to those used previously in order to detect any SNPs which may result in *C. dubliniensis* clade-specific 5FC resistance (Chapter 6).

Chapter 6

Investigation of the molecular mechanism(s) of clade-specific 5-fluorocytosine resistance in *Candida dubliniensis*

Part II:

The role of the *CdFCA1*-encoded cytosine deaminase in clade-specific 5-fluorocytosine resistance

6.1. Introduction

6.1.1. Molecular mechanisms of 5FC resistance in *C. albicans*

In the closest relative of *C. dubliniensis*; *C. albicans*, resistance to 5FC is mediated by a reduction in the activity of either the cytosine deaminase encoded by *FCA1*, or in the activity of UPRT encoded by *FUR1* (Polak & Scholer, 1975; Whelan & Kerridge, 1984; Hope *et al.*, 2004). Two different research groups reported that in the majority of 5FC resistance *C. albicans* isolates, 5FC resistance is associated with a homozygous single amino acid substitution, Arg101Cys, in UPRT (Dodgson *et al.*, 2004; Hope *et al.*, 2004). However, other 5FC-resistant *C. albicans* isolates lack this substitution (Hope *et al.*, 2004). One such isolate (MIC_{50} > 64 μg/ml) was reported to contain a homozygous Gly28Asp substitution in the cytosine deaminase gene, and a Ser29Leu amino acid substitution was also observed in the same gene of another *C. albicans* isolate exhibiting intermediate 5FC resistance (MIC_{50} 4 μg/ml) (Hope *et al.*, 2004).

6.1.2. Investigation of the molecular mechanism(s) of 5FC resistance in *C. dubliniensis*

In *C. dubliniensis* the DNA sequence of the *CdFUR1* gene encoding UPRT of four 5FC-resistant and five 5FC-susceptible isolates from the Middle East was determined previously, and while several SNPs were identified, no amino acid substitutions were observed between the isolates (Al Mosaid *et al.*, 2005). This finding was confirmed in the current study (Chapter 5, Section 5.3.5). As described previously, the DNA and amino acid sequence of the purine-cytosine permease genes *CdFCY21* and *CdFCY22* were compared between five 5FC-resistant isolates and two 5FC-susceptible isolates and identified only one common amino acid substitution in the 5FC-resistant isolates, occurring in the *CdFCY21* gene of 2/5 (Eg201 and Eg202) 5FC-resistant isolates sequenced (Chapter 5, Sections 5.3.1 and 5.3.2). Sequence analysis of the *FCY23* and *FCY24* encoded permeases was also carried out in order to identify any alterations that may play a role in 5FC resistance, despite their lesser importance in 5FC-transport, (Chapter 5, Sections 5.3.3 and 5.3.4). These analyses identified amino acid substitutions that occurred in all 5FC-resistant isolates and in neither of the 5FC-susceptible isolates sequenced, however, and an accumulation assay using radiolabelled

[^{3}H] 5FC confirmed the uptake and accumulation of 5FC in the 5FC-resisant isolates, arguing against a permease-mediated role in *C. dubliniensis* clade-specific 5FC resistance.

6.1.3. The yeast cytosine deaminase

The action of cytosine deaminase with its substrate, 5FC, has been investigated for use in cancer therapy as the product of this reaction, 5FU, is toxic to cells undergoing rapid cell division (Hayden *et al.*, 1998). Cancerous cells demonstrate a higher affinity for 5FU than healthy cells, and 5FU has been widely used in the treatment of malignancies. Using suicide gene therapy methods to introduce cytosine deaminase into cancerous cells and then treating these cells with 5FC has been used as a method of treating malignant gliomas in a murine model (Lv *et al.*, 2009).

The cytosine deaminase is the sole reason for 5FC toxicity to yeasts such as *Candida* and *Cryptococcus*. This homodimeric zinc metalloenzyme is not present in mammalian cells and so the non-toxic prodrug form, 5FC, presents no danger to humans. The cytosine deaminase is primarily responsible for the conversion of cytosine to uracil, and also converts 5FC to the toxic metabolite, 5FU. The enzyme is composed of five central β-sheets that are flanked by two α-helices on one side, and three α-helices on the other. At the centre of the tightly packed dimer there are two single active sites; each of these contain a single tetrahedrally coordinated zinc ion, that is bound to a water molecule in the absence of a substrate (Ireton *et al.*, 2003; Ko *et al.*, 2003; Yao *et al.*, 2005). During the deamination process, the nucleophilic attack of a Zn-bound hydroxide on the cytosine substrate results in the formation of a tetrahedral intermediate, this then decomposes releasing ammonia and the uracil end product (Yao *et al.*, 2005).

6.1.4. Tetracycline-inducible gene expression

The tetracycline-regulatable systems rely on the activity of a tetracycline repressor protein (TetR), which binds to the tetracycline operator (*tetO*) in the absence of tetracycline repressing expression of the *tet* genes (Hillen & Berens, 1994). The system is reversed in the presence of tetracycline, as the drug binds to TetR altering its ability to bind to *tetO*, and resulting in transcription of the *tet* genes. This system is known as the Tet-Off system, however, five amino acid changes in the TetR protein have been shown to reverse its activity (reverse tetracycline repressor protein; rTetR), so that binding of the rTetR protein to *tetO* results in expression of the tet genes as a

Tet-On system (Gossen *et al.*, 1995). Both the Tet-Off and Tet-On systems have been modified for use in eukaryotic species by fusion of the TetR or rTetR proteins to the activation domain of a transcription factor (tTA or rtTA, respectively) for the Gal4 protein of *S. cerevisiae* (Gossen & Bujard, 1992). This Tet-On system in eukaryotes enables tetracycline-inducible expression of any genes that are placed under the control of a minimal promoter in which all activating sequences have been removed and replaced by the *tetO* sequence in the presence of the eukaryotic transactivator (rtTA) (Gossen *et al.*, 1995). This system has been utilised by Park and Morshhäuser (2005) in the design of the pNIM1 plasmid which also contains the *SAT1* gene encoding a nourseothricin resistance marker and a sequence from the *C. albicans ADH1* gene to ensure recombination into this locus. This plasmid enables tetracycline-inducible expression of genes in this pNIM1 cassette integrated into the background *ADH1* locus of *C. albicans* (Park & Morschhäuser, 2005).

6.1.5. Objectives

The purpose of the present study was to investigate the role of cytosine deaminase in *C. dubliniensis* clade-specific 5FC resistance.

- To determine if bypassing the requirement of the cytosine deaminase resulted in 5FU-susceptibility, by carrying out broth microdilution assays using 5FC and 5FU.
- To compare the DNA and amino acid sequences of the *CdFCA1* gene in 5FC-resistant and 5FC-susceptible isolates, in order to identify any changes which may result in altered *CdFCA1* activity.
- To compare *CdFCA1* expression between 5FC-susceptible and 5FC-resistant isolates by quantitative real-time PCR analysis, in order to determine if 5FC resistance could be attributed to altered *CdFCA1* expression.
- To incorporate the *CdFCA1* gene from a 5FC-susceptible isolate (hereafter called $CdFCA1^{S}$) into the *ADH1* locus of a 5FC-resistant isolate, and from a 5FC-resistant isolate (hereafter called $CdFCA1^{R}$) into the *ADH1* locus of a 5FC-susceptible isolate using a tetracycline-inducible expression plasmid created for use in *C. albicans* (Park & Morschhäuser, 2005). These strains were used to determine if 5FC susceptibility/resistance could be induced in isolates upon acquisition and expression of the respective *CdFCA1* gene.

6.2. Materials and methods

6.2.1. Isolates and culture conditions

Twenty-one epidemiologically unrelated human *C. dubliniensis* isolates were included in the present study, including nine 5FC-susceptible isolates and 12 5FC-resistant isolates (Table 5.1) as previously reported (Sullivan *et al.*, 1995; Polacheck *et al.*, 2000; Al Mosaid *et al.*, 2001; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005). The *C. dubliniensis* type strain CD36, originally isolated from the oral cavity of a HIV-infected individual in Ireland was also included as a reference isolate because the complete genome sequence of this organism has been determined (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/). Yeast and bacterial strains were routinely cultured as described in Chapter 2, Section 2.1.1. Nourseothricin (Werner Bioagents, Gena, Germany) was prepared at a concentration of 200 mg/ml in Milli-Q® Biocel-purified water and doxycycline (Sigma-Aldrich Ltd.) was prepared in Milli-Q® Biocel-purified water at 10 mg/ml. Both of these drugs were stored at -20°C.

6.2.2. Susceptibility testing

Susceptibility testing was carried out according to the CLSI (2002) as previously described in Chapter 2, Section 2.7.

6.2.3. DNA extraction

Extraction of DNA from *Candida* isolates was carried out as previously described in Chapter 2, Section 2.3.3. Extraction of DNA from bacterial cells was carried out as described in Chapter 2, Section 2.5.3. Nucleic acids were ethanol precipitated and resuspended in 50 µl of molecular grade Milli-Q® Biocel-purified water (Millipore™).

6.2.4. PCR amplification, sequencing and sequence analysis of *CdFCA1*

The complete ORF of the *C. dubliniensis CdFCA1* gene was amplified from 12 5FC-resistant isolates and nine 5FC-susceptible isolates (Table 5.1) using the oligonucleotide primers FCA1F and FCA1R which incorporated *Sal*I and *Bgl*II restriction endonuclease recognition sites, respectively (Table 6.1). Reaction mixtures contained 100 ng of purified template DNA, 1 × Expand High Fidelity buffer, 2.5 U of Expand High Fidelity PCR system enzyme mix (Roche), 0.2 mM concentrations of each

deoxynucleoside triphosphate, (Promega) and 0.2 μM concentrations of each oligonucleotide (Sigma-Aldrich Ltd.). Reaction mixtures underwent an initial denaturation step of 94°C for 10 min, followed by 35 cycles of 94°C for 30 s, 52°C for 30 s and 72°C for 1 min, followed by a final elongation step of 72°C for 7 min. The 550 bp products were purified using the GenElute™ PCR Clean-up kit (Sigma-Aldrich Ltd.) and were sequenced on both strands using the same primers that had been used for amplification.

6.2.5. DNA sequencing and analysis

DNA sequencing reactions were performed commercially by Cogenics as previously described in Chapter 2, Section 2.6. Multiple DNA and amino acid sequence alignments of *CdFCA1* genes and their encoded proteins from 5FC-susceptible and 5FC-resistant *C. dubliniensis* isolates were carried out using the CLUSTAL W sequence alignment computer program (Thompson *et al.*, 1994) available at the EMBL-EBI website (http://www.ebi.ac.uk/).

6.2.6. Transformation of *C. dubliniensis*

Candida dubliniensis strains were transformed by electroporation as previously described (Köhler *et al.*, 1997; Staib *et al.*, 2001). Cells from YPD pre-cultures were diluted 10^{-4} in 50 ml of fresh YPD medium and grown overnight at 30°C to an optical density at 600 nm (OD_{600}) of 1.6 to 2.2, which yielded the best transformation efficiency. The cells were collected by centrifugation at 2,500 × *g* and resuspended in 8 ml of water. After addition of 1 ml of 10 × TE (100 mM Tris-HCl, 10 mM EDTA; pH 7.5) and 1 ml of 1 M lithium acetate (pH 7.5), the suspension was incubated in a rotary shaker at 150 rpm for 60 min at 30°C. A 250-μl volume of 1 M dithiothreitol was then added, and the cells were incubated for a further 30 min at 30°C with shaking. After addition of 40 ml of water, the cells were centrifuged at 2,500 × *g*, washed sequentially in 50 ml of ice-cold water and 10 ml of ice-cold 1 M sorbitol, resuspended in 50 μl of 1 M sorbitol, and kept on ice. Approximately 200 ng of the linear DNA fragments was mixed with 40 μl of electrocompetent cells, and electroporation was carried out in a Bio-Rad Gene Pulser (0.2 cm cuvette, 1.6 kV, 200 Ω, 25 μF) with a Bio-Rad Pulse Controller included in the circuit. Nourseothricin-resistant transformants were selected on YPD agar plates containing nourseothricin at a concentration of 100 μg/ml. Single

Table 6.1. Oligonucleotide primers used in the study of the role of the *CdFCA1* gene in 5FC resistance

Primer	Sequence (5′-3′) [b]	Function
ADH1F	ATGCAAGCAAGCTTATTCA	PCR screening of transformants[a]
cartTA	CGGCATACTATCAGTAGTAG	
SAT	CAATGCCGCCGAGAGTAAAG	
ADH1R	CCCAAGATCTTACCTTCTTCCATT	
FCA1F	GACGCGTCGACGATATCAACGATGACATTT	*CdFCA1* cloning and PCR screening
FCA1R	CGGGATCCAGATCTTTATTCTCCAATATCTTC	
RTFCA1F	AAACGCAGGAAGATTGCCAG	Gene expression analysis
RTFCA1R	TGGCCCCTGTACACATACTACATG	
RTACT1F	AGCTCCAGAAGCTTTGTTCAGACC	
RTACT1R	CCCAGGTATTGCTGAACGTATGCA	

[a] Regions of the pNIM1-*CdFCA1* cassette and background *ADH1* locus amplified during PCR screening of transformants are displayed in Fig. 6.1.

[b] The *Sal*I and *Bgl*II restriction endonuclease recognition sites incorporated into the FCA1F and FCA1R primer sequences, respectively, are shown in underlined typeface.

colonies were picked after 48 h of growth at 30°C, restreaked on the same medium, and, after verification of the correct allelic replacement, maintained on YPD agar plates.

6.2.7. Tetracycline-inducible *CdFCA1* expression in *C. dubliniensis*

The tetracycline-inducible gene expression plasmid pNIM1 developed by Park & Morschhäuser (2005) was adapted to investigate the inducible expression of *C. dubliniensis* $CdFCA1^S$ and $CdFCA1^R$ genes in both 5FC-susceptible and 5FC-resistant *C. dubliniensis* isolates. The pNIM1 cassette was originally designed to integrate into the *C. albicans* alcohol dehydrogenase encoding gene *ADH1*. The DNA sequence of the *C. albicans ADH1* gene (GenBank accession No. CaO19.3997) was used in a BLAST search against the *C. dubliniensis* genome sequence database (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/) in order to identify a homologue in *C. dubliniensis*. A high level of sequence homology (94%) is shared by the *ADH1* ORFs of *C. albicans* and *C. dubliniensis*. The tetracycline-inducible promoter, P_{tet}, included in pNIM1 (Park & Morschhäuser, 2005), was used to drive expression of the $CdFCA1^S$ and $CdFCA1^R$ genes individually in the *ADH1* locus of a *C. dubliniensis* isolate with the opposite 5FC-phenotype. The complete coding regions of the *CdFCA1* gene from the 5FC-susceptible *C. dubliniensis* isolate p7276 and the 5FC-resistant *C. dubliniensis* isolate SA113 were amplified from genomic DNA using the primer pair FCA1F/FCA1R (as described above). The amplimers were ligated into pGEM T-Easy vector I (Promega) vector DNA and transformed into *E. coli* strain DH5α. Plasmids were recovered from transformants using the GenElute™ plasmid miniprep kit (Sigma-Aldrich Ltd.), and the cloned DNA sequenced. The complete *CdFCA1* ORF was digested from pGEM T-Easy plasmids using the *Sal*I and *Bgl*II restriction endonuclease recognition sites which were introduced upstream and downstream of the ORF, respectively, using the FCA1F/FCA1R primer pair. The *CdFCA1* fragments were then gel purified using the Wizard SV Gel and PCR clean up system (Promega) and cloned between P_{tet} and T_{ACT1} in a *Sal*I-*Bgl*II digested pNIM1 separately; incorporating the *CdFCA1* ORFs into the pNIM1 cassette, in the place of the *GFP1* gene (Park & Morschhäuser, 2005), (Fig. 6.1). The resulting pNIM1-$CdFCA1^S$ and pNIM1-$CdFCA1^R$ plasmids were transformed into *E. coli* strain DH5α for replication of the plasmid prior to purification using the GenElute™ Plasmid Miniprep kit (Sigma-Aldrich Ltd), *Sac*II-*Apa*I linearisation of the cassette, gel purification and transformation into *C. dubliniensis* as described in Section 6.2.6 (Staib *et al.*, 2001). The

5FC-resistant isolates SA113 and SA109 were transformed with pNIM1-*CdFCA1*^{S}, and the 5FC-susceptible isolate p7276 was transformed with pNIM1-*CdFCA1*^{R}. All three of these *C. dubliniensis* isolates were also transformed with the pNIM1 cassette containing the *GFP1* gene instead of the *CdFCA1* gene as a control for the disruption of the *ADH1* locus. In order to confirm the correct integration of the complete pNIM1-*CdFCA1* cassette into the *ADH1* locus in transformant derivatives, a number of PCR amplifications were carried out. The primer pairs used in these PCR amplifications were ADH1F/cartTA, SAT/FCA1R, and FCA1F/ADH1R and SAT/ADH1R (Table 6.1 and Fig. 6.1). These stepwise amplifications revealed the presence and correct integration of the full pNIM1-*CdFCA1* cassette in transformant derivatives.

6.2.8. Southern hybridisation

Southern hybridisation analysis was used to confirm the integration of the pNIM1-*CdFCA1* cassette into the background chromosomal *ADH1* locus using two separate probes; a *CdFCA1*-directed probe and a pNIM1-directed probe which was directed towards the cartTA transactivator region and the *ADH1* locus.

6.2.8.1. Synthesis of DIG-labelled probes by PCR amplification

Probes were labelled with digoxigenin (DIG) in separate PCR amplifications using the PCR DIG labelling mix (Roche) and the primer pairs ADH1F/cartTA or FCA1F/FCA1R (Table 6.1 and Fig. 6.1). Both PCR amplifications were carried out in a reaction volume of 100 µl containing 100 ng of DNA, 3.5 mM $MgCl_2$, 18 pmol of each primer, 1 × colourless GoTaq® flexi buffer (Promega), 2.5 U GoTaq® DNA polymerase (Promega) and 1 × PCR DIG labelling mix (Roche). Reaction mixtures underwent an initial denaturation step of 94°C for 10 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 1 min. These 30 cycles were followed by a final elongation step at 72°C for 10 min. For amplification of the *CdFCA1*-directed probe, the annealing temperature was reduced to 52°C. Probes were purified using the GenElute™ PCR clean up kit (Sigma-Aldrich Ltd.) and denatured by incubation at 95°C for 10 min immediately before hybridisation.

6.2.8.2. Restriction endonuclease digestion and electrophoresis of genomic DNA

Large scale restriction endonuclease digestions of genomic DNA for Southern analysis were carried out in 60 µl volumes containing 45 µg DNA and 40 U of restriction endonuclease *Hind*III and the corresponding buffer at a final 1 ×

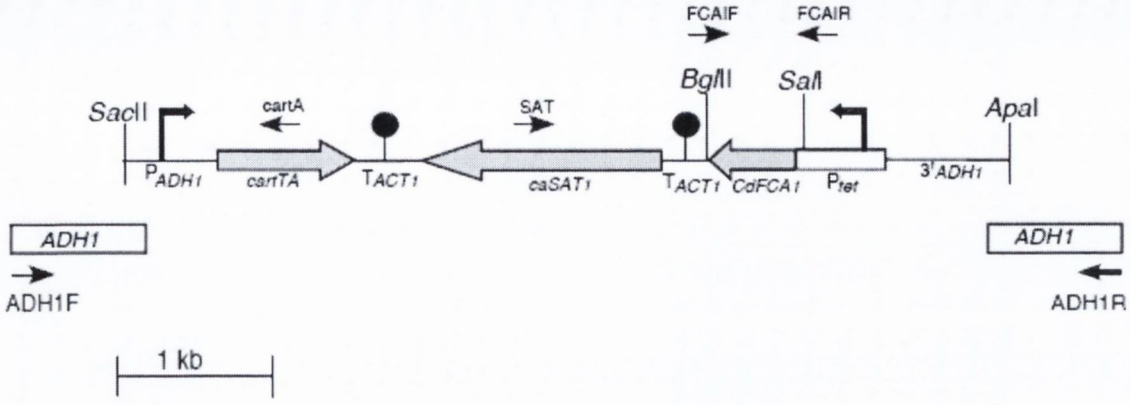


Figure 6.1. Structure and *ADH1* integration site of the pNIM1-*CdFCA1* cassette used in the tetracycline inducible expression transformation studies.

Restriction sites used in the excision of the *caGFP1* gene (replaced by the *CdFCA1* gene) and the excision of the entire pNIM1-*CdFCA1* cassette from the pNIM1 plasmid (Park & Morschhauser, 2005) are indicated. Transcription start sites and the directions of transcription are displayed using right-angled arrows. Terminator sequences are displayed as black hairpin loops which function in the termination of transcription of the *Candida*-adapted reverse tetracycline-dependent transactivator (cartTA) and the *CdFCA1* target gene. Primers used in the PCR screening of pNIM1-*CdFCA1* transformants are indicated by labelled arrows, and are listed in Table 6.1.

concentration as instructed by the manufacturer (Promega). Restriction endonuclease digestion was carried out at 37°C for 5 h. Horizontal 0.8% (w/v) agarose gels made up in 0.5 × TBE buffer were cast into horizontal gel trays. DNA loading dye was added to the restriction enzyme-digested DNA samples at a final concentration of 1 × and the samples were loaded into the gel wells. A DIG-labelled size standard (DNA molecular weight marker III; Roche) was loaded on each gel. Electrophoresis was carried out overnight at 20 V with a constant current.

6.2.8.3. Southern transfer of DNA from agarose gels

Following the separation of restriction endonuclease-digested DNA fragments by agarose gel electrophoresis; DNA was depurinated by soaking the gels in 250 mM HCl for no more than 10 mins, with gentle shaking. Following depurination, the DNA was denatured by soaking the gels in denaturation solution (500 mM NaOH, 1.5 M NaCl) for 15 min at room temperature. The denaturation step was carried out a second time, after which the gels were soaked in a neutralisation solution (500 mM Tris-HCl, 1.5 NaCl; pH 7.5) for 15 min with gentle shaking at room temperature. This neutralisation step was also carried out twice. DNA was transferred to positively charged nylon membranes (Boehringer, Mannheim, Germany) by capillary action using the Southern method (Southern, 1975). Following transfer, the membranes were rinsed in 2 × SSC, and DNA was crosslinked to the membrane using a crosslinker (CL-508, UVI tec, Cambridge, UK) set at 0.120 J/cm^2.

6.2.8.4. Hybridisation and detection of DIG-labelled probes

Hybridisation reactions were carried out in a rotary oven (Hybaid; Teddington, Middlesex, UK) in 25 × 3.5 cm bottles (Hybaid) by the method of Sambrook *et al.* (1989). Positively charged membranes were equlibrated in prehybridisation solution (5 × SSC, 0.1% (w/v) N-laurylsarcosine, 0.02% (w/v) SDS, 1 × blocking reagent [Roche]) that had been pre-warmed to 68°C for 2 h.

The probes were denatured in 50 µl volumes containing 600 ng of DIG-labelled DNA by heating at 95°C for 10 min and immediately cooling on ice. The denatured probe added to a 25 ml volume of prewarmed prehybridisation solution. The resulting hybridisation solution was incubated with the membranes at 68°C for 18 h with gentle rotation.

Unbound probe was removed from the membranes following hybridisation by washing the membranes (twice) in 200 ml of low stringency buffer (2 × SSC, 0.1% SDS) at room temperature for 10 min with vigorous shaking. These washes were followed by two washes in 100 ml of high stringency buffer (0.5 × SSC, 0.1% (w/v) SDS) which had been prewarmed to 68°C. Each high stringency wash was carried out at 68°C for 15 min with rotation. The membranes were then washed in 100 ml wash buffer (100 mM maleic acid, 150 mM NaCl, 0.4 M NaOH, 0.3% (v/v) Tween 20) at room temperature for 5 min with gentle shaking. After washing, the membranes were incubated in 100 ml 2 × blocking solution (0.8 × maleic acid buffer, 0.2 × blocking reagent) at room temperature for 2 h with gentle shaking. This was replaced with anti-DIG antibody solution (2 × blocking solution, 1/10,000 anti-DIG antibody-AP, Fab fragments [Roche]) which recognised the DNA-bound DIG-labelled probe. The membranes were incubated with antibody solution for 30 min at room temperature with gentle shaking, and were then rinsed in wash buffer twice for at least 15 min on each occasion.

Following washes, the membranes were equilibrated in detection buffer (100 mM Tris-HCl, 100 mM NaCl; pH 9.5) for 5 min and were transferred to Ziploc® Sandwich bags (Moon street, Bristol, UK) while still wet. A chemiluminesent substrate was added to the membranes, in order to detect the bound anti-DIG antibodies and emit visible light. A 500 µl volume of this substrate (0.25 M CPD-*Star,* ready-to-use [Roche]) was diluted in 500 µl of detection buffer and added to each membrane in the sandwich bags lowering the top sheet of the bag onto the membrane to exclude all air bubbles. Membranes were incubated at room temperature for 5 min and excess solution was removed from the bags, which were then placed in autoradiography cassettes with Kodak BioMax intensifying screens (Eastman Kodak Company, Rochester, New York, USA) and exposed to Kodak BioMax MS-1 X-ray films for approximately 1 h at room temperature. Autoradiograms were developed using Kodak GBX developer, and fixed in Kodak GBX fixer according to the protocols supplied by the manufacturer.

6.2.9. *CdFCA1* expression analysis

To monitor the relative gene expression of the *CdFCA1* gene in 5FC-resistant (SA113, SA109, Eg202) and 5FC-susceptible (Eg204, p7276) wild type *C. dubliniensis* isolates as well as in the DOX-inducible transformant derivatives, (SA113T1, SA109T1

and SA109T2), relative quantitative real-time PCR was carried out according to standard protocols.

6.2.9.1. RNA extraction and DNase treatment

In brief, RNA was extracted from isolates and transformant derivatives that were grown in YPD broth in the presence of DOX (15 μg/ml) only, or YPD broth in the presence of DOX (15 μg/ml), and sub-inhibitory concentrations of 5FC (6.4 ng/ml). The RNAs were extracted using the Qiagen RNeasy mini kit (Qiagen) which incorporates guanidine-isothiocyanate lysis with silica-membrane purification. Resulting RNA samples were treated with Turbo DNA-*free*™ (Ambion) according to the manufacturers' instructions in order to remove contaminating double stranded DNAs. The RNA samples were DNase-treated in 20 μl volumes containing 2 μl of 10 × Turbo DNase buffer and 1 μl of Turbo DNase. These samples were incubated at 37°C for 30 min, after which 2 μl of the resuspended DNase inactivation reagent was added to each sample and mixed. Samples were incubated at room temperature for 5 min with occasional mixing and were centrifuged at 10,000 × *g* for 2 min. Purified RNA was transferred to fresh tubes and stored at -20°C prior to reverse transcription to cDNA.

6.2.9.2. Reverse transcription of RNA to cDNA

The DNAase-treated RNA samples were then reverse transcribed to cDNA using the Superscript II reverse transcriptase kit (Invitrogen) according to the instructions of the manufacturer. Reverse transcription reactions were carried out in 11 μl volumes containing 1 μg total RNA, and 1 μl Oligo(dT)$_{12\text{-}18}$ (500 μg/ml) (Promega). Samples were heated to 70°C for 10 min and immediately cooled on ice for 1 min, after which 4 μl of 5 × First-Strand Buffer, 2 μl of 0.1 M DTT, 1 μl of RNaseOUT™ (all supplied by Invitrogen) and 1 μl of dNTP mix (10 mM each) (Promega) were added to each reaction. Samples were incubated at 42°C for 2 min and were held at this temperature whilst 1 μl (200 units) of SuperScript™ II RT was added to each. Reverse transcription was carried out at 42°C for 1 h, after which the reaction was stopped by heating to 70°C for 15 min. These cDNAs were used as the template in quantitative real-time PCR amplification.

6.2.9.3. Real-time PCR amplification efficiencies

Relative quantitative real-time PCR was carried out using two pairs of RT-PCR primers; one pair amplified the target *CdFCA1* gene, and the second pair amplified the

ACT1 gene as an internal expression control (Table 6.1). The comparative amplification efficiencies of these primers were assessed at 0.3 μM concentrations prior to RT-PCR using primer amplification efficiency plot analysis as previously described (Schmittgen & Livak, 2008). In brief, a series of two-fold cDNA dilutions were amplified in separate reactions using primers to amplify the *CdFCA1* target gene or the *ACT1* internal reference gene. Each amplification reaction was carried out in triplicate and the average C_T values for each primer pair were plotted against the concentration of input template cDNA. The real-time PCR efficiencies were calculated according to the equation $E=10^{(-1/slope)}$, where *E* is the efficiency of the primer pair (Pfaffl, 2001; Schmittgen & Livak, 2008).

6.2.9.4. Relative real-time PCR analysis of CdFCA1

Relative quantitative real-time PCRs were carried out in 15 μl reaction volumes containing 30 ng cDNA template, 0.3 μM of each primer, and 1 × Power SYBR® green mastermix (Applied Biosystems, Warrington, UK) in an ABI 7500 Real-Time PCR System (Applied Biosystems) according to the manufacturers recommended protocols. Real-time PCR cycling conditions consisted of an initial denaturation step of 95°C for 10 min, followed by 30 cycles of 95°C for 15 s, 59°C for 30 s and 72°C for 30 s. Data collection occurred during the annealing step at 59°C.

Data analysis was carried out using the $2^{(-\Delta CT)}$ as described by Livak and Schmittgen (2001). The average C_T was calculated for each set of replicates. The ΔC_T was determined as the (C_T *CdFCA1*- C_T *ACT1*) and fold expression changes were calculated as $2^{(-\Delta CT)}$ (ΔC_T Treated sample - ΔC_T Untreated sample). Statistical analysis was carried out in the form of two-tailed Student's *t*-tests, using Microsoft Office Excel version 2007.

6.3. Results

6.3.1. *In vitro* susceptibility testing

Nine *C. dubliniensis* isolates previously reported as 5FC-susceptible, and 12 *C. dubliniensis* isolates previously reported as 5FC-resistant by Al Mosaid *et al.* (2005) (Tables 5.1) were tested for susceptibility to 5FC using broth microdilution assays as described in Chapter 2, Section 2.7. All nine isolates previously reported as 5FC-susceptible were confirmed as such, exhibiting MIC values of ≤ 0.25 μg/ml. Similarly, the 12 previously reported 5FC-resistant isolates exhibited 5FC MIC values of ≥ 128 μg/ml. In an attempt to localise potential blocks or lesions in the 5FC metabolic pathway in 5FC-resistant isolates which may contribute to 5FC resistance, broth microdilution assays were also carried out using 5FU instead of 5FC (Table 5.1) as described in Chapter 2, Section 2.7. We hypothesised that if a block in the 5FC metabolic pathway occurred at the level of Fca1p, then 5FC-resistant isolates should be susceptible to 5FU (see Fig. 5.1). All nine 5FC-susceptible and 12 5FC-resistant isolates used in the study (Table 5.1) were tested for susceptibility to 5FU using a range of concentrations (0.25 - 128 μg/ml). All 21 isolates exhibited 5FU MIC values in the range of 8 to 32 μg/ml (Table 5.1). There was no correlation between 5FU MIC and susceptibility to 5FC. These findings indicated that a block(s) in the 5FC metabolic pathway occurred in resistant isolates at the level of cytosine deaminase or upstream of this enzyme (see Fig. 5.1).

6.3.2. Sequence analysis of *CdFCA1* in *C. dubliniensis*

The DNA sequence for the *C. albicans FCA1* gene, encoding Fca1p, (GenBank accession No. U55194) was used in a BLAST search against the *C. dubliniensis* genome sequence (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/) in order to identify its homologue in *C. dubliniensis*. In *C. dubliniensis*, Fca1p is encoded by the *CdFCA1* gene, which shares 89% nucleotide sequence identity with *FCA1* from *C. albicans,* and both contain an internal intron of 81 bp. In order to investigate whether a mutation(s) or deletion(s) was present in the *CdFCA1* gene encoding cytosine deaminase, the DNA sequences of the *CdFCA1* ORF was determined for 12 5FC-resistant isolates (Table 5.1), and nine 5FC-susceptible isolates (Table 5.1). For each isolate, the *CdFCA1* sequence was compared with that of the *CdFCA1* consensus

sequence of the 5FC-susceptible *C. dubliniensis* type strain CD36 (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/) (Fig. 6.2). Three SNPs were identified in the *CdFCA1* coding sequences of the 21 *C. dubliniensis* isolates investigated. Two synonymous SNPs were observed; one of these (position 264; A→T transversion) occurred in all isolates sequenced with the exception of 6/9 5FC-susceptible isolates (CD36, Eg203, Eg206, SA101, SA105 and SA115), and the second (position 390; T→C transition) occurred in one of the 5FC-susceptible isolates (p7276) only (Fig. 6.2). The third SNP was nonsynonymous (position 86; C→T transition); resulting in an amino acid substitution (Ser29Leu) in all 12 of the 5FC-resistant isolates tested, but was not present in the *CdFCA1* gene of the nine 5FC-susceptible isolates sequenced (Figs. 6.2 and 6.3). This radical substitution occurs in the β1 strand of the cytosine deaminase enzyme and is closely linked to an active site residue (Fig. 6.4), according to the yeast cytosine deaminase structure defined by Ko *et al.* (2003).

6.3.3. Tetracycline-inducible expression of *CdFCA1* in *C. dubliniensis*

6.3.3.1. Transformation of the tetracycline-inducible cassette in C. dubliniensis

In order to obtain direct evidence that the Ser29Leu substitution present in the Fca1p from 5FC-resistant *C. dubliniensis* isolates was responsible for the 5FC resistant phenotype in these isolates, the gene encoding cytosine deaminase from the 5FC-susceptible *C. dubliniensis* isolate p7276 (*CdFCA1*^{S}), which was originally recovered in Israel and which lacked the Ser29Leu substitution, was introduced into the *ADH1* locus of the two separate 5FC-resistant Saudi Arabian isolates SA109 (ITS genotype 3) and SA113 (ITS genotype 4) using the tetracycline-inducible cassette, pNIM1 (Park & Morschhäuser, 2005). Transformation of pNIM1-*CdFCA1*^{S} DNA into the 5FC-resistant isolates SA113 and SA109 yielded several transformants (hereafter referred to as pNIM1-*CdFCA1*^{S} transformants) with the correct integration of the pNIM1-*CdFCA1*^{S} cassette into the *ADH1* locus as determined by PCR analysis (Table 6.1 and Fig. 6.1). Transformation of the pNIM1-*CdFCA1*^{R} cassette into the 5FC-susceptible isolate p7276 yielded several transformants (hereafter referred to as pNIM1-*CdFCA1*^{R} transformants) with the correct integration of the pNIM1-*CdFCA1*^{R} cassette into the *ADH1* locus as determined by PCR analysis (Table 6.1 and Fig. 6.1). As transformation controls, the 5FC-resistant SA109 and SA113 parental isolates and the 5FC-susceptible parental

Figure 6.2. Alignment of *CdFCA1* coding DNA sequences from 5FC-resistant and 5FC-susceptible isolates.

[a] Sequences were compared to the consensus *CdFCA1* gene sequence obtained by a BLAST search of the *C. albicans FCA1* gene in the *C. dubliniensis* genome sequence database (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/).

Polymorphisms specific to 5FC-susceptible isolates only are indicated in blue typeface, SNPs specific to 5FC-resistant isolates only are indicated in red typeface, and SNPs detected in 5FC-susceptible and 5FC-resistant isolates are indicated in green typeface.

[b]The *CdFCA1* DNA sequences obtained for six of the 5FC-susceptible isolates (Table 5.1) were all identical and are indicated as one sequence (6 5FC-susceptible). Similarly, the *CdFCA1* DNA sequences for the 5FC-susceptible isolates p6785 and p7718 were identical and treated as one sequence (p6785/p7718).

[c] The *CdFCA1* DNA sequences obtained for the 12 5FC-resistant isolates (Table 5.1) were all identical and are indicated as one sequence (12 5FC-resistant).

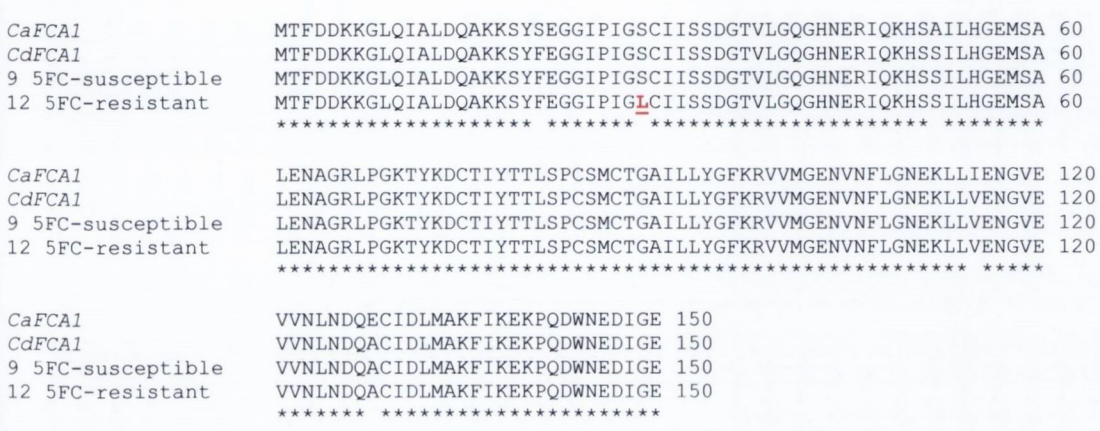


Figure 6.3. Alignment of *C. albicans* and *C. dubliniensis CdFCA1* amino acid sequences.

Sequences were compared to the consensus *CdFCA1* amino acid sequence of the type strain CD36, which was obtained by a BLAST search of the *C. albicans FCA1* gene in the *C. dubliniensis* genome sequence database (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/). The amino acid sequence of the *FCA1* gene in *C. albicans* (*CaFCA1*) was also included for comparison, and displayed 97 % identity with the reference *CdFCA1* sequence.

The *CdFCA1* amino acid sequences obtained for the nine 5FC-susceptible isolates (Table 5.1) are all identical and are indicated as one sequence (9 5FC-susceptible) as are the 12 5FC-resistant isolates (Table 5.1) which are also indicated as one sequence (12 5FC-resistant). Amino acid substitutions between 5FC-susceptible and 5FC-resistant *C. dubliniensis* isolates are indicated in red coloured, bold and underlined typeface.

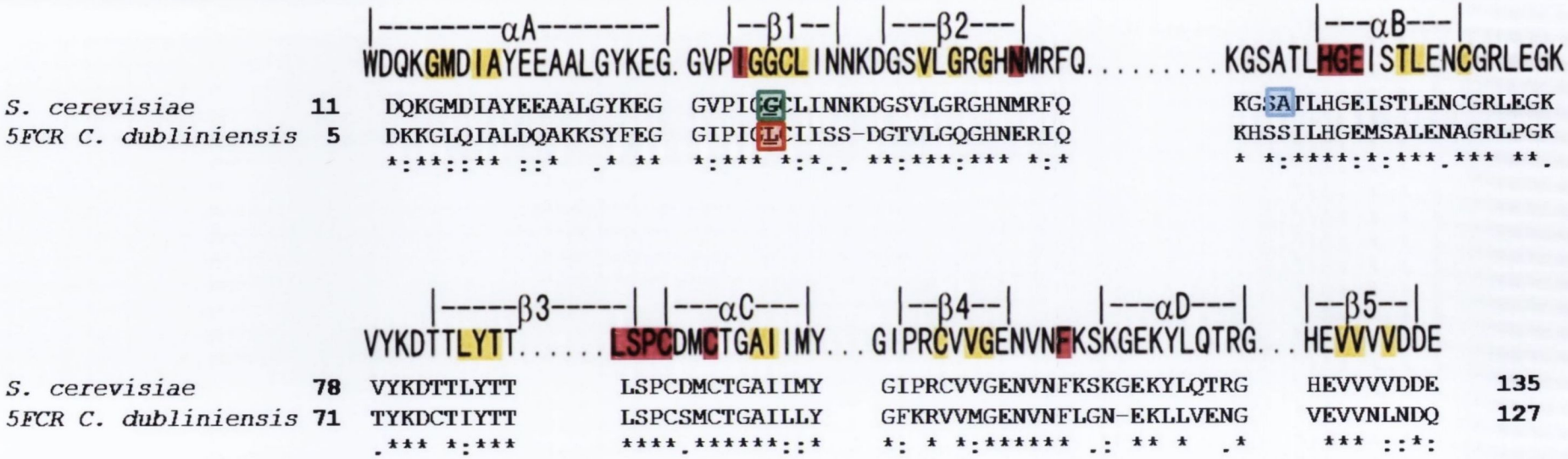


Figure 6.4. Structural domains of the cytosine deaminase of *S. cerevisiae*, and amino acid homology of the 5FC-resistant *C. dubliniensis* isolate SA113 (5FCR *C. dubliniensis*).

Hydrophobic core residues of the cytosine deaminase of *S. cerevisiae* are highlighted with a yellow background, and active site residues are highlighted with a red background (Ko *et al.*, 2003). A glycine [G] residue (highlighted in green) is present at position 34 of the cytosine deaminase of *S. cerevisiae.* This corresponds to the leucine [L] residue (highlighted in red) at position 29 of the 5FCR *C. dubliniensis* isolate, where the Ser29Leu substitution occurs in all 5FC-resistant *C. dubliniensis* isolates that have been examined to date. Interestingly, a hydrophobic alanine (A) residue (highlighted in blue) is present at position 58 in *S. cerevisiae*, this residue is closely linked to three active site residues and is conserved in *C. albicans* (see Fig. 6.3), but is replaced by a hydrophilic serine residue at this position in both 5FC-susceptible and 5FC-resistant isolates of *C. dubliniensis* (Fig. 6.3). The crystal structure of the cytosine deaminase has been determined (Ko *et al.*, 2003), and presuming that structural homology exists between *S. cerevisiae* and *C. dubliniensis*, as would be expected from the relative amino acid sequence conservation, the Ser29Leu transition occurs in a β strand in close proximity to an active site residue (isoleucine; position 27 in *C. dubliniensis*) that is conserved between the two species.

isolate p7276 were also transformed with the pNIM1 cassette containing the *GFP1* gene instead of the *CdFCA1* gene (Table 6.2).

6.3.3.2. Susceptibility testing of transformants

The 5FC-susceptibilties of the pNIM1-*CdFCA1*^{R} transformants were determined using broth microdilution assays using serial dilutions of 5FC from a concentration of 0.25 to 128 μg/ml. All of the transformants were 5FC-susceptible (MIC ≤ 0.25 μg/ml). This experiment was replicated with the addition of the tetracycline derivative, doxycycline (DOX) to the RPMI-1640 medium at a final concentration of 15 μg/ml, in order to induce expression of the pNIM1-*CdFCA1*^{R} cassette (Park & Morschhäuser, 2005). All of the DOX-induced transformants remained 5FC-susceptible (MIC ≤ 0.25 μg/ml).

The 5FC-susceptibilties of the pNIM1-*CdFCA1*^{S} transformants were determined using broth microdilution assays using serial dilutions of 5FC from a concentration of 0.25 to 128 μg/ml. All of these transformants were 5FC-resistant (MIC ≥ 128 μg/ml). In contrast, in a parallel series of experiments replicate broth microdilution assays were carried out with the addition of DOX (15 μg/ml) to the RPMI-1640 assay medium. This resulted in a dramatic change in the 5FC resistance phenotype of these transformants (SA109T1, SA109T2 and SA113T1) from being 5FC-resistant (MIC ≥ 128 μg/ml) to being 5FC-susceptible (MIC ≤ 0.25 μg/ml) (Table 6.2). Similar results were obtained in separate broth microdilution experiments using DOX at a final concentration of 30 μg/ml (Table 6.2).

The control pNIM1-*GFP1* transformants were also examined in broth microdilution assays in the presence and absence of DOX (15 μg/ml). The presence or absence of DOX in the broth microdilution medium had no effect on 5FC MIC values and all the transformant derivatives tested exhibited similar 5FC MIC values to their respective parental isolates (Table 6.2). These results strongly suggested that expression of the *CdFCA1*^{S} gene by the transformants harbouring the entire pNIM1-*CdFCA1*^{S} cassette (i.e., SA109T1, SA109T2 and SA113T1; Table 6.2) was responsible for the DOX-inducible 5FC susceptibility exhibited by these transformants.

6.3.3.3. Southern hybridisation

Definitive confirmation of the correct integration of the pNIM1-*CdFCA1*^{S} and pNIM1-*GFP1* cassettes was achieved by Southern hybridisation (Fig. 6.5).

Transformant derivatives containing the full and correct integration of the pNIM1-*CdFCA1* cassette resulted in a band of 4.6 kb. A DIG-labelled probe directed towards DNA sequence of the cartTA transactivator region of the pNIM1 cassette and the *ADH1* locus resulted in bands of 4.6 kb and 5.1 kb from the pNIM1-*CdFCA1* transformant derivatives SA113T1, SA109T1, and SA109T2, as well as from the pNIM1-*GFP1* control transformant derivatives SA113 pNIM1-GFP1T1, SA109 pNIM1-GFP1T1, SA109 pNIM1-GFP1T2 and p7276 pNIM1-GFP1T1 (Fig. 6.5A).

A second DIG-labelled probe was directed against the *CdFCA1* gene in a separate hybridisation experiment. Bands of 4.6 kb and 5.1 kb resulted from pNIM1-*CdFCA1* transformant derivatives SA113T1, SA109T1, and SA109T2 (Fig. 6.5B). These bands were absent from the pNIM1-*GFP1* control transformant derivatives and the parental isolates. The equal size (4.6 kb) of bands produced with both probes suggests that the ADH1/cartTA transactivator and *CdFCA1* sequences are both present on the same *Hind*III digested DNA fragment, confirming the correct integration of the pNIM1-*CdFCA1* and pNIM1-*GFP1* cassettes into the transformant derivatives SA113T1, SA109T1, and SA109T2 (Fig. 6.5A and B), and SA113 pNIM1-GFP1T1, SA109 pNIM1-GFP1T1, SA109 pNIM1-GFP1T2 and p7276 pNIM1-GFP1T1, respectively (Fig. 6.5A).

6.3.4. *CdFCA1* expression analysis

6.3.4.1. Comparative amplification efficiencies.

Real-time PCR primer amplification efficiencies were compared by plotting average C_T values obtained from replicate amplifications against the concentration of input template cDNA, for each primer pair at 0.3 μM concentrations (Fig. 6.6). Linear trendlines were added to each data series, and the slopes of these trendlines were obtained (Fig. 6.6). The slope of the line obtained by RTACT1 (Table 6.1) primer amplification was 0.9639, and the slope of the line obtained by RTFCA1 (Table 6.1) primer amplification was 0.895. The real-time PCR efficiencies were calculated according to the equation $E=10^{(-1/slope)}$, where E is the efficiency of the primer pair (Pfaffl, 2001; Schmittgen & Livak, 2008). The RTACT1 primer amplification efficiency was calculated as 2.6, and the RTFCA1 primer amplification efficiency was calculated as 2.5. These primer efficiencies are within 10% of each other and therefore these primers were used for relative real-time PCR analysis.

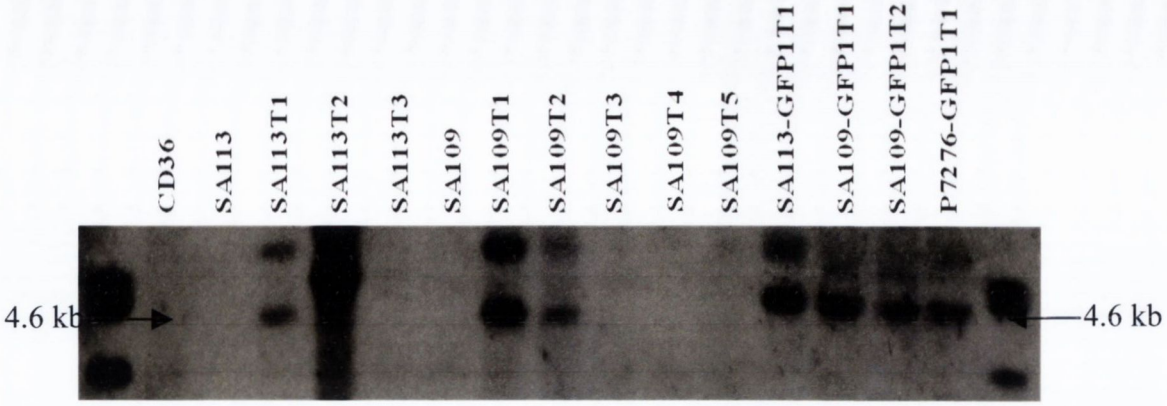
Table 6.2. Susceptibility of *C. dubliniensis* isolates and pNIM1-*FCA1/GFP1* transformant derivatives in the presence or absence of DOX

Isolate or transformant	5FC MIC_{50} (μg/ml)		
	No DOX	DOX (15 μg/ml)	DOX (30 μg/ml)
C. dubliniensis clinical isolates			
SA113	≥ 128	≥ 128	≥ 128
SA109	≥ 128	≥ 128	≥ 128
p7276	≤ 0.25	≤ 0.25	≤ 0.25
C. dubliniensis pNIM1-*CdFCA1*^{R} transformants			
p7276T1	≤ 0.25	≤ 0.25	≤ 0.25
p7276T2	≤ 0.25	≤ 0.25	≤ 0.25
p7276T3	≤ 0.25	≤ 0.25	≤ 0.25
C. dubliniensis pNIM1-*CdFCA1*^{S} transformants			
SA109T1	≥ 128	≤ 0.25	≤ 0.25
SA109T2	≥ 128	≤ 0.25	≤ 0.25
SA113T1	≥ 128	≤ 0.25	≤ 0.25
C. dubliniensis pNIM1-*GFP1* transformants			
SA113–GFP1T1	≥ 128	≥ 128	≥ 128
SA109-GFP1T1	≥ 128	≥ 128	≥ 128
SA109-GFP1T2	≥ 128	≥ 128	≥ 128
p7276-GFP1T1	≤ 0.25	≤ 0.25	≤ 0.25

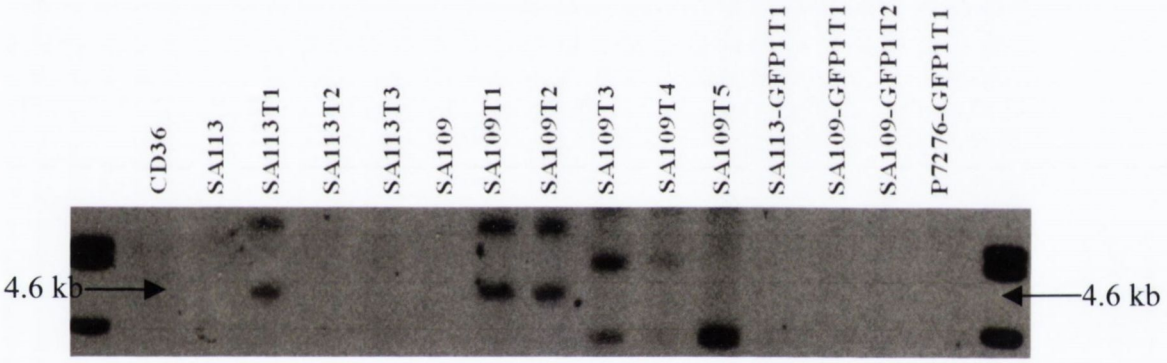
Figure 6.5. Southern hybridisation experiments using *Hind*III digested genomic DNA and two separate DIG-labelled probes to confirm pNIM1-integration into the *ADH1* locus of *C. dubliniensis*.

Separate hybridisations of the individual probes to the DNA of transformant derivatives which contained the full and correct integration of the pNIM1-*CdFCA1* cassette resulted in a band of 4.6 kb. The *C. dubliniensis* type strain CD36 and the two parental isolates (SA113 and SA109) were also included in both analyses, as were the pNIM1-*GFP1* control transformants. The DIG-labelled probe used in panel A was directed towards DNA sequence of the cartTA transactivator region of the pNIM1 cassette and the *ADH1* locus. A band of 4.6 kb was formed by hybridisation of this probe to DNA from pNIM1-*CdFCA1* transformant derivatives SA113T1, SA109T1, and SA109T2, as well as from the pNIM1-*GFP1* control transformant derivatives SA113 pNIM1-GFP1T1, SA109 pNIM1-GFP1T1, SA109 pNIM1-GFP1T2 and p7276 pNIM1-GFP1T1 (panel A). The DIG-labelled probe used in panel B was directed against the *CdFCA1* gene. A band of 4.6 kb was formed by hybridisation of this probe to DNA from pNIM1-*CdFCA1* transformant derivatives SA113T1, SA109T1, and SA109T2 (panel B). The 4.6 kb bands were absent from the pNIM1-*GFP1* control transformant derivatives due to the presence of the *GFP1* ORF instead of the *CdFCA1* ORF. A second band of 5.1 kb was formed by the pNIM1-*CdFCA1* transformant derivatives SA113T1, SA109T1, and SA109T2 and the pNIM1-GFP1 transformant derivatives SA113 pNIM1-GFP1T1, SA109 pNIM1-GFP1T1, SA109 pNIM1-GFP1T2 and p7276 pNIM1-GFP1T1 on hybridising to the ADH1/cartTA-directed probe (panel A). A band of similar size was also formed by the pNIM1-*CdFCA1* transformant derivatives SA113T1, SA109T1, and SA109T2 on binding the *CdFCA1*-directed probe. This band was absent in the pNIM1-GFP1 derivatives and the parental isolates (panel B). The equal size (4.6 kb) of bands produced with both probes suggests that the cartTA transactivator and *CdFCA1* sequences are both present on the same *Hind*III digested DNA fragment, confirming the correct integration of the pNIM1-*CdFCA1* cassette into the transformant derivatives SA113T1, SA109T1, and SA109T2.

A


CD36
SA113
SA113T1
SA113T2
SA113T3
SA109
SA109T1
SA109T2
SA109T3
SA109T4
SA109T5
SA113-GFP1T1
SA109-GFP1T1
SA109-GFP1T2
P7276-GFP1T1
4.6 kb
4.6 kb

B


CD36
SA113
SA113T1
SA113T2
SA113T3
SA109
SA109T1
SA109T2
SA109T3
SA109T4
SA109T5
SA113-GFP1T1
SA109-GFP1T1
SA109-GFP1T2
P7276-GFP1T1
4.6 kb
4.6 kb

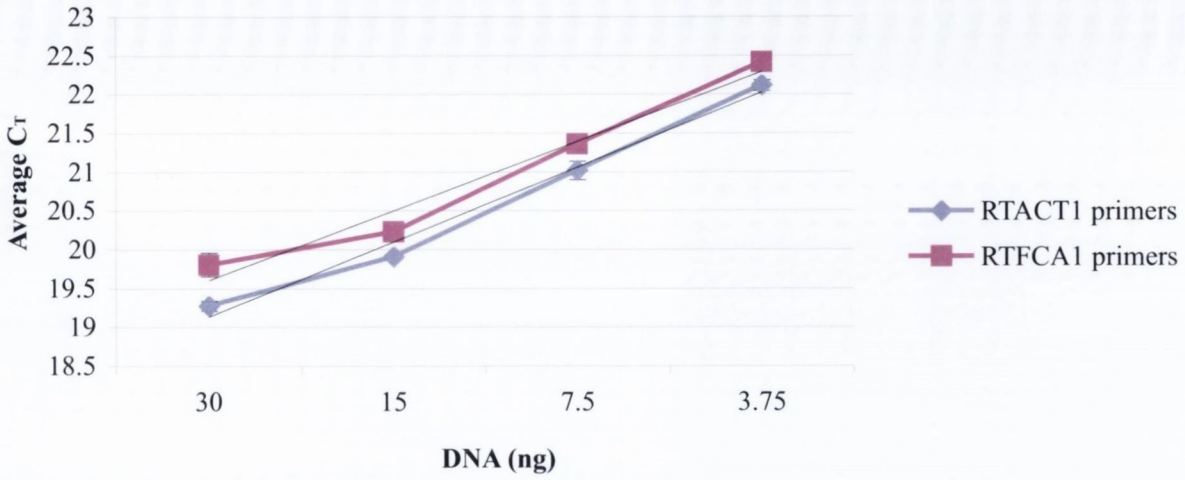


Figure 6.6. Amplification efficiencies of real-time PCR primers directed towards the *ACT1* internal reference gene and the *CdFCA1* target gene in *C. dubliniensis*. Average C_T values were obtained from three replicates were plotted against a series of DNA dilutions for each primer pair. Amplification efficiencies were analysed at a concentration of 0.3 μM for each of the forward and reverse primers. Standard deviations are denoted using error bars. Linear trendlines were added to each of the data series, and the equations of each of these trendlines were obtained. The equation of the RTACT1 trendline is y = 0.9639x+18.713, and the equation of the RTFCA1 trendline is y = 0.8952x+18.713. Amplification efficiencies were calculated according to the slopes of the trendlines according to the equation $E=10^{(-1/\text{slope})}$, where E is the efficiency of the primer pair (Schmittgen & Livak, 2008).

6.3.4.2. Relative quantitative real-time PCR analysis

A comparison of *CdFCA1* gene expression by the 5FC-susceptible (Eg204 and p7276) and 5FC-resistant (SA113, SA109 and Eg202) isolates and pNIM1-*CdFCA1*^{S} transformant derivatives SA113T1, SA109T1 and SA109T2 was undertaken following exposure to DOX (15 μg/ml), or following exposure to DOX (15 μg/ml) and a sub-inhibitory concentration (6.4 ng/ml) of 5FC. The expression of *CdFCA1* was analysed by relative quantitative real-time PCR, normalizing data to *ACT1* expression as an internal control.

In the presence of DOX, the average C_T values for the *CdFCA1* gene were 20.6 (Standard Deviation; SD 1.0) in the 5FC-resistant isolates, 20.1 (SD 0.6) in the 5FC-susceptible isolates, and 15.5 (SD 0.2) in the pNIM1-*CdFCA1*^{S} transformant derivatives (Table 6.3). The average ΔC_T values were calculated by subtraction of average C_T values of the *ACT1* gene from those of the target *CdFCA1* gene. Average values obtained were 3.5 (SD 1.0) for the 5FC-resistant isolates, 4.3 (SD 0.4) for the 5FC-susceptible isolates, and –0.2 (SD 0.3) for the pNIM1-*CdFCA1*^{S} transformant derivatives (Table 6.3). Relative *CdFCA1* expression in the presence of DOX and the absence of 5FC was analysed by comparison of $2^{(-\Delta C_T)}$ values for each strain examined (Fig. 6.7), and a two-tailed Student's *t*-test found no significant difference ($p=0.2$) in *CdFCA1* expression between 5FC-resistant or 5FC-susceptible isolates. In the presence of DOX, the expression of *CdFCA1* increased at least 4 fold (range 4.5–25.8) in the pNIM1-*CdFCA1* transformant derivatives (SA113T1, SA109T1 and SA109T2) in comparison to the parental isolates SA113 and SA109. These expression increases were all highly significant ($p<0.001$) according to two-tailed Student's *t*-tests.

In the presence of DOX and sub-inhibitory concentrations of 5FC, the average C_T values for the *CdFCA1* gene were 19.45 (SD 0.5) in the 5FC-resistant isolates, 20.11 (SD 0.3) in the 5FC-susceptible isolates, and 16.3 (SD 0.7) in the pNIM1-*CdFCA1*^{S} transformant derivatives (Table 6.4). The average ΔC_T values obtained were 4.2 (SD 0.3) for the 5FC-resistant isolates, 4.3 (SD 0.3) for the 5FC-susceptible isolates, and 0.2 (SD 0.8) for the pNIM1-*CdFCA1*^{S} transformant derivatives (Table 6.4). Relative *CdFCA1* expression levels in the presence of DOX and sub-inhibitory concentrations of 5FC were also analysed (Fig. 6.8). A two-tailed Student's *t*-test found no significant difference ($p=0.5$) in *CdFCA1* expression between 5FC-resistant or 5FC-susceptible isolates in the presence of 5FC. The expression of *CdFCA1* increased at least 5.2 fold (range 5.2–22.3) in the pNIM1-*CdFCA1* transformant derivatives (SA113T1, SA109T1

and SA109T2) in comparison to the parental isolates SA113 and SA109 due to the presence of DOX in the growth medium. These increases were all highly significant ($p<0.001$) according to a two-tailed Student's *t*-test.

These results indicate that alterations in *CdFCA1* expression do not play a significant role in mediating 5FC resistance. Upon exposure to 5FC (6.4 ng/ml), *CdFCA1* expression increased significantly ($p<0.001$) in both the 5FC-susceptible isolates (range 14.5–25 fold), and in the 5FC-resistant isolates (range 4–18 fold).

Table 6.3. Real-time PCR data for 5FC-resistant and 5FC-susceptible isolates and pNIM1-*CdFCA1*^{S} transformant derivatives grown in the presence of DOX (15 µg/ml)

Strain[a]	5FC-phenotype[b]	Average C_T[c] for indicated gene		ΔC_T[d]	$(2^{-\Delta C_T})$[e]
		ACT1	*CdFCA1*		
SA109	5FCR	19.4	21.6	2.2	0.2
SA113	5FCR	15.6	19.8	4.2	0.1
Eg202	5FCR	16.2	20.2	4.0	0.1
p7276	5FCS	16.0	20.6	4.7	0.0
Eg204	5FCS	15.7	19.6	3.9	0.1
SA109 T1	DOX/5FCS	15.6	15.4	-0.2	1.1
SA109 T2	DOX/5FCS	15.6	15.6	0.0	0.1
SA113 T1	DOX/5FCS	15.8	15.3	-0.5	1.4

[a] 5FC-resistant isolates and 5FC-susceptible isolates and pNIM1-*CdFCA1*^{S} transformant derivatives analysed in the study.

[b] 5FC-phenotypes of clinical isolates and DOX-inducible transformants; 5FCR, 5FC-resistant; 5FCS, 5FC-susceptible; DOX/5FCS, 5FC-susceptible in the presence of DOX.

[c] C_T is defined as the quantitative end point, i.e., the PCR cycle at which the threshold of the reporter dye crosses an arbitrarily placed threshold. Average C_T values were obtained from three replicate samples for each strain analysed.

[d] ΔC_T values were calculated by subtraction of average C_T values of the internal reference *ACT1* gene from those of the target *CdFCA1* gene.

[e] $2^{-\Delta C_T}$ The fold change in *CdFCA1* gene expression amongst strains according to relative real-time PCR analyses.

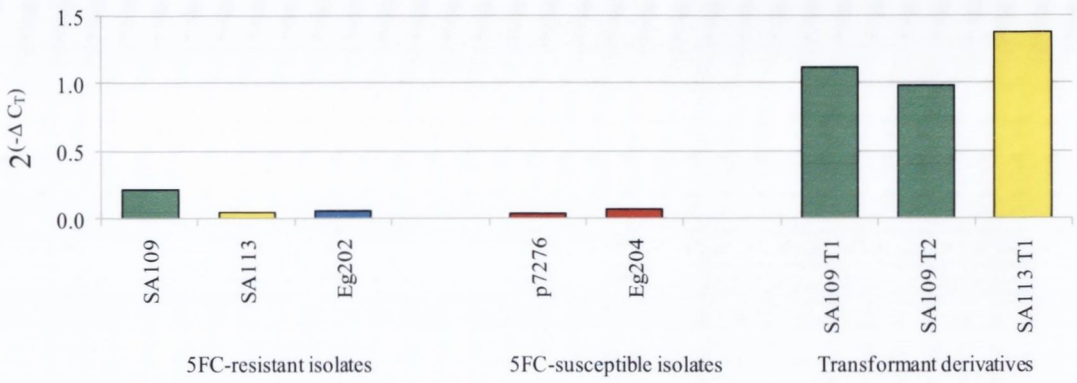


Figure 6.7. Relative expression of the *CdFCA1* gene in *C. dubliniensis* isolates and pNIM1-*CdFCA1*^{S} transformant derivatives grown in the presence of DOX.

Relative *CdFCA1* gene expression was analysed in 5FC-resistant and 5FC-susceptible isolates and pNIM1-*CdFCA1*^{S} transformant derivatives grown in the presence of DOX (15 μg/ml). Strains were analysed by comparison of $2^{(-\Delta C_T)}$ values in order to compare *CdFCA1* expression between 5FC-resistant isolates and 5FC-susceptible isolates, and to determine if *CdFCA1* gene expression was elevated in the pNIM1-*CdFCA1*^{S} transformant derivatives. Relative gene expression levels of the 5FC-resistant parental isolate SA109 and its resulting pNIM1-*CdFCA1*^{S} transformant derivatives SA109T1 and SA109T2 are shown in green, the 5FC-resistant parental isolate SA113 and its resulting pNIM1-*CdFCA1* transformant derivative SA113T1 are shown in yellow. Relative gene expression levels of an additional 5FC-resistant isolate (Eg202) is shown in blue, and are displayed in red for the two 5FC-susceptible isolates (p7276 and Eg204) also analysed in the study. In the presence of DOX, the expression of the *CdFCA1* gene increased at least 4 fold (range 4.5–25.8) in the pNIM1-*CdFCA1* transformant derivatives (SA113T1, SA109T1 and SA109T2) in comparison to the parental isolates SA113 and SA109.

Table 6.4. Real-time PCR data for 5FC-resistant and 5FC-susceptible isolates and pNIM1-*CdFCA1*^{S} transformant derivatives grown in the presence of DOX (15µg/ml) and sub-inhibitory concentrations of 5FC

Strain[a]	5FC-phenotype[b]	Average C_T[c] for indicated gene		ΔC_T[d]	$(2^{-\Delta C_T})$[e]
		ACT1	*CdFCA1*		
SA109	5FCR	15.7	19.3	3.6	0.1
SA113	5FCR	15.8	20.1	4.1	0.1
Eg202	5FCR	15.5	19.0	3.5	0.1
p7276	5FCS	15.8	20.3	4.5	0.0
Eg204	5FCS	15.9	19.9	4.0	0.1
SA109 T1	DOX/5FCS	17.1	16.8	-0.3	1.3
SA109 T2	DOX/5FCS	15.5	16.7	1.2	0.4
SA113 T1	DOX/5FCS	15.8	15.4	-0.4	1.3

[a] 5FC-resistant isolates and 5FC-susceptible isolates and pNIM1-*CdFCA1*^{S} transformant derivatives analysed in the study.

[b] 5FC-phenotypes of clinical isolates and DOX-inducible transformants; 5FCR, 5FC-resistant; 5FCS, 5FC-susceptible; DOX/5FCS, 5FC-susceptible in the presence of DOX.

[c] C_T is defined as the quantitative end point, i.e., the PCR cycle at which the threshold of the reporter dye crosses an arbitrarily placed threshold. Average C_T values were obtained from three replicate samples for each strain analysed.

[d] ΔC_T values were calculated by subtraction of average C_T values of the internal reference *ACT1* gene from those of the target *CdFCA1* gene.

[e] $(2^{-\Delta C_T})$ The fold change in *CdFCA1* expression amongst strains according to relative real-time PCR analyses.

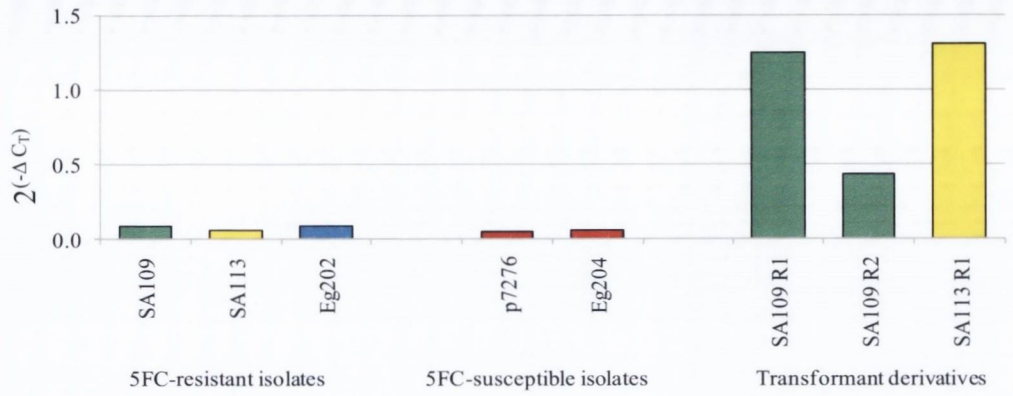


Figure 6.8. Relative expression of the *CdFCA1* gene in 5FC-resistant and 5FC-susceptible isolates and pNIM1-*CdFCA1*^{S} transformant derivatives grown in the presence of DOX and sub-inhibitory concentrations of 5FC.

Relative *CdFCA1* gene expression was analysed in 5FC-resistant and 5FC-susceptible isolates and pNIM1-*CdFCA1*^{S} transformant derivatives grown in the presence of DOX (15 µg/ml) and a sub-inhibitory concentration of 5FC (6.4 ng/ml). Strains were analysed by comparison of $2^{(-\Delta C_T)}$ values in order to compare *CdFCA1* expression between 5FC-resistant isolates and 5FC-susceptible isolates in the presence of 5FC in addition to DOX, and to determine if *CdFCA1* gene expression was further elevated in the pNIM1-*CdFCA1*^{S} transformant derivatives in the presence of 5FC in addition to DOX. Relative gene expression levels of the 5FC-resistant parental isolate SA109 and its resulting pNIM1-*CdFCA1* transformant derivatives SA109T1 and SA109T2 are shown in green, the 5FC-resistant parental isolate SA113 and its resulting pNIM1-*CdFCA1* transformant derivative SA113T1 are shown in yellow. Relative expression levels of an additional 5FC-resistant isolate (Eg202) is shown in blue, and are displayed in red for the two 5FC-susceptible isolates (p7276 and Eg204) also analysed in the study. In the presence of DOX and 5FC, the expression of the *CdFCA1* gene increased at least 5.2 fold (range 5.2–22.3) in the pNIM1-*CdFCA1* transformant derivatives (SA113T1, SA109T1 and SA109T2) in comparison to the parental isolates SA113 and SA109.

6.4. Discussion

Studies of pyrimidine salvage pathways (Fig. 5.1) and 5FC resistance mechanisms in yeast species have previously been undertaken with *S. cerevisiae* (Weber *et al.*, 1990; Erbs *et al.*, 1997; Kurtz *et al.*, 1999; Paluszynski *et al.*, 2006), *C. albicans* (Whelan & Kerridge, 1984; Erbs *et al.*, 1997; Hope *et al.*, 2004) and *C. lusitaniae* (Noël *et al.*, 2003; Chapeland-Leclerc *et al.*, 2005; Papon *et al.*, 2007). Investigations in *S. cerevisiae* have shown that disruption of the *FCY2* or *FUR1* gene can play a role in 5FC resistance, but only the *FCA1* gene is absolutely required for mediation of 5FC susceptibility (Paluzynski *et al.*, 2006, 2008). In *C. lusitaniae*, Papon *et al.* (2006) reported that inactivation of either the *FCA1* or the *FCY2* gene mediates 5FC resistance (MICs 128 μg/ml and 64 μg/ml, respectively), and promotes 5FC/fluconazole cross resistance (5FC MIC 4–32 μg/ml at a constant fluconazole concentration of 16 μg/ml) (Noël *et al.*, 2003; Chapeland-Leclerc *et al.*, 2005; Papon *et al.*, 2007). Further analysis has identified a nonsense mutation in the *FCY2* gene of seven isolates, resulting in a truncated purine-cytosine permease. In addition to this finding, a missense mutation (Met9Thr substitution) has been identified in the *FCY1* gene of four clinical *C. lusitaniae* isolates also exhibiting 5FC and 5FC/fluconazole cross resistance (Florent *et al.*, 2009). In *C. albicans*, two different research groups (Dodgson *et al.*, 2004; Hope *et al.*, 2004) identified a homozygous Arg101Cys amino acid substitution in the *FUR1*-encoded UPRT as the most common cause of high level 5FC resistance (MICs 8–64 μg/ml). Isolates that were heterozygous for this substitution exhibited reduced 5FC susceptibility (MICs 0.5–1 μg/ml) (Hope *et al.*, 2004). Furthermore, a homozygous Gly28Asp substitution in the Fca1p was suggested by Hope *et al.* as an alternative method of resistance in a single *C. albicans* isolate that did not harbour the UPRT-associated Arg101Cys substitution (Hope *et al.*, 2004). Finally, Hope *et al.* (2004) also described a single *C. albicans* isolate with a Ser29Leu substitution in the Fca1p, which exhibited an intermediate level of 5FC resistance (MIC_{50} 4 μg/ml). In the light of these previous studies we hypothesised that the *C. dubliniensis* pyrimidine salvage pathway very likely retains structural and functional homology with pyrimidine salvage pathways in other *Candida* species, as *C. dubliniensis* is the closest relative to *C. albicans* in the genus *Candida*. Therefore, we investigated the *CdFCA1*-encoded Fca1p (Fig. 5.1) as a possible cause of *C.*

dubliniensis clade-specific 5FC resistance. Initially broth microdilution assays were carried out using both 5FC and 5FU to determine whether the deamination step in the 5FC metabolic pathway was responsible for 5FC resistance. If the deamination step was responsible for 5FC resistance, bypassing its requirement in the metabolic pathway by exposing 5FC-resistant cells to 5FU should result in similar 5FC MICs in both 5FC-resistant and 5FC-susceptible isolates (see Fig. 5.1). This was shown to be the case, (Table 5.1) indicating that the Fca1p was very likely responsible for clade-specific 5FC resistance in *C. dubliniensis*.

A previous study (Al Mosaid *et al.*, 2005) examined the DNA sequence of the *CdFCA1* gene in 4 5FC-resistant isolates and 11 5FC-susceptible isolates in order to detect any SNPs which could account for 5FC resistance. The study identified a single SNP (position 258; A→ T transversion) that occurred in all four 5FC-resistant isolates examined exclusively, but this SNP did not result in an amino acid substitution in the Fca1p. In the present study, 12 5FC-resistant isolates and nine 5FC-susceptible isolates were examined using a different pair of primers (FCA1F/FCA1R) that amplified the entire *CdFCA1* ORF (Table 6.1). The previously reported SNP was not detected using these primers, and a further three SNPs were detected amongst the 21 isolates examined, one of which occurred exclusively in all 12 5FC-resistant isolates examined, and resulted in a Ser29Leu substitution. It is possible that the primers used for the previous study did not bind as efficiently as the primer pair used in the current study during amplification and DNA sequencing, which may have affected the resulting DNA sequence data.

The radical Ser29Leu substitution results in the replacement of a hydrophilic polar amino acid (Serine) with a hydrophobic non-polar residue (Leucine) in the β1 strand of the Fca1p enzyme and is closely linked to an active site residue, according to the yeast cytosine deaminase structure defined by Ko *et al.* (2003). This amino acid substitution may disrupt the quaternary structure of the enzyme, distorting the active site and inhibiting the conversion of the 5FC prodrug to its toxic form, 5FU. As mentioned above, a similar amino acid substitution was reported by Hope *et al.* (2004) in a single *C. albicans* isolate, however this isolate exhibited intermediate resistance (MIC_{50} 4 μg/ml) to 5FC (Hope *et al.*, 2004), in comparison to the high levels of 5FC resistance observed in the *C. dubliniensis* isolates displaying the Ser29Leu substitution in the present study. The difference in the levels of resistance to 5FC exhibited by the *C. albicans* isolate reported by Hope *et al.* (2004) and the 5FC-resistant *C. dubliniensis*

isolates reported here, all harbouring the same Ser29Leu substitution in Fca1p, may be due to the fact that different 5FC MIC determination methods were used in both studies; the EUCAST (2008) method was used for the *C. albicans* isolates and the CLSI method (2002) was used here for *C. dubliniensis* isolates. Alternatively, differences in *CdFCA1* and *FCA1* gene expression or post-transcriptional or post-translational modifications may be responsible for the differences in the levels of resistance to 5FC exhibited by isolates harboring the Fca1p Ser29Leu substitution between the two species.

The Fca1p encoded by *C. albicans* displays 60% amino acid identity with that of *S. cerevisiae* and 97% amino acid identity with that of 5FC-susceptible (MIC $\leq$ 0.25 μg/ml) *C. dubliniensis* type strain, CD36. A hydrophobic alanine residue is present at position 51 of the Fca1p of *C. albicans* that is conserved in the Fca1p of *S. cerevisiae*; this residue is closely linked to three active site residues according to the Fca1p structure (Fig. 6.4) defined by Ko *et al.* (2003). This hydrophobic alanine residue is replaced with a hydrophilic serine residue at the corresponding position of the Fca1p of both 5FC-resistant and 5FC-susceptible isolates of *C. dubliniensis* (Figs. 6.3 and 6.4). A structural alteration caused by this amino acid variation may alter the intrinsic Fca1p activity of *C. dubliniensis*, possibly explaining the different 5FC MIC values of the single *C. albicans* isolate (MIC_{50} 4 μg/ml) harbouring the Ser29Leu substitution reported by Hope *et al.* (2004) and the clade C3 *C. dubliniensis* isolates exhibiting high level 5FC resistance (MIC_{50} $\geq$ 128 μg/ml) harbouring the Ser29Leu substitution described in this study.

In order to obtain direct evidence that the Ser29Leu substitution present in the Fca1p from 5FC-resistant *C. dubliniensis* isolates was responsible for the 5FC-resistant phenotype in these isolates, the gene encoding cytosine deaminase from the 5FC-susceptible *C. dubliniensis* isolate p7276 (*CdFCA1*^{S}), which was originally recovered in Israel and which lacked the Ser29Leu substitution, was introduced into the *ADH1* locus of the two separate 5FC-resistant Saudi Arabian isolates SA109 (ITS genotype 3) and SA113 (ITS genotype 4) using the tetracycline-inducible cassette, pNIM1 (Park & Morschhäuser, 2005). Three transformant derivatives tested (SA109T1, SA109T2 and SA113T1) harbouring the complete pNIM1-*CdFCA1*^{S} cassette integrated into the *ADH1* locus exhibited DOX-inducible 5FC susceptibility on acquisition and expression of the *CdFCA1*^{S} gene (Table 6.2). Transformation of one of the *CdFCA1*^{R} alleles from the 5FC-resistant isolate SA113 into the *ADH1* locus of the 5FC-susceptible isolate p7276 was also carried out using a similar cassette. However, DOX-inducible expression of

the *CdFCA1*^{R} gene did not confer 5FC resistance to intrinsically 5FC-susceptible isolates, suggesting that the Ser29Leu substitution is recessive and does not affect the 5FC-phenotype in the presence of a wild-type allele. The present study suggests that the homozygous Ser29Leu substitution in the 5FC-resistant isolates renders the encoded Fca1p gene product non-functional, blocking the deamination of 5FC to the toxic metabolite 5FU.

No significant difference in *CdFCA1* expression was detected between the 5FC-susceptible and the 5FC-resistant *C. dubliniensis* isolates tested by quantitative real time PCR analysis in either the presence (Fig. 6.8) or absence (Fig. 6.7) of sub-inhibitory concentrations of 5FC in the growth medium. These results indicated that lack of, or reduced *CdFCA1* expression was not responsible for 5FC resistance in 5FC-resistant isolates following exposure to sub-inhibitory concentrations of 5FC, and that the Ser29Leu substitution is very likely the sole method of *CdFCA1*-mediated 5FC resistance in the *C. dubliniensis* isolates investigated. Previous studies have shown that free pyrimidines often present in peptones present in some brands of culture media can antagonise the activity of 5FC (Doern *et al.*, 1986). Antagonism was not observed in the present study with YPD-grown cultures used in expression studies. Following the addition of a sub-inhibitory concentration of 5FC (i.e., 6.4 ng/ml) to YPD-grown cultures, quantitative real-time PCR experiments consistently showed that both 5FC-susceptible and 5FC-resistant *C. dubliniensis* isolates exhibited significant upregulation of *CdFCA1* expression, 14.5–20 fold and 4–18 fold, respectively.

All Cd25 fingerprint Group III *C. dubliniensis* isolates tested so far exhibit high level 5FC resistance, all were originally recovered in Saudi Arabia, Egypt or Israel and all belong to MLST clade C3 (Al Mosaid *et al.*, 2005; Chapter 4, this thesis). The close genetic relationship shared by these isolates is reflected by identical *CdFCA1* DNA sequences and high level resistance to 5FC. It is highly likely that an identical mechanism is used to mediate 5FC resistance in all of these isolates. MLST C3 clade (Cd25 fingerprint Group III) *C. dubliniensis* isolates can be subdivided into ITS genotypes 3 and 4 based on the nucleotide sequence of the ITS region of the rDNA operon (Al Mosaid *et al.*, 2005), although clade C3 isolates of both ITS genotypes exhibit high level 5FC resistance. In the present study clade C3 *C. dubliniensis* isolates SA109 (ITS genotype 3) and SA113 (genotype 4) both were transformed with, the pNIM1-*CdFCA1*^{S} cassette and both yielded transformant derivatives (SA109T1, SA109T2, and SA113T1) that exhibited DOX-inducible 5FC susceptibility. These

findings support our view that clade-specific 5FC resistance in *C. dubliniensis* is mediated by a common molecular mechanism, i.e., the presence of the homozygous Ser29Leu substitution in the Fca1p.

This is not the first report of a clade-specific SNP that has resulted in the alteration of a protein involved in antifungal drug resistance in *C. dubliniensis*. In 2002 Moran *et al.* (2002) reported that 58% of ITS genotype 1 *C. dubliniensis* isolates (Cd25-group I, MLST clade C1) harboured a TAG nonsense mutation in the *CDR1* gene encoding an ABC transporter. In *C. albicans*, upregulation of *CDR1* is the most common mechanism of fluconazole resistance, whereas in *C. dubliniensis* the most common mechanism of fluconazole resistance involves overexpression of the *MDR1* gene encoding a multidrug transporter (Moran *et al.*, 1998). These studies highlight that despite the close phylogenetic relationship between *C. dubliniensis* and *C. albicans,* resistance to particular antifungal drugs can be due to different mechanisms in the two species.

In conclusion, the results of this study demonstrate that presence of a Ser29Leu substitution in the Fca1p of *C. dubliniensis* isolates is responsible for clade-specific resistance to 5FC. Isolates belonging to *C. dubliniensis* clade C3 have only been recovered in individuals of Arab ethnicity in Saudi Arabia, Egypt and Israel (Al Mosaid *et al.*, 2005). Resistance to 5FC has not yet been reported in *C. dubliniensis* isolates from other countries around the world apart from Kuwait (Pfaller *et al.*, 1999; Ahmad *et al.*, 2004; Al Mosaid *et al.*, 2005). In 2004 the recovery of two 5FC-resistant isolates of *C. dubliniensis* from Kuwait was reported (Ahmad *et al.*, 2004). Due to Kuwait's close proximity to Saudi Arabia, it is likely that these isolates also belong to *C. dubliniensis* MLST clade C3.

Chapter 7
General Discussion

7.1. General Discussion

The present study has offered some insights into the prevalence and population structure of *C. dubliniensis.* It has demonstrated that the true prevalence of *C. dubliniensis* can be underestimated when solely using culture-based techniques, and that direct PCR analysis of broth enriched clinical samples is the most reliable method to undertake such studies.

The analysis of the population structure of *C. dubliniensis* using the MLST method has offered a reliable alternative to the laborious DNA fingerprinting method using the *C. dubliniensis*-specific complex probe Cd25. It has also enabled direct comparisons of the population structure of *C. dubliniensis* with its significantly more pathogenic relative, *C. albicans*. Such a comparison has shown that the population structure of *C. dubliniensis* is significantly less divergent than *C. albicans*, which suggests that *C. dubliniensis* is less able to adapt to specific host environments than its closest relative, perhaps explaining, at least in part, the lower virulence of *C. dubliniensis* in comparison to *C. albicans*.

This study also identified for the first time a Ser29Leu substitution in the Fca1p of *C. dubliniensis* isolates that was responsible for high level 5FC resistance in isolates belonging to MLST clade C3 and Cd25 group III.

7.1.1. Prevalence of *C. dubliniensis*

A combination of culture- and molecular-based methods was used to stringently determine the prevalence of *C. dubliniensis* in the normal healthy and immunocompromised populations, as well as in avian-excrement-associated samples. Similar to results obtained from previous studies, this opportunistic pathogen was observed to preferentially colonise immunocompromised individuals in the present study, being found to be less prevalent in normal healthy individuals.

A previous study that investigated the prevalence of *C. dubliniensis* in the normal healthy population and HIV-infected population in South Africa yielded results that contradicted the results of all previous *C. dubliniensis* prevalence studies (Blignaut *et al.*, 2003). The authors reported that *C. dubliniensis* was more prevalent in normal healthy individuals than in the HIV-infected patients. It seems very unusual that *C. dubliniensis* preferentially colonises healthy individuals rather than immunocompromised individuals, as this species is an opportunistic pathogen. Blignaut

et al. (2003) suggested that in this region, *C. dubliniensis* colonisation was influenced by race rather than HIV-status. If this report is accurate, the increased prevalence of *C. dubliniensis* in the normal healthy population in South Africa suggests that *C. dubliniensis* strains in this region are more highly adapted for colonising healthy human hosts than isolates recovered from other geographical regions, or, that white individuals living in this geographical location are more prone to *C. dubliniensis* colonisation than black individuals. The unusual prevalence rates observed in South Africa have not been confirmed to date by any other studies, and it is important that the prevalence of *C. dubliniensis* in South Africa and other African countries is investigated. It may be possible that some racial or ethnic populations are predisposed to *C. dubliniensis* colonisation. Variations in diet or lifestyle amongst different racial or ethnic populations living in similar geographical locations may enable such preferential colonisation. Specific racial and distinct ethnic groups such as Aboriginal tribes, Inuits, Druze, Jews and Arabs merit such investigation, and should be screened along with samples from individuals who are not part of these groups but who live in the same geographical locations. Such studies should use both culture- and molecular-based methods as described in this study. Further investigation of South African *C. dubliniensis* isolates or isolates recovered from specific racial or ethnic groups may also reveal the presence of additional Cd25 fingerprint groups or MLST clades. It is worth noting that Cd25 group III was recently identified by Al Mosaid *et al.* (2005) whilst examining *C. dubliniensis* isolates recovered from individuals of Arab ethnic origin from the Middle East. Isolates belonging to this Cd25 group demonstrate high level 5FC resistance, which has not yet been reported in isolates recovered from other geographical regions or ethnic groups.

The recovery of *C. dubliniensis* from the surface of *I. uriae* (ticks) that were living in seabird excrement on a cliff face of the Great Saltee Island was reported for the first time by Nunn *et al.* (2007) during the course of the present study. The high *C. dubliniensis* recovery rate reported by Nunn *et al.* (2007) suggested that the natural host of *C. dubliniensis* may not be human, and that its natural niche may be the gastrointestinal tract of avian species. A subsequent prevalence study undertaken as part of the present study examined recently deposited seabird excrement and the data obtained suggested that this was not the case as only 3/134 (2.2%) samples yielded *C. dubliniensis*. The design of the *C. dubliniensis* MLST scheme provided and interesting opportunity to investigate the genetic relatedness between *C. dubliniensis* isolates recovered from avian-excrement-associated sources and humans, in order to determine

if transmission can occur between birds and humans. The MLST analysis of isolates recovered from humans and avian-excrement-associated samples in Ireland showed that all avian-excrement-associated isolates belonged to MLST clade C1, the most predominant *C. dubliniensis* MLST clade, and that isolates recovered from humans and avian-excrement-associated samples were very closely related. However, despite the close genetic relationship between humans and avian-excrement-associated *C. dubliniensis* isolates, the avian-excrement-associated isolates formed a distinct subgroup within this clade using many different clustering techniques. Analyses of the *MTL* and *CDR1* loci supported the formation of this subgroup, as all but one of the avian-excrement-associated isolates were *MTL* a/a homozygous and contained the homozygous TAG polymorphism in the *CDR1* gene. Together, these data suggested that the avian-excrement-associated isolates were highly clonal, despite the recovery of isolates from two separate locations which are separated by 150 km. Notably, one of the avian-excrement-associated isolates (AV7) recovered from the campus of Trinity College, Dublin, was indistinguishable from several *C. dubliniensis* isolates recovered from humans. This suggested that transmission of *C. dubliniensis* between human and avian hosts was possible. In the present prevalence study, all three excrement samples which yielded *C. dubliniensis* were collected on the campus of Trinity College, Dublin, which is an area that is heavily populated with humans, in contrast to the Great Saltee Island. It is possible that the seabirds inhabiting the Great Saltee Island came into contact with sewage somewhere off the Irish coast and became contaminated with human *C. dubliniensis* isolates, which may explain the greater *C. dubliniensis* isolate yields at this location.

Furthermore, it may be that the natural host of *C. dubliniensis* is the *I. uriae* tick rather than the gastrointestinal tract of seabirds, which would explain the increased *C. dubliniensis* recovery rate at the Great Saltee Island. However, there were no ticks present in the *C. dubliniensis*-positive avian-excrement-associated samples recovered in Dublin. Further analysis of such ticks recovered in the absence of seabird excrement is necessary in order to determine this. It may be possible that *C. dubliniensis* can be recovered from soil. However, it is unlikely that soil was the true source of the avian-excrement-associated isolates recovered in the present study. One of the avian-excrement-associated *C. dubliniensis* isolates included in the present study (AV6) was recovered from a nest in Trinity College Dublin, and another such isolate (AV5) was recovered directly from a gull.

7.1.2. Population structure analysis of *C. dubliniensis*

Advances in DNA sequencing technology in recent years have led to the development of MLST, which has been shown to be a reliable and reproducible alternative method to DNA fingerprinting using complex probes for investigating population structure and epidemiology of many *Candida* species (Bougnoux *et al.*, 2002; Dodgson *et al.*, 2003; Robles *et al.*, 2004; Tavanti *et al.*, 2005a, 2005b) despite originally being developed for use in bacterial species. In the present study, MLST was investigated as an alternative method for population structure analysis of *C. dubliniensis*.

The population structure of *C. dubliniensis* as defined by MLST (scheme D; Chapter 4, this thesis) correlated well with that defined by the Cd25 fingerprinting probe (Fig. 1.2), identifying the presence of three main clades. All isolates identified as ITS genotype 1 grouped in Cd25 group I or to MLST clade C1, (Joly *et al.*, 1999; Gee *et al.*, 2002; Al Mosaid *et al.*, 2005; Chapter 4, this thesis). Isolates identified as ITS genotype 2 belonged to Cd25 group II (Al Mosaid *et al.*, 2005). These ITS genotype 2 isolates were found to belong to MLST clade C2 (Chapter 4, this thesis), however previously identified ITS genotype 3 isolates belonging to Cd25 group II (Al Mosaid *et al.*, 2005) were regrouped to clade C3 using the MLST method (Chapter 4, this thesis). Isolates previously identified as ITS genotypes 3 or 4 that were recovered in the Middle East belonged to Cd25 group III. Similarly, these isolates belonged to MLST clade C3 (Chapter 4, this thesis).

The Cd25 fingerprinting, ITS genotype determination and MLST methods showed a good degree of correlation. These three different techniques are distinct from each other in the manner in which they assess the population structure of *C. dubliniensis*. The Cd25 fingerprinting method determines genetic relatedness according to the dispersion of species-specific, repetitive DNA sequences which are dispersed throughout the *C. dubliniensis* genome. ITS genotype determination is based on the nucleotide sequence of the ITS region of the rRNA operon, and MLST examines the abundance of SNPs in eight housekeeping genes. Why is the population structure of *C. dubliniensis* so robust, showing only slight variation when using different analytical techniques? It is most likely due to the clonal nature of *C. dubliniensis* evolution. Isolates belonging to each Cd25 group, ITS genotype or MLST clade are highly related to each other in every aspect of genomic analysis. Gee *et al.* (2002) reported that Cd25

group definition could be supported by electrophoretic karyotype analysis. All 23 Cd25 group II isolates examined by Gee *et al.* (2002) contained a chromosome-sized DNA band of approximately 1.7 Mb present in their karyotype profiles. This band was absent in all but five of 65 Cd25 group I isolates examined in the same study. The main difference between these methods of population structure analyses lies in the different discriminatory powers of each. The discriminatory power of the *C. dubliniensis* MLST scheme is lower than that of the Cd25 fingerprinting method, but is higher than that of the ITS genotyping method. However, ITS genotype determination can be easily combined with either MLST data or Cd25 fingerprinting to yield more information on the genetic relatedness of isolates within each MLST clade or Cd25 group.

In addition to the ability of MLST to define the population structure of *C. dubliniensis*, the high level of sequence homology (90%) shared by *C. albicans* and *C. dubliniensis* enabled the application of an identical MLST scheme to both *C. albicans* and *C. dubliniensis* as a method of comparative population structure analysis, and to assess the evolutionary relatedness of the two species (Fig. 4.2). Comparative population structure analysis revealed that the *C. albicans* MLST clades were significantly more divergent than those of *C. dubliniensis* (Fig. 4.2), exhibiting significantly higher levels of sequence variation and sites of heterozygosity. The lower levels of sequence variation observed in *C. dubliniensis* in comparison to *C. albicans* may be a result of the significantly smaller number of *C. dubliniensis* isolates investigated in the present study in comparison to *C. albicans* isolates. However, the *C. dubliniensis* isolates investigated in the present study were recovered from a wide range of geographical locations around the world (Fig. 3.1), and included isolates recovered from normal healthy individuals, and HIV-infected individuals, as well as from patients with cystic fibrosis, malignancies, diabetes, AIDS, or who were taking systemic steroids. *C. dubliniensis* isolates included in the study were also recovered from a broad range of anatomical sites, and 14 isolates included were recovered from avian-excrement-associated samples. These findings suggest that the population structure of *C. dubliniensis* is truly less divergent than that of *C. albicans*, despite both species evolving from a common ancestor. In order to investigate this hypothesis further, isolates of *C. dubliniensis* recovered in Africa, Scandinavia, Iceland, New Zealand and Asia should be investigated by MLST, as these geographical region have not been well represented in the present analysis (Fig. 3.1) and may possibly reveal the presence of

additional *C. dubliniensis* clades. Similarly, the inclusion of isolates from distinct racial and ethnic populations may also reveal the presence of additional subpopulations.

Previous population studies of *C. albicans* using the species-specific complex DNA fingerprinting probe Ca3 or using MLST have revealed the enrichment of certain clades with isolates from specific geographical locations (Blignaut *et al.*, 2002; Odds *et al.*, 2007). *C. dubliniensis* isolates belonging to MLST clade C3 or Cd25 group III were all recovered from individuals of Arab ethnic origin in the Middle East, suggesting an enrichment of clade C3 or Cd25 group III among isolates recovered from this geographical location and/or from humans of Arab ethnicity. In fact, no isolates belonging to this clade have been recovered from regions outside of the Middle East to date. All of the isolates belonging to Cd25 group III are ITS genotypes 3 or 4, and most exhibit high level intrinsic resistance to the antifungal agent 5FC (Al Mosaid *et al.*, 2005). Even the nature of clade-specific 5FC resistance displays more divergence in *C. albicans* than in *C. dubliniensis* due to the appearance of reduced 5FC susceptibilities in more than one *C. albicans* MLST clade (Pujol *et al.*, 2004; Odds *et al.*, 2007), and the variation in *C. albicans* MIC values from reduced 5FC susceptibility to high level 5FC resistance (Dodgson *et al.*, 2004; Hope *et al.*, 2004).

7.1.3. Clade-specific 5FC resistance in *C. dubliniensis*

The population structure analysis of *C. dubliniensis* has also offered some insights into the molecular mechanism(s) of 5FC resistance in the species. Due to the close genetic relatedness of clade C3 5FC-resistant isolates, the identical high MIC values exhibited by all 5FC-resistant isolates, the recovery of all such isolates in the Middle East and the clonal nature of *C. dubliniensis* reproduction, it was hypothesised that an identical molecular mechanism was used by all clade C3 *C. dubliniensis* isolates to mediate 5FC resistance.

Although the most common mechanism of 5FC resistance in *C. albicans* has been reported to be the Arg101Cys amino acid substitution in the UPRT, other amino acid substitutions in the Fca1p have also been associated with 5FC resistance in isolates which lack the Arg101Cys substitution (Dodgson *et al.*, 2004; Hope *et al.*, 2004). This contrasts with the clade-specific 5FC resistance of *C. dubliniensis*, of which all 5FC-resistant isolates examined to date belong to the same clade (Cd25group III or MLST clade C3), exhibit identical 5FC MICs, and contain the Ser29Leu substitution in the Fca1p. The Ser29Leu substitution was the only difference in the Fca1p between 12

5FC-resistant isolates and nine 5FC-susceptible isolates. Transformation of the *CdFCA1*^{R} gene from the 5FC-resistant isolate SA113 into the *ADH1* locus of the 5FC-susceptible isolate p7276 using the tetracycline-inducible expression plasmid pNIM1 had no effect on the recipient isolate. In the presence of DOX, p7276 remained 5FC-susceptible, despite having a copy of the highly expressed *CdFCA1*^{R} gene. Transformation of the *CdFCA1*^{S} gene from the 5FC-susceptible isolate p7276 into the *ADH1* locus of the 5FC-resistant isolates SA113 and SA109 using the same plasmid demonstrated that the 5FC-resistant phenotype of the transformant derivatives of SA113 and SA109 could be controlled by the presence or absence of DOX in the growth media. This demonstrated that 5FC resistance was mediated solely by the *CdFCA1* gene, as acquisition and expression of a functional *CdFCA1*^{S} gene mediates 5FC susceptibility. The present study demonstrates that the Ser29Leu substitution is a recessive mechanism of 5FC resistance, i.e., it must occur in both alleles of the *CdFCA1* gene in order to result in high level 5FC resistance. Site directed mutagenesis could have been used as an alternative method to demonstrate the recessive nature of the Ser29Leu substitution. Using this method, conversion of one of the TTA codons in the *CdFCA1*^{R} gene of a 5FC-resistant isolate to a TCA codon would most likely result in total or reduced 5FC susceptibility.

Quantitative real-time PCR analysis demonstrated no significant difference in *CdFCA1* expression between 5FC-resistant and 5FC-susceptible isolates, suggesting that the Ser29Leu substitution was the sole cause of 5FC resistance in these isolates. The present study suggests that the homozygous Ser29Leu substitution distorts the active site of the dimeric Fca1p, rendering it non-functional and thus blocking the deamination of 5FC to the toxic metabolite 5FU.

The Ser29Leu substitution appears to offer no obvious advantage to the Cd25 group III or MLST clade C3 isolates from the Middle East, as 5FC is not used in clinical settings. It is more likely that a conserved amino acid substitution occurring in the *CdFCA1* gene of these isolates coincidentally gives rise to 5FC resistance, rather than an evolutionary adaptation of these isolates to survive in the presence of 5FC during clinical treatment. This is not the first report of a clade-specific SNP that has resulted in the alteration of a protein involved in antifungal resistance in *C. dubliniensis*. Moran *et al.* (2002) reported a TAG polymorphism that occurred in the *CdCDR1* gene of 58% of ITS genotype 1 *C. dubliniensis* isolates, resulting in a truncated *CdCDR1*-encoded ABC transporter. Restoration of this SNP to a TAT codon in the *CdCDR1* gene

conferred the ability to mediate resistance to multiple drugs (Moran *et al.*, 2002). The Ser29Leu substitution in Fca1p and the TAG polymorphism in *CdCDR1* is a reflection of the relatively large numbers of pseudogenes found in the *C. dubliniensis* genome (115 in the genome of the type strain CD36 which has been sequenced) (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/), relative to *C. albicans* (8 pseudogenes in strain SC5314) (http://www.candidagenome.org/) (Jackson *et al.*, 2009).

7.1.4. Future directions in *C. dubliniensis* research

A thorough and widespread *C. dubliniensis* prevalence study is still lacking to date in comparison to *C. albicans*. In-depth analyses of a wide range of anatomical sites in both the normal healthy population and a wider range of immunocompromised patients should be carried out enabling the determination of the true niche of this species in humans. Investigation of the prevalence of *C. dubliniensis* in specific racial and distinct ethnic groups should also be carried out. Furthermore, a more widespread analysis of the prevalence of *C. dubliniensis* in non-human sources is needed. Such a study should screen botanical, avian, reptilian, amphibian and non-human mammalian sources in order to definitively determine if the true host of this species is human.

Any isolates that are recovered from these studies should be investigate by MLST in order to identify the presence of any further MLST clades. The enrichment of MLST clade 4 with *C. albicans* isolates from South Africa (Odds *et al.*, 2007) and the unusual *C. dubliniensis* prevalence rates reported from the same continent (Blignaut *et al.*, 2002) suggests that isolates recovered from South Africa may be genetically distinct from those recovered elsewhere. It is therefore, very important that *C. dubliniensis* isolates from this geographical region are investigated using MLST in order to accurately determine the true population structure of this species.

The *C. dubliniensis* MLST scheme recommended by the present study is not as discriminatory as the Cd25 fingerprinting method, or as the *C. albicans* MLST method. In order to maximise the discriminatory power of this method, perhaps some of the MLST loci which show lower levels of nucleotide sequence divergence should be substituted with loci that show more variation. For example, the present study noted high levels of SNPs in the *CdFCY21–CdFCY24* genes (Chapter 5, this thesis). Recently, a study by Jackson *et al.* (2009) reported that the *C. dubliniensis* genome is less polymorphic than that of *C. albicans*, and that the *C. dubliniensis* chromosome 6 contained the highest frequency of SNPs. In the recommended *C. dubliniensis* MLST

scheme, only the *ALA1* locus is located on chromosome 6, and the addition of a second locus from the same scheme might lend a higher discriminatory power to the *C. dubliniensis* MLST method. Another possibility is to supplement the recommended *C. dubliniensis* MLST scheme with microsatellite length polymorphism analysis, which has been shown a similar discriminatory ability to MLST in *C. albicans* (Garcia-Hermoso *et al.*, 2007).

A recent phylogenetic study of 42 complete fungal genomes has hinted that *C. tropicalis* is most likely the ancestor of both *C. albicans* and *C. dubliniensis* (Fitzpatrick *et al.*, 2006). According to the present comparative population structure analysis (Chapter 4, this thesis), it could be suggested that *C. dubliniensis* has evolved from its ancestor more recently than *C. albicans* has. This would explain why the population structure of *C. dubliniensis* is considerably less divergent than that of *C. albicans*, as *C. dubliniensis* may have had less time to evolve and diversify. However, the recent study by Jackson *et al.* (2009) has shown that since evolving from their common ancestor, *C. albicans* has expanded its genetic repertoire of certain gene families such as the *SAP*s, whereas *C. dubliniensis* has experienced reductive evolution, undergoing gene deletion and pseudogenisation (Jackson *et al.*, 2009). It appears that *C. albicans* is evolving to be more pathogenic, whereas *C. dubliniensis* is shedding many of the genes that associated with *Candida* pathogenicity in humans. It may be possible that while *C. albicans* is adapting to become a more successful human commensal and opportunistic pathogen, *C. dubliniensis* is adapting to become a successful environmental species. It has been previously shown that *C. dubliniensis* can produce chlamydospores more readily and abundantly than *C. albicans* in certain conditions (Gilfillan *et al.*, 1998; Staib & Morchäuser, 1999; Al Mosaid *et al.*, 2003; Khan *et al.*, 2004), and it is worth noting that in most of these conditions rely on the aqueous extract of various plant seeds or leaves.

References

Abele-Horn, M., Kopp, A., Sternberg, U., Ohly, A., Dauber, A., Russwurm, W., Büchinger, W., Nagengast, O. & Emmerling, P. (1996). A randomized study comparing fluconazole with amphotericin B/5-fluorocytosine for the treatment of systemic *Candida* infections in intensive care patients. *Infection* **24**, 426–432.

Agapow, P.M. & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* **1**, 101–102.

Ahmad, S., Khan, Z., Mokaddas, E. & Khan, Z.U. (2004). Isolation and molecular identification of *Candida dubliniensis* from non-human immunodeficiency virus-infected patients in Kuwait. *J Med Microbiol* **53**, 633–637.

Akins, R.A. (2005). An update on antifungal targets and mechanisms of resistance in *Candida albicans*. *Med Mycol* **43**, 285–318.

Al Mosaid, A., Sullivan, D., Salkin, I. F., Shanley, D. & Coleman, D.C. (2001). Differentiation of *Candida dubliniensis* from *Candida albicans* on staib agar and caffeic acid-ferric citrate agar. *J Clin Microbiol* **39**, 323–327.

Al Mosaid, A., Sullivan, D.J. & Coleman, D.C. (2003). Differentiation of *Candida dubliniensis* from *Candida albicans* on Pal's agar. *J Clin Microbiol* **41**, 4787–4789.

Al Mosaid, A., Sullivan, D.J., Polacheck, I., Shaheen, F.A., Soliman, O., Al Hedaithy, S., Al Thawad, S., Kabadaya, M. & Coleman, D.C. (2005). Novel 5-fluorocytosine-resistant clade of *Candida dubliniensis* from Saudi Arabia and Egypt identified by Cd25 fingerprinting. *J Clin Microbiol* **43**, 4026–4036.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Alvarez, M.I., Suárez, B.L. & Caicedo, L.D. (2009). Isolation of *Candida dubliniensis* for the first time in Cali, Colombia, and its identification with phenotyping methods. *Mycopathologia* **167**, 19–24.

Alves, S.H., Milan, E.P., de Laet Sant'Ana, P., Oliveira, L.O., Santurio, J.M. & Colombo, A.L. (2002). Hypertonic sabouraud broth as a simple and powerful test for *Candida dubliniensis* screening. *Diagn Microbiol Infect Dis* **43**, 85–86.

Anderson, J.B. (2005). Evolution of antifungal-drug resistance: mechanisms and pathogen fitness. *Nat Rev Microbiol* **3**, 547–556.

Asakura, K., Iwaguchi, S., Homma, M., Sukai, T., Higashide, K. & Tanaka, K. (1991). Electrophoretic karyotypes of clinically isolated yeasts of *Candida albicans* and *C. glabrata*. *J Gen Microbiol* **137**, 2531–2538.

Asmundsdóttir, L.R., Erlendsdottir, H., Haraldsson, G., Guo, H., Xu, J. & Gottfredsson, M. (2008). Molecular epidemiology of candidemia: evidence of clusters of smoldering nosocomial infections. *Clin Infect Dis* **47**, 17–24.

Back-Brito, G.N., Mota, A.J., Vasconcellos, T.C., Querido, S.M.R., Jorge, A.O.C., Reis, A.S.M., Balducci, I. & Koga-Ito, C.Y. (2009). Frequency of *Candida* spp. in the oral cavity of Brazilian HIV-positive patients and correlation with CD4 cell counts and viral load. *Mycopathologia* **167**, 81–87.

Bain, J.M., Tavanti, A., Davidson, A.D., Jacobsen, M.D., Shaw, D., Gow, N.A.R. & Odds, F.C. (2007). Multilocus sequence typing of the pathogenic fungus *Aspergillus fumigatus*. *J Clin Microbiol* **45**, 1469–1477.

Baixench, M.T., Aoun, N., Desnos-Ollivier, M., Garcia-Hermoso, D., Bretagne, S., Ramires, S., Piketty, C. & Dannaoui, E. (2007). Acquired resistance to echinocandins in *Candida albicans*: case report and review. *J Antimicrob Chemother* **59**, 1076–1083.

Baixench, M.T., Viguié, C. & Paugam, A. (2008). Isolation of *Candida dubliniensis* in a French hospital mycology laboratory. *J Med Microbiol* **57**, 258–259.

Bart-Delabesse, E., Boiron, P., Carlotti, A. & Dupont, B. (1993). *Candida albicans* genotyping in studies with patients with AIDS developing resistance to fluconazole. *J Clin Microbiol* **31**, 2933–2937.

Bille, J. (2000). Mechanisms and clinical significance of antifungal resistance. *Int J Antimicrob Agents* **16**, 331–333.

Blignaut, E., Pujol, C., Lockhart, S., Joly, S. & Soll, D.R. (2002). Ca3 fingerprinting of *Candida albicans* isolates from human immunodeficiency virus-positive and healthy individuals reveals a new clade in South Africa. *J Clin Microbiol* **40**, 826–836.

Blignaut, E., Pujol, C., Joly, S. & Soll, D.R. (2003). Racial distribution of *Candida dubliniensis* colonization among South Africans. *J Clin Microbiol* **41**, 1838–1842.

Boerlin, P., Boerlin-Petzold, F., Durussel, C., Addo, M., Pagani, J.L., Chave, J.P. & Bille, J. (1995). Cluster of oral atypical *Candida albicans* isolates in a group of human immunodeficiency virus-positive drug users. *J Clin Microbiol* **33**, 1129–1135.

Borg-von Zepelin, M., Niederhaus, T., Gross, U., Seibold, M., Monod, M. & Tintelnot, K. (2002). Adherence of different *Candida dubliniensis* isolates in the presence of fluconazole. *AIDS* **16**, 1237–1244.

Bostock, A., Khattak, M.N., Matthews, R. & Burnie, J. (1993). Comparison of PCR fingerprinting, by random amplification of polymorphic DNA, with other molecular typing methods for *Candida albicans*. *J Gen Microbiol* **139**, 2179–2184.

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.

Bougnoux, M–E., Tavanti, A., Bouchier, C., Gow, N.A.R., Magnier, A., Davidson, A. D., Maiden, M.C.J., D'Enfert, C. & Odds, F.C. (2003). Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. *J Clin Microbiol* **41**, 5265–5266.

Bougnoux, M–E., Aanensen, D. M., Morand, S., Théraud, 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., Diogo, D., François, N., Sendid, B., Veirmeire, S., Colombel, J.F., Bouchier, C., Van Kruiningen, H., d'Enfert, C. & Poulain, D. (2006). Multilocus sequence typing reveals intrafamilial transmission and microevolutions of *Candida albicans* isolates from the human digestive tract. *J Clin Microbiol* **44**, 1810–1820.

Branchini, M.L., Pfaller, M.A., Rhine-Chalberg, J., Frempong, T. & Isenberg, H. D. (1994). Genotypic variation and slime production among blood and catheter isolates of *Candida parapsilosis*. *J Clin Microbiol* **32**, 452–456.

Brena, S., delCarmen Rubio, M., Salesa, R., Iglesias, I., Gil, J., Rezusta, A., Moragues, M.D. & Pontón, J. (2004). Genotypes of *Candida dubliniensis* in clinical isolates. *Rev Iberoam Micol* **21**, 20–23.

Brito Gamboa, A., Mendoza, M., Fernández, A. & Díaz, E. (2006). Detection of *Candida dubliniensis* in patients with candidiasis in Caracas, Venezuela. *Rev Iberoam Micol* **23**, 81–84.

Buck, J. D. (1990). Isolation of *Candida albicans* and halophilic *Vibrio* spp. from aquatic birds in Connecticut and Florida. *Appl Environ Microbiol* **56**, 826–828.

Bujdáková, H., Melkusová, S., Soji, I., Mokrás, M. & Mikami, Y. (2004). Discrimination between *Candida albicans* and *Candida dubliniensis* isolated from HIV-positive patients by using commercial method in comparison with PCR assay. *Folia Microbiol (Praha)* **49**, 484–490.

Cafarchia, C., Camarda, A., Romito, D., Campolo, M., Quaglia, N.C., Tullio, D. & Otranto, D. (2006). Occurrence of yeasts in cloacae of migratory birds. *Mycopathologia* **161**, 229–234.

Carr, M.J., Clarke, S., O'Connell, F., Sullivan, D.J., Coleman, D.C. & O'Connell, B. (2005). First reported case of endocarditis caused by *Candida dubliniensis*. *J Clin Microbiol* **43**, 3023–3026.

Caugant, D.A. & Sandven, P. (1993). Epidemiological analysis of *Candida albicans* strains by multilocus enzyme electrophoresis. *J Clin Microbiol* **31**, 215–220.

Chapeland-Leclerc, F., Bouchoux, J., Goumar, A., Chastin, C., Villard, J. & Noël, T. (2005). Inactivation of the *FCY2* gene encoding purine-cytosine permease promotes cross-resistance to flucytosine and fluconazole in *Candida lusitaniae*. *Antimicrob Agents Chemother* **49**, 3101–3108.

Chen, K–W., Chen, Y–C., Lo, H–J., Odds, F.C., Wang, T–H., Lin, C–Y. & Li, S–Y. (2006). Multilocus sequence typing for analyses of clonality of *Candida albicans* strains in Taiwan. *J Clin Microbiol* **44**, 2172–2178.

Clinical and Laboratory Standards Institute. (2002). Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts; Approved Standard, M27-A2 *Clinical and Laboratory Standards Institute, Wayne, Pa.*

Coleman, D.C., Bennett, D.E., Sullivan, D.J., Gallagher, P.J., Henman, M.C., Shanley, D.B. & Russell, R.J. (1993). Oral *Candida* in HIV infection and AIDS: new perspectives/new approaches. *Crit Rev Microbiol* **19**, 61–82.

Coleman, D.C., Sullivan, D.J., Harrington, B., Haynes, K., Henman, M., Shanley, D., Bennett, D., Moran, G., McCreary, C., & O'Neill, L. (1997a). Molecular and phenotypic analysis of *Candida dubliniensis*: a recently identified species linked with oral candidosis in HIV-infected and AIDS patients. Oral Dis **3 Suppl 1**, S96–101.

Coleman, D.C., Sullivan, D.J., Bennett, D.E., Moran, G.P., Barry, H.J. & Shanley, D.B. (1997b). Candidiasis: the emergence of a novel species, *Candida dubliniensis*. *AIDS* **11**, 557–567.

Coleman, D.C., Rinaldi, M.G., Haynes, K.A., Rex, J.H., Summerbell, R.C., Anaissie, E.J., Li, A. & Sullivan, D.J. (1998). Importance of *Candida* species other than *Candida albicans* as opportunistic pathogens. *Med Mycol* **36 Suppl 1**, 156–165.

Colombo, A.L., Nucci, M., Park, B.J., Nouér, S.A., Arthington-Skaggs, B., da Matta, D.A., Warnock, D. & Morgan, J. and for the Brazilian Network Candidemia Study. (2006). Epidemiology of candidemia in Brazil: a nationwide sentinel surveillance of candidemia in eleven medical centers. *J Clin Microbiol* **44**, 2816–2823.

Cowen, L.E., Sirjusingh, C., Summerbell, R.C., Walmsley, S., Richardson, S., Kohn, L.M. & Anderson, J.B. (1999). Multilocus genotypes and DNA fingerprints do not predict variation in azole resistance among clinical isolates of *Candida albicans*. *Antimicrob Agents Chemother* **43**, 2930–2938.

Desnos-Ollivier, M., Bretagne, S., Bernède, C., Robert, V., Raoux, D., Chachaty, E., Forget, E., Lacroix, C. & Dromer, F. and the YEASTS group (2008). Clonal population of flucytosine-resistant *Candida tropicalis* from blood cultures, Paris, France. *Emerg Infect Dis* **14**, 557–565.

Dib, J.C., Dube, M., Kelly, C., Rinaldi, M.G. & Patterson, J.E. (1996). Evaluation of pulsed-field gel electrophoresis as a typing system for *Candida rugosa*: comparison of karyotype and restriction fragment length polymorphisms. *J Clin Microbiol* **34**, 1494–1496.

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**, 5709–5717.

Dodgson, A.R., Dodgson, K.J., Pujol, C., Pfaller, M.A. & Soll, D.R. (2004). Clade-specific flucytosine resistance is due to a single nucleotide change in the *FUR1* gene of *Candida albicans*. *Antimicrob Agents Chemother* **48**, 2223–2227.

Doebbeling, B.N., Hollis, R.J., Isenberg, H.D., Wenzel, R.P. & Pfaller, M.A. (1991). Restriction fragment analysis of a *Candida tropicalis* outbreak of sternal wound infections. *J Clin Microbiol* **29**, 1268–1270.

Doebbeling, B.N., Lehmann, P.F., Hollis, R.J., Wu, L.C., Widmer, A.F., Voss, A. & Pfaller, M.A. (1993). Comparison of pulsed-field gel electrophoresis with isoenzyme profiles as a typing system for *Candida tropicalis*. *Clin Infect Dis* **16**, 377–383.

Doern, G.V., Tubert, T.A., Chapin, K. & Rinaldi, M.G. 1986. Effect of medium composition on results of macrobroth dilution antifungal susceptibility testing of yeasts. *J Clin Microbiol* **24**, 507–11.

Doi, M., Mizuguchi, I., Homma, M. & Tanaka, K. (1994). Electrophoretic karyotypes of *Candida* yeasts recurrently isolated from single patients. *Microbiol Immunol* **38**, 19–23.

Donnelly, S.M., Sullivan, D.J., Shanley, D.B. & Coleman, D.C. (1999). Phylogenetic analysis and rapid identification of *Candida dubliniensis* based on analysis of *ACT1* intron and exon sequences. *Microbiology* **145**, 1871–1882.

Edelmann, A., Kruger, M. & Schmid, J. (2005). Genetic relationship between human and animal isolates of *Candida albicans*. *J Clin Microbiol* **43**, 6164–6166.

Elie, C. M., Lott, T. J., Reiss, E. & Morrison, C. J. (1998). Rapid identification of *Candida* species with species-specific DNA probes. *J Clin Microbiol* **36**, 3260–3265.

Enjalbert, B., Moran, G.P., Vaughan, C., Yeomans, T., Maccallum, D.M., Quinn, J., Coleman, D.C., Brown, A.J. & Sullivan, D.J. (2009). Genome-wide gene expression profiling and a forward genetic screen show that differential expression of the sodium ion transporter Ena21 contributes to the differential tolerance of *Candida albicans* and *Candida dubliniensis* to osmotic stress. *Mol Microbiol* **72**, 216–228.

Erbs, P., Exinger, F. & Jund, R. (1997). Characterization of the *Saccharomyces cerevisiae FCY1* gene encoding cytosine deaminase and its homologue *FCA1* of *Candida albicans*. *Curr Genet* **31**, 1–6.

EUCAST definitive document EDef 7.1. (2008). EUCAST definitive document EDef 7.1: method for the determination of broth dilution MICs of antifungal agents for fermentative yeasts. *Clin Microbiol Infect* **14,** 398-405.

Fell, J.W. (1993). Rapid identification of yeast species using three primers in a polymerase chain reaction. *Mol Mar Biol Biotechnol* **2**, 174–180.

Ferreira, T., Brèthes, D., Pinson, B., Napias, C. & Chevallier, J. (1997). Functional analysis of mutated purine-cytosine permease from *Saccharomyces cerevisiae*. A possible role of the hydrophilic segment 371-377 in the active carrier conformation. *J Biol Chem* **272**, 9697–9702.

Ferreira, T., Chevallier, J., Paumard, P., Napias, C. & Brèthes, D. (1999a). Screening of an intragenic second-site suppressor of purine-cytosine permease from *Saccharomyces cerevisiae*. Possible role of Ser272 in the base translocation process. *Eur J Biochem* **260**, 22–30.

Ferreira, T., Napias, C., Chevallier, J. & Brèthes, D. (1999b). Evidence for a dynamic role for proline376 in the purine-cytosine permease of *Saccharomyces cerevisiae*. *Eur J Biochem* **263**, 57–64.

Fitzpatrick, D. A., Logue, M. E., Stajich, J. E. & Butler, G. (2006). A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol Biol* **6**, 99.

Florent, M., Noël, T., Ruprich-Robert, G., Da Silva, B., Fitton-Ouhabi, V., Chastin, C., Papon, N. & Chapeland-Leclerc, F. (2009). Nonsense and missense mutations in *FCY2* and *FCY1* genes are responsible for flucytosine resistance and flucytosine-fluconazole cross-resistance in clinical isolates of *Candida lusitaniae*. *Antimicrob Agents Chemother* **53**, 2982–90.

Fotedar, R. & Al-Hedaithy, S.S.A. (2003). *Candida dubliniensis* at a university hospital in Saudi Arabia. *J Clin Microbiol* **41**, 1907–1911.

Galgiani, J.N. & Stevens, D.A. (1976). Antimicrobial susceptibility testing of yeasts: a turbidimetric technique independent of inoculum size. *Antimicrob Agents Chemother* **10**, 721–728.

Garcia-Hermoso, D., Cabaret, O., Lecellier, G., Desnos-Ollivier, M., Hoinard, D., Raoux, D., Costa, J. M., Dromer, F. & Bretagne, S. (2007). Comparison of microsatellite length polymorphism and multilocus sequence typing for DNA-Based typing of *Candida albicans*. *J Clin Microbiol* **45**, 3958-3963.

Gee, S.F., Joly, S., Soll, D.R., Meis, J.F.G.M., Verweij, P.E., Polacheck, I., Sullivan, D.J. & Coleman, D.C. (2002). Identification of four distinct genotypes of *Candida dubliniensis* and detection of microevolution *in vitro* and *in vivo*. *J Clin Microbiol* **40**, 556–574.

Ghannoum, M.A. & Rice, L.B. (1999). Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. *Clin Microbiol Rev* **12**, 501–517.

Gil-Lamaignere, C., Roilides, E., Hacker, J. & Müller, F.M. (2003). Molecular typing for fungi--a critical review of the possibilities and limitations of currently and future methods. *Clin Microbiol Infect* **9**, 172–185.

Gilfillan, G.D., Sullivan, D.J., Haynes, K., Parkinson, T., Coleman, D.C. & Gow, N.A. (1998). *Candida dubliniensis*: phylogeny and putative virulence factors. *Microbiology* **144**, 829–838.

Gossen, M. & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci USA* **89**, 5547–5551.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. & Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769.

Gräser, Y., Volovsek, M., Arrington, J., Schönian, G., Presber, W., Mitchell, T.G. & Vilgalys, R. (1996). Molecular markers reveal that population structure of the human pathogen *Candida albicans* exhibits both clonality and recombination. *Proc Natl Acad Sci USA* **93**, 12473–12477.

Groll, A.H., De Lucca, A.J. & Walsh, T.J. (1998). Emerging targets for the development of novel antifungal therapeutics. *Trends Microbiol* **6**, 117–124.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557–580.

Hannula, J., Saarela, M., Dogan, B., Paatsama, J., Koukila-Kähkölä, P., Pirinen, S., Alakomi, H.L., Perheentupa, J. & Asikainen, S. (2000). Comparison of virulence factors of oral *Candida dubliniensis* and *Candida albicans* isolates in healthy people and patients with chronic candidosis. *Oral Microbiol Immunol* **15**, 238–244.

Hasenclever, H.F. & Kogan, R.M. (1975). *Candida albicans* associated with the gastrointestinal tract of the common pigeon (*Columbia livia*). *Sabouraudia* **13**, 116–120.

Hayden, M.S., Linsley, P.S., Wallace, A.R., Marquardt, H. & Kerr, D.E. (1998). Cloning, overexpression, and purification of cytosine deaminase from *Saccharomyces cerevisiae*. *Protein Expr Purif* **12**, 173–184.

Hazen, K.C. (1995). New and emerging yeast pathogens. *Clin Microbiol Rev* **8**, 462–478.

Heidelberger, C., Chaudhuri, N.K., Danneberg, P., Mooren, D., Griesbach, L., Duschinsky, R., Schnitzer, R.J., Pleven, E. & Scheiner, J. (1957). Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature* **179**, 663–666.

Heidelberger, C., Griesbach, L., Montag, B.J., Mooren, D., Cruz, O., Schnitzer, R.J. & Grunberg, E. (1958). Studies on fluorinated pyrimidines. II. Effects on transplanted tumors. *Cancer Res* **18**, 305–317.

Hillen, W. & Berens, C. (1994). Mechanisms underlying expression of *Tn10* encoded tetracycline resistance. *Annu Rev Microbiol* **48**, 345–369.

Hope, W.W., Tabernero, L., Denning, D.W. & Anderson, M.J. (2004). Molecular mechanisms of primary resistance to flucytosine in *Candida albicans*. *Antimicrob Agents Chemother* **48**, 4377–4386.

Hoskins, J. & Butler, J.S. (2008). RNA-based 5-fluorouracil toxicity requires the pseudouridylation activity of Cbf5p. *Genetics* **179**, 323–330.

Hoyer, L.L., Fundyga, R., Hecht, J.E., Kapteyn, J.C., Klis, F.M. & Arnold, J. (2001). Characterization of agglutinin-like sequence genes from non-albicans *Candida* and phylogenetic analysis of the *ALS* family. *Genetics* **157**, 1555–1567.

Hull, C.M. & Johnson, A.D. (1999). Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*. *Science* **285**, 1271–1275.

Hunter, P.R. (1990). Reproducibility and indices of discriminatory power of microbial typing methods. *J Clin Microbiol* **28**, 1903–1905.

Ireton, G.C., Black, M.E. & Stoddard, B.L. (2003). The 1.14 A crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure* **11**, 961–972.

Jabra-Rizk, M.A., Baqui, A.A., Kelley, J.I., Falkler, W.A. Jr., Merz, W.G. & Meiller, T.F. (1999). Identification of *Candida dubliniensis* in a prospective study of patients in the United States. *J Clin Microbiol* **37**, 321–326.

Jabra-Rizk, M.A., Falkler, W.A. Jr., Merz, W.G., Baqui, A.A.M.A., Kelley, J.I. & Meiller, T.F. (2000). Retrospective identification and characterization of *Candida dubliniensis* isolates among *Candida albicans* clinical laboratory isolates from human immunodeficiency virus (HIV)-infected and non-HIV-infected individuals. *J Clin Microbiol* **38**, 2423–2426.

Jabra-Rizk, M.A., Ferreira, S.M.S., Sabet, M., Falkler, W.A., Merz, W.G. & Meiller, T.F. (2001). Recovery of *Candida dubliniensis* and other yeasts from human immunodeficiency virus-associated periodontal lesions. *J Clin Microbiol* **39**, 4520–4522.

Jackson, A.P., Gamble, J.A., Moran, G.P., Saunders, D., Harris, D., Aslett, M., Barrell, J.F., Butler, G., Citiulo, F., Coleman, D.C., de Groot, P.W.J., Goodwin, T.J., Keane, T., Quail, M.A., Munro, C.A., Pain, A., Poulter, R.T., Rajandream, M.-A., Renauld, H., Spiering, M.J., Tivey, A., Yeomans, T., Gow, N.A.R., Barrell, B., Sullivan, D.J. & Berriman, M. (2009). Comparative genomics of *Candida dubliniensis* and *C. albicans*. *Genome Res.* Oct 27. Epub ahead of print, doi:10.1101/gr.097501.109.

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.

Jacobsen, M.D., Bougnoux, M–E., d'Enfert, C. & Odds, F.C. (2008). Multilocus sequence typing of *Candida albicans* isolates from animals. *Res Microbiol* **159**, 436–440.

Jewtuchowicz, V.M., Mujica, M.T., Brusca, M.I., Sordelli, N., Malzone, M.C., Pola, S.J., Iovannitti, C.A. & Rosa, A.C. (2008). Phenotypic and genotypic identification of *Candida dubliniensis* from subgingival sites in immunocompetent subjects in Argentina. *Oral Microbiol Immunol* **23**, 505–509.

Jolley, K.A., Feil, E.J., Chan, M–S. & Maiden, M.C.J. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**, 1230–1231.

Joly, S., Pujol, C., Schröppel, K. & Soll, D.R. (1996). Development of two species-specific fingerprinting probes for broad computer-assisted epidemiological studies of *Candida tropicalis*. *J Clin Microbiol* **34**, 3063–3071.

Joly, S., Pujol, C., Rysz, M., Vargas, K. & Soll, D.R. (1999). Development and characterization of complex DNA fingerprinting probes for the infectious yeast *Candida dubliniensis*. *J Clin Microbiol* **37**, 1035–1044.

Kelly, S.L., Lamb, D.C., Kelly, D.E., Loeffler, J. & Einsele, H. (1996). Resistance to fluconazole and amphotericin in *Candida albicans* from AIDS patients. *Lancet* **348**, 1523–1524.

Kelly, S.L., Lamb, D.C., Kelly, D.E., Manning, N.J., Loeffler, J., Hebart, H., Schumacher, U. & Einsele, H. (1997). Resistance to fluconazole and cross-resistance to amphotericin B in *Candida albicans* from AIDS patients caused by defective sterol delta 5,6-desaturation. *FEBS Lett* **400**, 80–82.

Khan, Z.U., Ahmad, S., Mokaddas, E. & Chandy, R. (2004). Tobacco agar, a new medium for differentiating *Candida dubliniensis* from *Candida albicans*. *J Clin Microbiol* **42**, 4796-4798.

Khlif, M., Sellami, H., Sellami, A., Chelly, H., Makni, F., Bouaziz, M. & Ayadi, A. (2009). *Candida dubliniensis*: first identification in Sfax Hospital, Tunisia. *Mycoses* **52**, 171–175.

Kirkpatrick, W.R., Revankar, S.G., Mcatee, R.K., Lopez-Ribot, J.L., Fothergill, A.W., McCarthy, D.I., Sanche, S.E., Cantu, R.A., Rinaldi, M.G. & Patterson, T.F. (1998). Detection of *Candida dubliniensis* in oropharyngeal samples from human immunodeficiency virus-infected patients in North America by primary CHROMagar candida screening and susceptibility testing of isolates. *J Clin Microbiol* **36**, 3007–3012.

Ko, T–P., Lin, J–J., Hu, C–Y., Hsu, Y–H., Wang, A–H-J. & Liaw, S–H. (2003). Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. *J Biol Chem* **278**, 19111–19117.

Köhler, G.A., White, T.C. & Agabian, N. (1997). Overexpression of a cloned IMP dehydrogenase gene of *Candida albicans* confers resistance to the specific inhibitor mycophenolic acid. *J Bacteriol* **179**, 2331–2338.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.

Kurtz, J.E., Exinger, F., Erbs, P. & Jund, R. (1999). New insights into the pyrimidine salvage pathway of *Saccharomyces cerevisiae*: requirement of six genes for cytidine metabolism. *Curr Genet* **36**, 130–136.

Kwon-Chung, K.J., Riggsby, W.S., Uphoff, R.A., Hicks, J.B., Whelan, W.L., Reiss, E., Magee, B.B. & Wickes, B.L. (1989). Genetic differences between type I and type II *Candida stellatoidea*. *Infect Immun* **57**, 527–532.

Lamb, D.C., Kelly, D.E., Schunck, W–H., Shyadehi, A.Z., Akhtar, M., Lowe, D.J., Baldwin, B.C. & Kelly, S.L. (1997). The mutation T315A in *Candida albicans* sterol 14α-demethylase causes reduced enzyme activity and fluconazole resistance through reduced affinity. *J Biol Chem* **272**, 5682–5688.

Lehmann, P.F., Lin, D. & Lasker, B.A. (1992). Genotypic identification and characterization of species and strains within the genus *Candida* by using random amplified polymorphic DNA. *J Clin Microbiol* **30**, 3249–3254.

Lennox, E.S. (1955). Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* **1**, 190–206.

Liguori, G., Lucariello, A., Colella, G., De Luca, A. & Marinelli, P. (2007). Rapid identification of *Candida* species in oral rinse solutions by PCR. *J Clin Pathol* **60**, 1035–1039.

Lockhart, S.R., Joly, S., Pujol, C., Sobel, J.D., Pfaller, M.A. & Soll, D.R. (1997). Development and verification of fingerprinting probes for *Candida glabrata*. *Microbiology* **143**, 3733–3746.

Lockhart, S.R., Joly, S., Vargas, K., Swails-Wenger, J., Enger, L. & Soll, D.R. (1999). Natural defenses against *Candida* colonization breakdown in the oral cavities of the elderly. *J Dent Res* **78**, 857–868.

Lockhart, S.R., Daniels, K.J., Zhao, R., Wessels, D. & Soll, D.R. (2003). Cell biology of mating in *Candida albicans*. *Eukaryot Cell* **2**, 49–61.

Lott, T.J. & Scarborough, R.T. (2008). Development of a MLST-biased SNP microarray for *Candida albicans*. *Fungal Genet Biol* **45**, 803–811.

Lv, S–Q., Zhang, K–B., Zhang, E.E., Gao, F–Y., Yin, C–L., Huang, C–J., He, J–Q. & Yang, H. (2009). Antitumor efficiency of the cytosine deaminase/5-fluorocytosine suicide gene therapy system on malignant gliomas: an *in vivo* study. *Med Sci Monit* **15**, BR13–20.

Magee, B.B., Sanchez, M.D., Saunders, D., Harris, D., Berriman, M. & Magee, P.T. (2008). Extensive chromosome rearrangements distinguish the karyotype of the hypovirulent species *Candida dubliniensis* from the virulent *Candida albicans*. *Fungal Genet Biol* **45**, 338–350.

Maiden, M.C.J., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. & Spratt, B.G. (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.

Marichal, P., Vanden Bossche, H., Odds, F.C., Nobels, G., Warnock, D.W., Timmerman, V., Van Broeckhoven, C., Fay, S. & Mose-Larsen, P. (1997). Molecular biological characterization of an azole-resistant *Candida glabrata* isolate. *Antimicrob Agents Chemother* **41**, 2229–2237.

McCullough, M., Ross, B. & Reade, P.C. (1995a). Oral *Candida albicans* from patients infected with the human immunodeficiency virus and characterization of a genetically distinct subgroup of *Candida albicans*. *Aust Dent J* **40**, 91–97.

McCullough, M., Ross, B. & Reade, P. (1995b). Characterization of genetically distinct subgroup of *Candida albicans* strains isolated from oral cavities of patients infected with human immunodeficiency virus. *J Clin Microbiol* **33**, 696–700.

McCullough, M.J., Clemons, K.V. & Stevens, D.A. (1999). Molecular epidemiology of the global and temporal diversity of *Candida albicans*. *Clin Infect Dis* **29**, 1220–1225.

Meis, J.F.G.H., Ruhnke, M., De Pauw, B.E., Odds, F.C., Siegert, W. & Verweij, P.E. (1999). *Candida dubliniensis* candidemia in patients with chemotherapy-induced neutropenia and bone marrow transplantation. *Emerg Infect Dis* **5**, 150–153.

Melkusová, S., Lisalova, M., Pavlík, P. & Bujdáková, H. (2005). The first clinical isolates of *Candida dubliniensis* in Slovakia. *Mycopathologia* **159**, 369–371.

Metwally, L., Walker, M.J., Coyle, P.V., Hay, R.J., Hedderwick, S., McCloskey, B.V., O'Neill, H.J., Webb, C.H. & McMullan, R. (2007). Trends in candidemia and antifungal susceptibility in a university hospital in Northern Ireland 2001-2006. *J Infect* **55**, 174–178.

Momani, O.M. & Qaddoomi, A. (2005). Identification of *Candida dubliniensis* in a diagnostic microbiology laboratory. *East Mediterr Health J* **11**, 366–371.

Montour, L., Tey, R. & Xu, J. (2003). Isolation of *Candida dubliniensis* in an aboriginal community in Ontario, Canada. *J Clin Microbiol* **41**, 3423–3426.

Moran, G.P., Sullivan, D.J., Henman, M.C., McCreary, C.E., Harrington, B.J., Shanley, D.B. & Coleman, D.C. (1997). Antifungal drug susceptibilities of oral *Candida dubliniensis* isolates from human immunodeficiency virus (HIV)-infected and non-HIV-infected subjects and generation of stable fluconazole-resistant derivatives in vitro. *Antimicrob Agents Chemother* **41**, 617–623.

Moran, G.P., Sanglard, D., Donnelly, S.M., Shanley, D.B., Sullivan, D.J. & Coleman, D.C. (1998). Identification and expression of multidrug transporters responsible for fluconazole resistance in *Candida dubliniensis*. *Antimicrob Agents Chemother* **42**, 1819–1830.

Moran, G.P., Sullivan, D.J., Morschhäuser, J. & Coleman, D.C. (2002). The *Candida dubliniensis CdCDR1* gene is not essential for fluconazole resistance. *Antimicrob Agents Chemother* **46**, 2829–2841.

Moran, G.P., Stokes, C., Thewes, S., Hube, B., Coleman, D.C. & Sullivan, D.J. (2004). Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence-associated genes in *Candida dubliniensis*. *Microbiology* **150**, 3363–3382.

Mosca, C.O., Moragues, M.D., Llovo, J., Al Mosaid, A., Coleman, D.C. & Pontón, J. (2003). Casein agar: a useful medium for differentiating *Candida dubliniensis* from *Candida albicans*. *J Clin Microbiol* **41**, 1259–1262.

Mosca, C.O., Moragues, M.D., Brena, S., Rosa, A.C. & Pontón, J. (2005). Isolation of *Candida dubliniensis* in a teenager with denture stomatitis. *Med Oral Patol Oral Cir Bucal* **10**, 25–31.

Noël, T., François, F., Paumard, P., Chastin, C., Brèthes, D. & Villard, J. (2003). Flucytosine-fluconazole cross-resistance in purine-cytosine permease-deficient *Candida lusitaniae* clinical isolates: indirect evidence of a fluconazole uptake transporter. *Antimicrob Agents Chemother* **47**, 1275–1284.

Nolte, F.S., Parkinson, T., Falconer, D.J., Dix, S., Williams, J., Gilmore, C., Geller, R. & Wingard, J.R. (1997). Isolation and characterization of fluconazole- and amphotericin B-resistant *Candida albicans* from blood of two patients with leukemia. *Antimicrob Agents Chemother* **41**, 196–199.

Normark, S. & Schönebeck, J. (1972). *In vitro* studies of 5-fluorocytosine resistance in *Candida albicans* and *Torulopsis glabrata*. *Antimicrob Agents Chemother* **2**, 114–121.

Nunn, M.A., Schaefer, S.M., Petrou, M.A. & Brown, J.R. (2007). Environmental source of *Candida dubliniensis*. *Emerg Infect Dis* **13**, 747–750.

Odds, F.C. (1988). Ecology of *Candida* and epidemiology of candidosis. In *Candida and Candidosis*, pp. 68-92. Edited by F.C. Odds. London: Bailière Tindall.

Odds, F.C. & Bernaerts, R. (1994). CHROMagar Candida, a new differential isolation medium for presumptive identification of clinically important *Candida* species. *J Clin Microbiol* **32**, 1923–1929.

Odds, F.C., Van Nuffel, L. & Dams, G. (1998). Prevalence of *Candida dubliniensis* isolates in a yeast stock collection. *J Clin Microbiol* **36**, 2869–2873.

Odds, F.C., Davidson, A.D., Jacobsen, M.D., Tavanti, A., Whyte, J.A., Kibbler, C.C., Ellis, D.H., Maiden, M.C.J., Shaw, D.J. & Gow, N.A.R. (2006). *Candida albicans* strain maintenance, replacement, and microvariation demonstrated by multilocus sequence typing. *J Clin Microbiol* **44**, 3647–3658.

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*. *Eukaryot Cell* **6**, 1041–1052.

Odds, F.C. & Jacobsen, M.D. (2008). Multilocus sequence typing of pathogenic *Candida* species. *Eukaryot Cell* **7**, 1075–1084.

Orozco, A.S., Higginbotham, L.M., Hitchcock, C.A., Parkinson, T., Falconer, D., Ibrahim, A.S., Ghannoum, M.A. & Filler, S.G. (1998). Mechanism of fluconazole resistance in *Candida krusei*. *Antimicrob Agents Chemother* **42**, 2645–2649.

Paluszynski, J.P., Klassen, R., Rohe, M. & Meinhardt, F. (2006). Various cytosine/adenine permease homologues are involved in the toxicity of 5-fluorocytosine in *Saccharomyces cerevisiae*. *Yeast* **23**, 707–715.

Papon, N., Noël, T., Florent, M., Gibot-Leclerc, S., Jean, D., Chastin, C., Villard, J. & Chapeland-Leclerc, F. (2007). Molecular mechanism of flucytosine resistance in

Candida lusitaniae: Contribution of the *FCY2, FCY1* and *FUR1* genes to 5-fluorouracil and fluconazole cross-resistance. *Antimicrob Agents Chemother* **51**, 369–371.

Park, S., Wong, M., Marras, S.A., Cross, E.W., Kiehn, T.E., Chaturvedi, V., Tyagi, S. & Perlin, D.S. (2000). Rapid identification of *Candida dubliniensis* using a species-specific molecular beacon. *J Clin Microbiol* **38**, 2829–2836.

Park, Y.N. & Morschhäuser, J. (2005). Tetracycline-inducible gene expression and gene deletion in *Candida albicans*. *Eukaryot Cell* **4**, 1328–1342.

Pasqualotto, A.C., Nedel, W.L., Machado, T.S. & Severo, L.C. (2005). A 9-year study comparing risk factors and the outcome of paediatric and adults with nosocomial candidaemia. *Mycopathologia* **160**, 111–116.

Peltroche-Llacsahuanga, H., Döhmen, H. & Haase, G. (2002). Recovery of *Candida dubliniensis* from sputum of cystic fibrosis patients. *Mycoses* **45**, 15–18.

Perea, S., López-Ribot, J.L., Wickes, B.L., Kirkpatrick, W.R., Dib, O.P., Bachmann, S.P., Keller, S.M., Martinez, M. & Patterson, T.F. (2002). Molecular mechanisms of fluconazole resistance in *Candida dubliniensis* isolates from human immunodeficiency virus-infected patients with oropharyngeal candidiasis. *Antimicrob Agents Chemother* **46**, 1695–1703.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45.

Pfaller, M.A., Messer, S.A., Gee, S., Joly, S., Pujol, C., Sullivan, D.J., Coleman, D.C. & Soll, D.R. (1999). *In vitro* susceptibilities of *Candida dubliniensis* isolates tested against the new triazole and echinocandin antifungal agents. *J Clin Microbiol* **37**, 870–872.

Pfaller, M.A. & Diekema, D.J.; International Fungal Surveillance Participant Group. (2004). Twelve years of fluconazole in clinical practice: global trends in species

distribution and fluconazole susceptibility of bloodstream isolates of *Candida*. *Clin Microbiol Infect* **10 Suppl 1**, 11–23.

Pincus, D.H., Coleman, D.C., Pruitt, W.R., Padhye, A.A., Salkin, I.F., Geimer, M., Bassel, A., Sullivan, D.J., Clarke, M. & Hearn, V. (1999). Rapid identification of *Candida dubliniensis* with commercial yeast identification systems. *J Clin Microbiol* **37**, 3533–3539.

Pinjon, E., Sullivan, D., Salkin, I., Shanley, D. & Coleman, D. (1998). Simple, inexpensive, reliable method for differentiation of *Candida dubliniensis* from *Candida albicans*. *J Clin Microbiol* **36**, 2093–2095.

Polacheck, I., Strahilevitz, J., Sullivan, D., Donnelly, S., Salkin, I.F. & Coleman, D.C. (2000). Recovery of *Candida dubliniensis* from non-human immunodeficiency virus-infected patients in Israel. *J Clin Microbiol* **38**, 170–174.

Polak, A. & Scholer, H.J. (1975). Mode of action of 5-fluorocytosine and mechanisms of resistance. *Chemotherapy* **21**, 113–130.

Pontón, J., Rüchel, R., Clemons, K.V., Coleman, D.C., Grillot, R., Guarro, J., Aldebert, D., Ambroise-Thomas, P., Cano, J., Carrillo-Munoz, A.J., Gene, J., Pinel, C., Stevens, D.A. & Sullivan, D.J. (2000). Emerging pathogens. *Med Mycol* **38 Suppl 1**, 225–236.

Powderly, W.G. (1992). Mucosal candidiasis caused by non-albicans species of *Candida* in HIV-positive patients. *AIDS* **6**, 604–605.

Pressler, B.M., Vaden, S.L., Lane, I.F., Cowgill, L.D. & Dye, J.A. (2003). *Candida* spp. urinary tract infections in 13 dogs and seven cats: predisposing factors, treatment, and outcome. *J Am Anim Hosp Assoc* **39**, 263–270.

Pujol, C., Reynes, J., Renaud, F., Raymond, M., Tibayrenc, M., Ayala, F.J., Janbon, F., Mallié, M. & Bastide, J–M. (1993). The yeast *Candida albicans* has a clonal mode of reproduction in a population of infected human immunodeficiency virus-positive patients. *Proc Natl Acad Sci USA* **90**, 9456–9459.

Pujol, C., Renaud, F., Mallié, M., de Meeûs, T. & Bastide, J–M. (1997). Atypical strains of *Candida albicans* recovered from AIDS patients. *J Med Vet Mycol* **35**, 115–121.

Pujol, C., Pfaller, M. & Soll, D.R. (2002). Ca3 fingerprinting of *Candida albicans* bloodstream isolates from the United States, Canada, South America, and Europe reveals a European clade. *J Clin Microbiol* **40**, 2729–2740.

Pujol, C., Messer, S.A., Pfaller, M. & Soll, D.R. (2003). Drug resistance is not directly affected by mating type locus zygosity in *Candida albicans*. *Antimicrob Agents Chemother* **47**, 1207–1212.

Pujol, C., Pfaller, M.A. & Soll, D.R. (2004). Flucytosine resistance is restricted to a single genetic clade of *Candida albicans*. *Antimicrob Agents Chemother* **48**, 262–266.

Raut, S.H. & Varaiya, A. (2009). Differentiation of *Candida dubliniensis* on chrom agar and Pal's agar. *Indian J Med Microbiol* **27**, 55–58.

Redding, S.W., Bailey, C.W., Lopez-Ribot, J.L., Kirkpatrick, W.R., Fothergill, A.W., Rinaldi, M.G. & Patterson, T.F. (2001). *Candida dubliniensis* in radiation-induced oropharyngeal candidiasis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* **91**, 659–662.

Robles, J.C., Koreen, L., Park, S. & Perlin, D.S. (2004). Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of *Candida albicans*. *J Clin Microbiol* **42**, 2480–2488.

Romeo, O. & Criseo, G. (2009). Molecular epidemiology of *Candida albicans* and its closely related yeasts *Candida dubliniensis* and *Candida africana*. *J Clin Microbiol* **47**, 212–214.

Rustad, T.R., Stevens, D.A., Pfaller, M.A. & White, T.C. (2002). Homozygosity at the *Candida albicans MTL* locus associated with azole resistance. *Microbiology* **148**, 1061–1072.

Ryder, N.S., Wagner, S. & Leitner, I. (1998). *In vitro* activities of terbinafine against cutaneous isolates of *Candida albicans* and other pathogenic yeasts. *Antimicrob Agents Chemother* **42**, 1057–1061.

Sahand, I.H., Moragues, M.D., Eraso, E., Villar-Vidal, M., Quindós, G. & Pontón, J. (2005). Supplementation of CHROMagar Candida medium with Pal's medium for rapid identification of *Candida dubliniensis*. *J Clin Microbiol* **43**, 5768–5770.

Salkin, I.F., Pruitt, W.R., Padhye, A.A., Sullivan, D., Coleman, D. & Pincus, D.H. (1998). Distinctive carbohydrate assimilation profiles used to identify the first clinical isolates of *Candida dubliniensis* recovered in the United States. *J Clin Microbiol* **36**, 1467.

Sambrook, J.F., Fritsch E.F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual. 2nd Edition.* **Cold Spring Harbour Laboratory Press. New York.**

Sandven, P., Bevanger, L., Digranes, A., Haukland, H.H., Mannsåker, T. & Gaustad, P. and the Norwegian Yeast Study Group. (2006). Candidemia in Norway (1991 to 2003): results from a nationwide study. *J Clin Microbiol* **44**, 1977–1981.

Sanger, F., Nicklen, S. & Coulson, A.R. (1992). DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology* **24**, 104–108.

Sanglard, D., Kuchler, K., Ischer, F., Pagani, J.L., Monod, M. & Bille, J. (1995). Mechanisms of resistance to azole antifungal agents in *Candida albicans* isolates from AIDS patients involve specific multidrug transporters. *Antimicrob Agents Chemother* **39**, 2378–2386.

Sanglard, D., Ischer, F., Monod, M. & Bille, J. (1996). Susceptibilities of *Candida albicans* multidrug transporter mutants to various antifungal agents and other metabolic inhibitors. *Antimicrob Agents Chemother* **40**, 2300–2305.

Sanglard, D., Ischer, F., Monod, M. & Bille, J. (1997). Cloning of *Candida albicans* genes conferring resistance to azole antifungal agents: characterization of *CDR2*, a new multidrug ABC transporter gene. *Microbiology* **143**, 405–416.

Sanglard, D., Ischer, F., Calabrese, D., Majcherczyk, P.A. & Bille, J. (1999). The ATP binding cassette transporter gene *CgCDR1* from *Candida glabrata* is involved in the resistance of clinical isolates to azole antifungal agents. *Antimicrob Agents Chemother* **43**, 2753–2765.

Sanglard, D. (2002). Clinical relevance of mechanisms of antifungal drug resistance in yeasts. *Enferm Infecc Microbiol Clin* **20**, 462–469.

Sanglard, D. & Bille, J. (2002). Current understanding of the modes of action of and resistance mechanisms to conventional and emerging antifungal agents for treatment of *Candida* infections. In *Candida and Candidiasis*. Edited by R. A. Calderone. Washington DC: ASM Press.

Scherer, S. & Stevens, D.A. (1988). A *Candida albicans* dispersed, repeated gene family and its epidemiologic applications. *Proc Natl Acad Sci USA* **85**, 1452–1456.

Schmittgen, T.D. & Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101–1108.

Schoofs, A., Odds, F.C., Colebunders, R., Ieven, M. & Goossens, H. (1997). Use of specialised isolation media for recognition and identification of *Candida dubliniensis* isolates from HIV-infected patients. *Eur J Clin Microbiol Infect Dis* **16**, 296–300.

Schumacher, M.A., Carter, D., Scott, D.M., Roos, D.S., Ullman, B. & Brennan, R.G. (1998). Crystal structures of *Toxoplasma gondii* uracil phosphoribosyltransferase reveal the atomic basis of pyrimidine discrimination and prodrug binding. *EMBO J* **17**, 3219–3232.

Schumacher, M.A., Bashor, C.J., Song, M.H., Otsu, K., Zhu, S., Parry, R.J., Ullman, B. & Brennan, R.G. (2002). The structural mechanism of GTP stabilized

oligomerization and catalytic activation of the *Toxoplasma gondii* uracil phosphoribosyltransferase. *Proc Natl Acad Sci USA* **99**, 78–83.

Sebti, A., Kiehn, T.E., Perlin, D., Chaturvedi, V., Wong, M., Doney, A., Park, S. & Sepkowitz, K.A. (2001). *Candida dubliniensis* at a cancer center. *Clin Infect Dis* **32**, 1034–1038.

Séron, K., Blondel, M–O., Haguenauer-Tsapis, R. & Volland, C. (1999). Uracil-induced down-regulation of the yeast uracil permease. *J Bacteriol* **181**, 1793–1800.

Shen, S., Samaranayake, L.P., Yip, H.K. & Dyson, J.E. (2002). Bacterial and yeast flora of root surface caries in elderly, ethnic Chinese. *Oral Dis* **8**, 207–217.

Shin, J.H., Shin, D.H., Song, J.W., Kee, S.J., Suh, S.P. & Ryang, D.W. (2001). Electrophoretic karyotype analysis of sequential *Candida parapsilosis* isolates from patients with persistent or pecurrent fungemia. *J Clin Microbiol* **39**, 1258–1263.

Smith, J.M., Smith, N.H., O'Rourke, M. & Spratt, B.G. (1993). How clonal are bacteria? *Proc Natl Acad Sci USA* **90**, 4384–4388.

Soll, D.R. (2000). The ins and outs of DNA fingerprinting the infectious fungi. *Clin Microbiol Rev* **13**, 332–370.

Soll, D.R. & Pujol, C. (2003). *Candida albicans* clades. *FEMS Immunol Med Microbiol* **39**, 1–7.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503–517.

Staib, P. & Morschhäuser, J. (1999). Chlamydospore formation on Staib agar as a species-specific characteristic of *Candida dubliniensis*. *Mycoses* **42**, 521–524.

Staib, P., Moran, G.P., Sullivan, D.J., Coleman, D.C. & Morschhäuser, J. (2001). Isogenic strain construction and gene targeting in *Candida dubliniensis*. *J Bacteriol* **183**, 2859–2865.

Stiller, R.L., Bennett, J.E., Scholer, H.J., Wall, M., Polak, A. & Stevens, D.A. (1983). Correlation of *in vitro* susceptibility test results with *in vivo* response: flucytosine therapy in a systemic candidiasis model. *J Infect Dis* **147**, 1070–1077.

Stokes, C., Moran, G.P., Spiering, M.J., Cole, G.T., Coleman, D.C. & Sullivan, D.J. (2007). Lower filamentation rates of *Candida dubliniensis* contribute to its lower virulence in comparison with *Candida albicans*. *Fungal Genet Biol* **44,** 920–931.

Stolz, J. & Vielreicher, M. (2003). *Tpn1p*, the plasma membrane vitamin B6 transporter of *Saccharomyces cerevisiae*. *J Biol Chem* **278**, 18990–18996.

Sullivan, D.J., Bennett, D., Henman, M., Harwood, P., Flint, S., Mulcahy, F., Shanley, D. & Coleman, D.C. (1993). Oligonucleotide fingerprinting of isolates of *Candida* species other than *C. albicans* and of atypical *Candida* species from human immunodeficiency virus-positive and AIDS patients. *J Clin Microbiol* **31**, 2124–2133.

Sullivan, D.J., Westerneng, T.J., Haynes, K.A., Bennett, D.E. & Coleman, D.C. (1995). *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. *Microbiology* **141**, 1507–1521.

Sullivan, D.J., Haynes, K., Bille, J., Boerlin, P., Rodero, L., Lloyd, S., Henman, M. & Coleman, D.C. (1997). Widespread geographic distribution of oral *Candida dubliniensis* strains in human immunodeficiency virus-infected individuals. *J Clin Microbiol* **35**, 960–964.

Sullivan, D.J. & Coleman, D.C. (1998). *Candida dubliniensis*: characteristics and identification. *J Clin Microbiol* **36**, 329–334.

Sullivan, D.J., Moran, G.P., Donnelly, S., Gee, S., Pinjon, E., McCartan, B., Shanley, D.B. & Coleman, D.C. (1999). *Candida dubliniensis*: An update. *Rev Iberoam Micol* **16**, 72–76.

Sullivan, D.J., Moran, G.P., Pinjon, E., Al-Mosaid, A., Stokes, C., Vaughan, C. & Coleman, D.C. (2004). Comparison of the epidemiology, drug resistance mechanisms, and virulence of *Candida dubliniensis* and *Candida albicans*. *FEMS Yeast Res* **4**, 369–376.

Tassel, D. & Madoff, M.A. (1968). Treatment of *Candida* sepsis and *Cryptococcus meningitis* with 5-fluorocytosine. A new antifungal agent. *JAMA* **206**, 830–832.

Tavanti, A., Gow, N.A.R., Senesi, S., Maiden, M.C. & Odds, F.C. (2003). Optimization and validation of multilocus sequence typing for *Candida albicans*. *J Clin Microbiol* **41**, 3765–3776.

Tavanti, A., Gow, N.A.R., Maiden, M.C.J., Odds, F.C. & Shaw, D.J. (2004). Genetic evidence for recombination in *Candida albicans* based on haplotype analysis. *Fungal Genet Biol* **41**, 553–562.

Tavanti, A., Davidson, A.D., Fordyce, M.J., Gow, N.A.R., Maiden, M.C.J. & Odds, F.C. (2005a). 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.J., Shaw, D.J., Gow, N.A.R. & Odds, F.C. (2005b). Multilocus sequence typing for differentiation of strains of *Candida tropicalis*. *J Clin Microbiol* **43**, 5593–5600.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Tintelnot, K., Haase, G., Seibold, M., Bergmann, F., Staemmler, M., Franz, T. & Naumann, D. (2000). Evaluation of phenotypic markers for selection and identification of *Candida dubliniensis*. *J Clin Microbiol* **38**, 1599–1608.

Urzúa, B., Hermosilla, G., Gamonal, J., Morales-Bozo, I., Canals, M., Barahona, S., Cóccola, C. & Cifuentes, V. (2008). Yeast diversity in the oral microbiota of subjects with periodontitis: *Candida albicans* and *Candida dubliniensis* colonize the periodontal pockets. *Med Mycol* **46**, 783–793.

Us, E. & Cengiz, S.A. (2007). Prevalence and phenotypic evaluation of *Candida dubliniensis* in pregnant women with vulvovaginal candidosis in a university hospital in Ankara. *Mycoses* **50**, 13–20.

Vanden Bossche, H. (1997). Mechanisms of antifungal resistance. *Rev Iberoam Micol* **14**, 44–49.

Vandeputte, P., Larcher, G., Bergès, T., Renier, G., Chabasse, D. & Bouchara, J. P. (2005). Mechanisms of azole resistance in a clinical isolate of *Candida tropicalis*. *Antimicrob Agents Chemother* **49**, 4608–4615.

Vidotto, V., Mantoan, B., Pugliese, A., Pontón, J., Quindós, G., Aoki, S. & Ito-Kuwa, S. (2003). Adherence of *Candida albicans* and *Candida dubliniensis* to buccal and vaginal cells. *Rev Iberoam Micol* **20**, 52–54.

Vilela, M.M., Kamei, K., Sano, A., Tanaka, R., Uno, J., Takahashi, I., Ito, J., Yarita, K. & Miyaji, M. (2002). Pathogenicity and virulence of *Candida dubliniensis*: comparison with *C. albicans*. *Med Mycol* **40**, 249–257.

Waldorf, A.R. & Polak, A. (1983). Mechanisms of action of 5-fluorocytosine. *Antimicrob Agents Chemother* **23**, 79–85.

Warnock, D.W. (2007). Trends in the epidemiology of invasive fungal infections. *Nippon Ishinkin Gakkai Zasshi* **48**, 1–12.

Weber, E., Rodriguez, C., Chevallier, M.R. & Jund, R. (1990). The purine-cytosine permease gene of *Saccharomyces cerevisiae*: primary structure and deduced protein sequence of the *FCY2* gene product. *Mol Microbiol* **4**, 585–596.

Weems, J.J.,Jr. (1992). *Candida parapsilosis*: epidemiology, pathogenicity, clinical manifestations, and antimicrobial susceptibility. *Clin Infect Dis* **14**, 756–766.

Whelan, W.L. & Kerridge, D. (1984). Decreased activity of UMP pyrophosphorylase associated with resistance to 5-fluorocytosine in *Candida albicans*. *Antimicrob Agents Chemother* **26**, 570–574.

White, T.C. (1997). The presence of an R467K amino acid substitution and loss of allelic variation correlate with an azole-resistant lanosterol 14alpha demethylase in *Candida albicans*. *Antimicrob Agents Chemother* **41**, 1488–1494.

White, T.C., Marr, K.A. & Bowden, R.A. (1998). Clinical, cellular, and molecular factors that contribute to antifungal drug resistance. *Clin Microbiol Rev* **11**, 382–402.

Willis, A.M., Coulter, W.A., Sullivan, D.J., Coleman, D.C., Hayes, J.R., Bell, P.M. & Lamey, P.J. (2000). Isolation of *C. dubliniensis* from insulin-using diabetes mellitus patients. *J Oral Pathol Med* **29**, 86–90.

Wisplinghoff, H., Seifert, H., Tallent, S.M., Bischoff, T., Wenzel, R.P. & Edmond, M.B. (2003a). Nosocomial bloodstream infections in pediatric patients in United States hospitals: epidemiology, clinical features and susceptibilities. *Pediatr Infect Dis J* **22**, 686–691.

Wisplinghoff, H., Seifert, H., Wenzel, R.P. & Edmond, M.B. (2003b). Current trends in the epidemiology of nosocomial bloodstream infections in patients with hematological malignancies and solid neoplasms in hospitals in the United States. *Clin Infect Dis* **36**, 1103–1110.

Wisplinghoff, H., Bischoff, T., Tallent, S.M., Seifert, H., Wenzel, R.P. & Edmond, M.B. (2004). Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin Infect Dis* **39**, 309–317.

Wrobel, L., Whittington, J.K., Pujol, C., Oh, S–H., Ruiz, M.O., Pfaller, M.A., Diekema, D.J., Soll, D.R. & Hoyer, L.L. (2008). Molecular phylogenetic analysis of a geographically and temporally matched set of *Candida albicans* isolates from humans and nonmigratory wildlife in central Illinois. *Eukaryot Cell* **7**, 1475–1486.

Yang, C.W., Barkham, T.M., Chan, F.Y. & Wang, Y. (2003). Prevalence of *Candida* species, including *Candida dubliniensis*, in Singapore. *J Clin Microbiol* **41**, 472–474.

Yang, Y–L. & Lo, H–J. (2001). Mechanisms of antifungal agent resistance. *J Microbiol Immunol Infect* **34**, 79–86.

Yao, L., Sklenak, S., Yan, H. & Cukier, R.I. (2005). A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase. *J Phys Chem B* **109**, 7500–7510.

Publications

JOURNAL OF CLINICAL MICROBIOLOGY, Feb. 2008, p. 652–664 Vol. 46, No. 2
0095-1137/08/$08.00+0 doi:10.1128/JCM.01574-07

Multilocus Sequence Typing Reveals that the Population Structure of *Candida dubliniensis* Is Significantly Less Divergent than That of *Candida albicans*▽†

Brenda A. McManus,[1] David C. Coleman,[1]* Gary Moran,[1] Emmanuelle Pinjon,[1] Dorothée Diogo,[2] Marie-Elisabeth Bougnoux,[2] Silvia Borecká-Melkusova,[3] Helena Bujdáková,[3] Philip Murphy,[4] Christophe d'Enfert,[2] and Derek J. Sullivan[1]

Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental School & Hospital, Trinity College Dublin, University of Dublin, Dublin 2, Republic of Ireland[1]; Unité Biologie et Pathogénicité Fongiques, INRA USC2019, Institut Pasteur, Paris, France[2]; Department of Microbiology and Virology, Faculty of Natural Sciences, Comenius University, Mlynská dolina B-2, 842 15 Bratislava, Slovak Republic[3]; and Department of Laboratory Medicine, the Adelaide and Meath Hospital, Dublin, Incorporating the National Children's Hospital, Tallaght, Dublin 24, Ireland[4]

Received 8 August 2007/Returned for modification 4 November 2007/Accepted 20 November 2007

The pathogenic yeast *Candida dubliniensis* is phylogenetically very closely related to *Candida albicans*, and both species share many phenotypic and genetic characteristics. DNA fingerprinting using the species-specific probe Cd25 and sequence analysis of the internal transcribed spacer (ITS) region of the ribosomal gene cluster previously showed that *C. dubliniensis* is comprised of three major clades comprising four distinct ITS genotypes. Multilocus sequence typing (MLST) has been shown to be very useful for investigating the epidemiology and population biology of *C. albicans* and has identified many distinct major and minor clades. In the present study, we used MLST to investigate the population structure of *C. dubliniensis* for the first time. Combinations of 10 loci previously tested for MLST analysis of *C. albicans* were assessed for their discriminatory ability with 50 epidemiologically unrelated *C. dubliniensis* isolates from diverse geographic locations, including representative isolates from the previously identified three Cd25-defined major clades and the four ITS genotypes. Dendrograms created by using the unweighted pair group method with arithmetic averages that were generated using the data from all 10 loci revealed a population structure which supports that previously suggested by DNA fingerprinting and ITS genotyping. The MLST data revealed significantly less divergence within the *C. dubliniensis* population examined than within the *C. albicans* population. These findings show that MLST can be used as an informative alternative strategy for investigating the population structure of *C. dubliniensis*. On the basis of the highest number of genotypes per variable base, we recommend the following eight loci for MLST analysis of *C. dubliniensis*: Cd*AAT1b*, Cd*ACC1*, Cd*ADP1*, Cd*MPIb*, Cd*RPN2*, Cd*SYA1*, exCd*VPS13*, and exCd*ZWF1b*, where "Cd" indicates *C. dubliniensis* and "ex" indicates extended sequence.

Candida dubliniensis is a pathogenic yeast species that is phenotypically, genetically, and phylogenetically very closely related to *Candida albicans*, the yeast species most commonly associated with infection in humans (49, 51). Despite their close relationship, *C. albicans* is significantly more pathogenic (49, 50). *C. dubliniensis* is most commonly associated with oral carriage and infection in human immunodeficiency virus (HIV)-infected and diabetic patients, although it has been identified as a minor constituent of the commensal floras in the oral cavities of healthy individuals (40, 49, 50). Although *C. dubliniensis* has also been recovered from patients with systemic infections, its incidence is far lower than that of *C. albicans*. While the latter is responsible for 40 to 60% of cases of candidemia, *C. dubliniensis* has been identified in only 1 to 2% of blood culture yeast samples (11, 15, 27, 29–31). These epidemiological data are reflected in the results of animal infection model studies that demonstrated that *C. albicans* is significantly more pathogenic than *C. dubliniensis* (22, 47, 56). The reasons for the lower virulence of *C. dubliniensis* relative to that of *C. albicans* have not been investigated in detail; however, recently published data showed that *C. dubliniensis* has a reduced capacity to produce hyphae that results in lower levels of colonization and tissue invasion (47).

In order to be able to perform meaningful and informative epidemiological studies of *Candida* isolates, it is essential to be able to discriminate between unrelated strains of the species of interest. Ideally, strain differentiation methods should be highly discriminatory, reproducible, and suitable for the analysis of large numbers of isolates. To date, DNA fingerprinting using the species-specific, semirepetitive-sequence-containing DNA probe Cd25 has been the most widely applied and informative tool used for *C. dubliniensis* epidemiology and population analysis. When first developed, data generated using this probe showed that *C. dubliniensis* is comprised of two distinctive major clades, termed Cd25 fingerprint groups I and II (21,

* Corresponding author. Mailing address: Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental School & Hospital, Trinity College—Dublin, Dublin 2, Republic of Ireland. Phone: 353 1 6127276. Fax: 353 1 6127295. E-mail: dcoleman@dental.tcd.ie.

† Supplemental material for this article may be found at http://jcm.asm.org/.

▽ Published ahead of print on 5 December 2007.

26). Cd25 group I isolates are all closely related, with an average similarity coefficient (S_{AB}) value of approximately 0.8 (range, 0.8 to 0.86), where A and B represent two different strains. These isolates comprise the majority of isolates investigated to date recovered in many countries around the world, mainly from HIV-infected individuals (4, 21, 26). Furthermore, sequence analysis of the internally transcribed spacer (ITS) region of the rRNA gene cluster revealed that Cd25 group I isolates consist of a single ITS genotype: ITS genotype 1 (21). In contrast, Cd25 group II isolates are more diverse, with an average S_{AB} value of 0.52 (range, 0.07 to 0.67) (4), and consist of three separate ITS genotypes (ITS genotypes 2 to 4), which correspond to distinct subclades within the Cd25 group II fingerprinting clade (21). More recently, a third major clade, termed Cd25 group III, was identified among isolates from Egypt and Saudi Arabia and displayed an average S_{AB} value of 0.35 (range, 0.16 to 0.54) (4). The DNA fingerprints of Cd25 group III isolates are very distinctive relative to those of isolates from Cd25 groups I and II, and ITS sequence analysis revealed that they belong to ITS genotypes 3 and 4 (4). All Cd25 group III isolates examined to date exhibit resistance to 5-flucytosine (4).

DNA fingerprinting analysis of large numbers of *C. albicans* isolates using the species-specific, repetitive-sequence-containing Ca3 probe has demonstrated that the population structure of the species is complex, with five major clades, some of which appear to be associated with specific geographic locations (6, 42, 46). In comparison, the population structure of *C. dubliniensis* determined with the Cd25 probe is significantly less complex (4, 21, 26). Although DNA fingerprinting has been shown to be a very useful tool in the molecular epidemiological analysis of *C. albicans* and *C. dubliniensis* populations, it is time-consuming, expensive, and not conducive to interlaboratory comparisons. There are many other molecular strain-typing techniques that have been applied to the analysis of *Candida* species (e.g., karyotyping and randomly amplified polymorphic DNA fingerprinting, etc. [45]). However, all of these methodologies also suffer from drawbacks, particularly in relation to reproducibility. In the late 1990s, multilocus sequence typing (MLST), a technique based on the nucleotide sequence analysis of a set of housekeeping genes, was developed for the population analysis of several bacterial species (28). This technique has also been applied to the analysis of the diploid yeast *C. albicans* (9, 10, 54) and to other *Candida* species (17, 25, 53). In 2002, Bougnoux et al. identified six housekeeping gene loci that allowed accurate and reproducible discrimination between unrelated *C. albicans* isolates (9). A study by Tavanti et al. used four of these loci and a further four loci and also showed high levels of discrimination (54). These two groups of researchers subsequently revised the combination of loci used in the MLST analysis of *C. albicans*, aiming to identify the minimum number of loci required to maintain the high discriminatory standard of the scheme (10). An agreed consensus scheme between the different laboratories examined seven loci, *AAT1a*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b*, and allowed the application of MLST to the analysis of *C. albicans* epidemiology and population structure (7–9, 16, 36, 52). MLST analysis has since been demonstrated to be as sensitive as DNA fingerprinting (43), and due to the nature of the data (i.e., DNA sequences of specific loci), these can be used to create a large database generated by multiple laboratories (7). In the case of *C. albicans*, it has also been shown that the strain groupings identified by MLST correlate with clades of *C. albicans* organisms identified using the species-specific DNA fingerprinting probe Ca3 (52). When applied to specific epidemiological studies, the use of MLST has identified intrafamilial transmission of *C. albicans* from the human digestive tract, strain maintenance, replacement, and microevolution (8).

The purpose of the present study was to determine the usefulness of MLST for investigating the epidemiology and population structure of *C. dubliniensis*. In addition, we hypothesized that if the same set of MLST loci currently used for *C. albicans* could be applied to the analysis of *C. dubliniensis*, then comparative sequence data could be used to provide valuable information concerning the evolutionary relatedness of the two species.

MATERIALS AND METHODS

***C. dubliniensis* isolates.** The 50 epidemiologically unrelated *C. dubliniensis* isolates used in this study are shown in Table 1. Isolates were selected from diverse geographical locations and included representatives of the three Cd25 major fingerprinting clades and all four ITS genotypes. The majority of the isolates studied have been described previously; however, a number of new clinical isolates were also included (Table 1). These new isolates were initially presumptively identified on the basis of dark-green colony coloration on *Candida*-selective chromogenic agar (Oxoid Ltd., Hampshire, United Kingdom) and on the basis of hyphal fringe production on Pal's agar after 48 h of growth at 30°C as described by Al Mosaid et al. (3). The identities of presumptive *C. dubliniensis* isolates were further confirmed using the API ID 32C yeast identification system (bioMérieux, Paris, France) (37). Definitive identification of *C. dubliniensis* was confirmed by PCR using primers specific for the Cd*ACT1*-associated intron as described previously (18).

Isolates were routinely cultured on potato dextrose agar (Oxoid) medium, pH 5.6, at 37°C. Liquid cultures were grown overnight in yeast extract-peptone-dextrose broth at 37°C in an orbital incubator (Gallenkamp, Leicester, United Kingdom) at 200 rpm.

***C. albicans* isolates.** A selection of 50 *C. albicans* isolates (Table 2) was chosen to represent a range of the 17 MLST clades recently described by Odds et al. (35). Sequence data for each of the 50 *C. albicans* isolates were available at http://test1.mlst.net/ for the seven collaborative consensus MLST loci, *AAT1a*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b* (35), and at http://calbicans.mlst.net/ for the *RPN2* locus (9). All locus sequences were treated in a manner identical to that used for the *C. dubliniensis* sequence data, as described below.

Chemicals, enzymes, and oligonucleotides. Analytical-grade or molecular biology-grade chemicals were purchased from Sigma-Aldrich Ireland Ltd. (Tallaght, Dublin, Ireland) or Fisher Scientific Ltd. (Loughborough, United Kingdom). Enzymes were purchased from the Promega Corporation (Madison, WI) or New England Biolabs Inc. (Beverly MA) and used according to the manufacturers' instructions. Custom-synthesized oligonucleotides were purchased from Sigma Genosys Biotechnologies Europe Ltd. (Pampisford, Cambridgeshire, United Kingdom).

***C. dubliniensis* DNA isolation.** Isolates were grown overnight in 5 ml of yeast extract-peptone-dextrose broth, and cells from 1.5 ml of culture were harvested by centrifugation. DNA was extracted from the cells as described by Gallagher et al. (20).

Genotyping of *C. dubliniensis*. Template DNA was tested in separate PCR amplification experiments with each of the primer pairs G1F/G1R, G2F/G2R, G3F/G3R, and G4F/G4R to identify the ITS genotype of the isolate (21). Genotypes are ascribed based on the nucleotide sequences of the ITS1 and ITS2 regions and of the intervening 5.8S rRNA gene (57). Template DNAs from the four reference *C. dubliniensis* isolates (CD36, genotype 1; Can4, genotype 2; CD519, genotype 3, and p7718, genotype 4) previously described by Gee et al. (21) were used in control amplification reactions. Each PCR was carried out with one pair of ITS genotype-specific primers and the universal fungal primers RNAF/RNAR (19), which amplify approximately 610 bp from all fungal large-subunit rRNA genes and were used as an internal positive control. Genotyping experiments were performed on a minimum of two

TABLE 1. *C. dubliniensis* isolates investigated by MLST analysis[c]

Isolate	Country of origin	Yr of isolation	Underlying patient condition	Sample	MLST clade[a]	ITS genotype	Cd25 fingerprint group	Reference(s) or source
CD36[b]	Ireland	1988	HIV pos	Oral	1	1	I	21, 51
CD06033[b]	Ireland	2006	CF	Sputum	1	1	ND	This study
CD06031	Ireland	2006	CF	Sputum	1	1	ND	This study
CD060213	Ireland	2006	CF	Sputum	1	1	ND	This study
CD06041	Ireland	2006	CF	Sputum	1	1	ND	This study
CD06038[b]	Ireland	2006	CF	Sputum	1	1	ND	This study
CD06045	Ireland	2006	CF	Sputum	1	1	ND	This study
CD06037[b]	Ireland	2006	CF	Sputum	2	2	ND	This study
CD06036[b]	Ireland	2006	CF	Sputum	2	2	ND	This study
CD060215[b]	Ireland	2006	CF	Sputum	3	3	ND	This study
CD604[b]	France	2000	HIV asymptomatic	Throat	1	1	ND	This study
CD603[b]	France	2000	HIV asymptomatic	Throat	1	1	ND	This study
CM1[b]	Australia	1991	HIV pos	Oral	1	1	I	21, 51
SA105	Saudi Arabia	2002	Diabetes	Oral	1	1	I	4
SA115	Saudi Arabia	2002	Diabetes	Oral	1	1	I	4
SA116	Saudi Arabia	2002	Diabetes	Oral	1	1	I	4
SA108[b]	Saudi Arabia	2002	Diabetes	Oral	3	3	III	4
SA100[b]	Saudi Arabia	2002	Leukemia	Oral	3	3	III	4
SA121[b]	Saudi Arabia	2002	S/P renal tx	Oral	3	4	III	4
Eg202	Egypt	2002	Cancer	Oral	3	4	III	4
Eg203	Egypt	2002	Cancer	Oral	1	1	I	4
Eg204	Egypt	2002	Cancer	Oral	1	1	I	4
Eg207	Egypt	2002	Diabetes	Oral	3	4	III	4
p6785[b]	Israel	1999	HIV neg	Urine	3	3	II	4, 39
p7276[b]	Israel	1999	HIV neg	Respiratory tract	3	3	II	4
p7718	Israel	1999	HIV neg	Wound	3	4	III	4, 21
CD71	Argentina	1994	HIV pos	Oral	1	1	I	21, 48
CD98923	India	1998	HIV pos	Oral	1	1	I	2
B1324	United States	1998	NA	Tongue	1	1	ND	This study
B341	United States	1998	NA	Throat	1	1	ND	This study
PM6-2	Chile	2006	NA	NA	1	1	ND	This study
P2	Switzerland	1993	HIV pos	Oral	1	1	I	4, 21
1504	Slovakia	2005	NA	Tonsular swab	1	1	ND	32
8882	Slovakia	2005	Congenital hydrocephalus	Tonsular swab	1	1	ND	32
9097	Slovakia	2005	Leukemia	Tonsular swab	1	1	ND	32
966/3(1)[b]	Slovakia	2005	NA	Sputum	1	1	ND	32
966/3(2)	Slovakia	2005	NA	Sputum	1	1	ND	32
CCY29-177-1	Slovakia	2005	Pelvic organ inflammation	Cervical swab	1	1	ND	32
MLNIH0479	Thailand	2005	Leukemia	Blood	1	1	ND	13
MLNIH0720	Thailand	2005	TB, anemia, diabetes	Oral	1	1	ND	13
49831	Japan	2005	Anemia	Sputum	2	2	ND	13
IFM49883	Japan	2005	NA	NA	2	2	ND	13
IFM0492	Thailand	2005	Cancer	Blood	1	1	ND	13
IFM49832	Japan	2005	Diabetes	Sputum	1	1	ND	13
Can4	Canada	1996	HIV pos	Oral	2	2	II	21, 38
CD539	United Kingdom	1994	AIDS	Oral	2	2	II	21, 38
CBS2747	The Netherlands	1952	HIV neg	Sputum	2	2	II	21, 31
CD514	Ireland	1995	HIV neg	Oral	3	3	II	21
CD519	Ireland	1997	AIDS	Oral	3	3	II	21
CD75004	United Kingdom	1975	Diabetes	Oral	2	2	II	21, 38

[a] MLST clades assigned in the present study.
[b] Isolate included in assessment of sequence stability and MLST reproducibility.
[c] HIV pos, HIV infected; HIV neg, HIV negative; CF, cystic fibrosis; S/P renal tx, stage post-renal transplantation; NA, not available; TB, tuberculosis; ND, not determined.

occasions, with each isolate being tested with separately prepared *C. dubliniensis* template DNA.

Selection of loci for MLST analysis. Due to the high level of sequence homology between the majority of *C. albicans* and *C. dubliniensis* open reading frames (33), all loci previously examined for the purpose of MLST analysis in *C. albicans* were also investigated for their potential use with *C. dubliniensis*. The usefulness of the six genes in *C. albicans* examined as described by Bougnoux et al. (9) (i.e., *ACC1*, *VPS13*, *GLN4*, *ADP1*, *RPN2* and *SYA1*) and the additional four loci described by Tavanti et al. (54) (i.e., *AAT1a*, *AAT1b*, *MPIb*, and *ZWF1b*) was assessed. The *C. albicans* MLST locus sequences were used in separate BLAST searches against the *C. dubliniensis* genome sequence database (the Wellcome Trust Sanger Institute *C. dubliniensis* genome sequence project, http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/). For each MLST locus, sequences were aligned using the CLUSTAL W sequence alignment computer program (55), and the *C. albicans* primer binding regions were identified in the alignment. Any nucleotide differences in the primer binding regions between the

TABLE 2. *Candida albicans* isolates used in the MLST study

Isolate	MLST clade[a]	Reference
APRHCRC1	4	35
APRURC3	6	35
APRURM3	1	35
APRURM6	9	35
APRURM7	10	35
BCHURC10	10	35
BCHURC15	11	35
BCHURS24	S	35
BCHURS9	4	35
CCHHCRM11	4	35
CCHHCRM4	14	35
CCHHCRM7	9	35
CCHURM1	15	35
CCHURM12	9	35
CCHURM4	11	35
CCHURM9	16	35
CCHURS10	9	35
CP04	1	35
CP05	S	35
CP06	1	35
CP08	1	35
CP12	4	35
CP15	3	35
CP54	2	35
CP58	4	35
CP85	4	35
DPC111	9	8
DPC118	1	8
DPC168	11	8
DPC18	12	8
DPC2	11	8
DPC206	1	8
DPC208	5	8
DPC22	12	8
DPC25	3	8
DPC28	11	8
DPC35	17	8
DPC37	1	8
DPC44	11	8
DPC47	3	8
DPC5	11	8
DPC55	1	8
DPC6	11	8
DPC65	11	8
DPC66	9	8
DPC70	11	8
DPC81	2	8
DPC90	4	8

[a] MLST clades correspond to those recently identified by Odds et al. (35). S stands for singletons, i.e., isolates that were not assigned to any of the 17 clades.

species were adjusted in the corresponding *C. dubliniensis* oligonucleotides to facilitate optimum amplification. All of the *C. dubliniensis*-optimized primer pairs yielded a single PCR product of the expected size (ranging in size from 400 bp to 700 bp) (Table 3).

PCR amplification and sequence determination. PCR assays were carried out in 50-μl reaction volumes containing a 200 μM concentration of each deoxynucleoside triphosphate, 1.25 U of GoTaq polymerase (Promega), 10 μl (1×) of GoTaq FlexiBuffer (Promega), 3 μM $MgCl_2$, 100 pmol of each primer, and 1 ng of the DNA template. PCR products were purified using a QIAquick 96-well PCR purification kit (Qiagen Science, MD) and were sequenced on both strands using the same primers that had been used previously for amplification. DNA sequencing reactions were performed commercially by Cogenics (Essex, United Kingdom) using an ABI 3730*xl* DNA analyzer.

Sequence analysis and sequence type determination. Sequence analysis was performed by examination of chromatogram files using the ABI prism Seqscape software, version 2.0 (Applied Biosystems, Foster City, CA). The sequences of all of the loci examined are provided in the supplemental material. Numbers were assigned to unique genotypes for each locus, and genotype numbers were then combined to yield a diploid sequence type (DST) number. All genotype numbers and DST numbers are also available in the supplemental material. Maximum-parsimony trees and dendrograms based on analysis by the unweighted pair group method with arithmetic averages (UPGMA) were constructed using the Bionumerics software package, version 4.6 (Applied Maths.NV, Sint-Martens-Latem, Belgium), based on concatenated *C. albicans* and *C. dubliniensis* MLST sequences. The discriminatory power of each MLST scheme was determined using Hunter's formula (24).

Linkage disequilibrium and clonality. Linkage disequilibrium was assessed using the index of association, as described by Smith et al. (44) and as calculated with the Multilocus 1.3 software package, available at http://www.agapow.net/software/multilocus/ (1), using genotype numbers for all loci (scheme C) from all 50 *C. dubliniensis* isolates. The levels of significance for nonrandom association between loci were computed under the null hypothesis of a freely recombining population (panmixia).

Stability and reproducibility of the MLST method. The stability and reproducibility of the sequence data at each MLST locus were assessed by carrying out the sequence analysis in duplicate on two randomly selected isolates. For each locus and isolate, the duplicate DNA extractions, PCRs, and sequencing reactions were carried out independently. Resulting sequence duplicates for each isolate were compared to each other.

RESULTS

Development of an MLST scheme for *C. dubliniensis*. All loci previously examined for the purpose of analyzing MLSTs in *C. albicans* were also investigated for their potential use with *C. dubliniensis*. Homologous genes were found in each case, and comparison of the sequences of the complete open reading frames of the orthologous pairs revealed homologies ranging from 89% to 94%, while the parts of the genes that were analyzed for MLST purposes displayed 88% to 100% sequence identity (Table 3).

The current consensus MLST scheme for *C. albicans* examines 2,883 nucleotides from seven loci: *AAT1a*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b*. For the purposes of this study, this scheme is referred to as scheme A (Table 4). The corresponding scheme A loci were examined in the 50 *C. dubliniensis* isolates included in this study. The scheme A loci resulted in 23 single nucleotide polymorphisms (SNPs) from the equivalent 2,883 nucleotides (0.8%) and identified 33 genotypes. In scheme A, only one site (i.e., position 186 in *SYA1*) showed a polymorphism at the same position in both *C. albicans* and *C. dubliniensis*, whereas all other variable sites were at different locations in the two species. Tavanti et al. analyzed eight loci (54) (*AAT1a*, *AAT1b*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b*) for sites of heterozygosity among 50 isolates of *C. albicans*. The group identified 87 sites that displayed polymorphisms, 71 of which (81.6%) also displayed heterozygosity. The same loci were studied for sites of heterozygosity in the 50 isolates of *C. dubliniensis* (Table 1), and of the 30 sites displaying variability, 9 (30%) displayed heterozygosity.

Due to the low levels of polymorphism observed in the *C. dubliniensis* loci examined, we investigated whether increasing the length of the sequences examined had the potential to increase the discriminatory power of the method. To achieve this, additional nucleotide sequence data (range, 100 bp to 280 bp) at each of the loci in both the 5′ and 3′ directions of the original sequence fragment were also analyzed for the potential presence of polymorphic sites. Sequence analysis of the extended fragments in Cd*VPS13* and Cd*ZWF1b* (exCd*VPS13*

TABLE 3. Oligonucleotide primers and sequences used for *C. dubliniensis* MLST analysis

Locus	Amplicon size (bp) in *C. dubliniensis*	No. of bases analyzed in *C. dubliniensis* (bp)[a]	Coordinates of[b]:		Analyzed segment sequence homology (%)[c]	Primer[d]	Analyzed segment start (5′) and end (3′) points in *C. dubliniensis*[e]
			C. albicans amplicon	*C. dubliniensis* amplicon			
Cd*AAT1a*	480	373	+31–+508	+35–+497	91	5′-ATCAAACTACTAAATT TTTGAC-3′ (forward) 5′-CGGCAACATGATTA GCCC-3′ (reverse)	5′-ATTGAAA 3′-CGATTT
Cd*AAT1b*	491	341	+736–+1,226	+742–+1,233	95	5′-ATGGCTTATCAAGGT TTTGC-3′ (forward) 5′-GTAGCATAAACTGA ATAATCA-3′ (reverse)	5′-TTAACTAA 3′-TTGGGATCA
Cd*ACC1*	519	407	+3,184–+3,702	+3,190–+3,708	93	5′-GCCAGAGAAATTTT GATCCAATGT-3′ (forward) 5′-TTCATCAACATCATC CAAGTG-3′ (reverse)	5′-TTTTGAGAT 3′-TACAAGA
Cd*ADP1*	537	443	+868–+1,404	+868–+1,404	90	5′-GAGCCAAGTATGAA TGACTTG-3′ (forward) 5′-TTGATCAACAAACCC GATAAT-3′ (reverse)	5′-TACGTTGCAA 3′-GGAAATCCAA
Cd*GLN4*	483	404	+82–+564	+82–+564	100	5′-GAGATAGTTAAGAA TAAAAAAGTTG-3′ (forward) 5′-GTCTCTTTCGTCTTTA GGACCCAATC-3′ (reverse)	5′-TCTGCTTTA 3′-TTCAAACC
Cd*MPIb*	486	375	+406–+891	+406–+892	90	5′-ACCAGAAATGGCC-3′ (forward) 5′-GCAGCCATACATTCA ATTAT-3′ (reverse)	5′-TTTAAGC 3′-GGGAAGCA
Cd*RPN2*	447	306	+1,012–+1,458	+1,015–+1,461	90	5′-TTTATGCATGCTGGT ACTACTGATG-3′ (forward) 5′-TAACCCCATACTCAA AGCAGCAGCCT-3′ (reverse)	5′-TTGGTCCAAG 3′-GTCTTTACGA
Cd*SYA1*	543	391	+2,284–+2,826	+2,284–+2,826	90	5′-AGAAGAATAGTTGC TCTTACTG-3′ (forward) 5′-GTTGCCCTTACCACC AGCTTT-3′ (reverse)	5′-TAAATCCAAG 3′-AGTCTGTATCT
Cd*VPS13* (exCd*VPS13*)	741	403 (675)	+4,854–+5,594	+4,858–+5,598	88	5′-CGTTGAGAGATATTC GACTT-3′ (forward) 5′-ACGGATCGATCGCC AATCC-3′l (reverse)	5′-CCTTGATATG 3′-AAAATCTTGG (5′-AGAGCAAACG) (3′-AAACCTTGG)

Cd*ZWF1b* (exCd*ZWF1b*)	702	491 (621)	+787–+1,469	+766–+1,468	93	5′-GTTTCATTTGATCCT GAAGC-3′ (forward) 5′-GCCATTGATAAGTAC CTGGAT-3′ (reverse)	5′-CAAACCAGG 3′-TAGAATTAC (5′-AAAGTTTTA AAA) (3′-GAAAATATTT GAAA)

[a] Numbers in parentheses are for extended sequence fragments analyzed for polymorphisms.
[b] Nucleotide coordinates for MLST amplicons are numbered based on the adenine residue of the ATG start codon at the 5′ end of the gene designated +1.
[c] Sequence homology between *C. dubliniensis* and *C. albicans* at the sequenced fragments.
[d] Nucleotide bases shown in bold typeface and underlined denote nucleotide bases that vary between *C. albicans* and *C. dubliniensis*.
[e] Loci for which additional sequence data were analyzed in both the 5′ and 3′ directions of the original sequence fragment (range, 100 bp to 280 bp) due to the presence of additional polymorphic sites (underlined and bold). Sequences in parentheses are those used with the extended sequence fragments.

and exCd*ZWF1b*, respectively) revealed the presence of additional SNPs (Table 3). The inclusion of these two extended sequences together with the sequences of the other five loci generated a second MLST scheme, termed scheme B (Table 4), which was made up of the sequences of the Cd*AAT1a*, Cd*ACC1*, Cd*ADP1*, Cd*MPIb*, Cd*SYA1*, exCd*VPS13*, and exCd*ZWF1b* loci. Examination of exCd*VPS13* revealed one extra SNP, which gave rise to two extra genotypes (Table 5), and exCd*ZWF1b* yielded two further SNPs, which in turn gave rise to one extra genotype. Scheme B resulted in a new total of 3,285 nucleotides, 26 (0.8%) of which displayed SNPs, resulting in four additional genotypes (Table 5).

In a further attempt to improve the discriminatory power of this method for *C. dubliniensis*, three other sequence segments in *C. dubliniensis* that were previously investigated for possible use in the MLST analysis of *C. albicans* were also analyzed, although these were not included in the final consensus *C. albicans* MLST scheme. Sequences from these additional loci, Cd*AAT1b*, Cd*GLN4*, and Cd*RPN2* (Table 3), comprised an additional 1,051 nucleotides, 11 of which displayed SNPs (1.05%). The 10 sequence fragments (the scheme B loci and the three additional loci) were all analyzed together in a third scheme, termed scheme C (Table 4), which was based on the sequences of the Cd*AAT1a*, Cd*AAT1b*, Cd*ACC1*, Cd*ADP1*, Cd*GLN4*, Cd*MPIb*, Cd*RPN2*, Cd*SYA1*, exCd*VPS13*, and exCd*ZWF1b* loci. Cd*GLN4* displayed four of the further 11 SNPs; however, it gave rise to only two genotypes among the 50 isolates investigated. Similarly, Cd*AAT1b* displayed four SNPs, giving rise to five genotypes, and the Cd*RPN2* locus contained three polymorphic sites, which gave rise to three genotypes. Scheme C analyzed a total of 4,336 nucleotides, 37 of which (0.85%) displayed SNPs, resulting in a total of 48 genotypes from 50 *C. dubliniensis* isolates. The minimum number of MLST loci for maximum discrimination among isolates of *C. dubliniensis* was determined according to the highest number of genotypes per variable base in each locus (Table 4 and Table 5) and is referred to as scheme D, consisting of the eight loci Cd*AAT1b*, Cd*ACC1*, Cd*ADP1*, Cd*MPIb*, Cd*RPN2*, Cd*SYA1*, exCd*VPS13*, and exCd*ZWF1b*. Scheme D displayed a total of 32 SNPs, resulting in 40 genotypes from the 50 isolates of *C. dubliniensis* investigated. Polymorphic sites and resulting genotypes are summarized in Tables 4 and 5.

Nucleotide polymorphisms and amino acid changes. The effect of nucleotide polymorphism on the resulting amino acid sequence was investigated by mapping the triplet codons for each gene fragment and examining the effect that the SNP had on each codon in question. Nucleotide polymorphisms and amino acid substitutions are summarized in Table 6. Of the 37 SNPs identified in the 10 *C. dubliniensis* loci examined (with substitution of Cd*VPS13* for exCd*VPS13* and Cd*ZWF1b* for exCd*ZWF1b*), 13 (35%) resulted in nonsynonymous amino acid changes, while all of the remaining nucleotide changes resulted in synonymous polymorphisms. Seven of the 13 nonsynonymous polymorphisms affected the resulting amino acid substantially, i.e., affecting the pH (4 of 7) or polarity (3 of 7) of the amino acid.

Stability and reproducibility of *C. dubliniensis* MLST. Sequence analysis was performed in duplicate for each of two randomly selected isolates (Table 1) per MLST locus, using independently prepared DNA, PCR, and sequencing reac-

TABLE 4. Summary of loci used in individual MLST schemes

MLST scheme	Locus[a]												No. of polymorphic sites	No. of DSTs	Discriminatory power[b]
	AAT1a	*AAT1b*	*ACC1*	*ADP1*	*GLN4*	*MPIb*	*RPN2*	*SYA1*	*VPS13*	(ex)*VPS13*	*ZWF1b*	(ex)*ZWF1b*			
A	+	−	+	+	−	+	−	+	+	−	+	−	23	20	0.899
B	+	−	+	+	−	+	−	+	−	+	−	+	26	22	0.901
C	+	+	+	+	+	+	+	+	−	+	−	+	37	26	0.910
D	−	+	+	+	−	+	+	+	−	+	−	+	32	25	0.909
E	+	−	+	+	−	+	+	+	+	−	+	−	26	22	0.906

[a] + and − denote the presence and absence, respectively, of individual loci in each MLST scheme.
[b] Determined using Hunter's formula (24).

tions. For each MLST locus, duplicate sequences for each isolate showed 100% sequence identity, displaying full conservation of both SNPs and sites of heterozygosity at each locus.

***C. dubliniensis* strain differentiation by MLST.** Examination of the *C. albicans* MLST locus set (scheme A) in the 50 *C. dubliniensis* isolates investigated identified 20 unique DSTs based on the unique combinations of the genotype numbers for the seven loci examined. Application of Hunter's formula (24) to this data set infers that MLST using scheme A has a discriminatory power of 0.899 when applied to *C. dubliniensis*, compared with a value of 0.996 when applied to *C. albicans* (35). Extension of the Cd*VPS13* and Cd*ZWF1b* loci (scheme B) resulted in a further 2 DSTs, and adding the Cd*AAT1b*, Cd*GLN4*, and Cd*RPN2* loci (scheme C in Table 4) resulted in a further 4 DSTs, giving a total of 26 DSTs from the 10 loci. Extending the Cd*ZWF1b* and Cd*VPS13* gene fragments and including the three other loci in scheme C increased the discriminatory power to 0.9102 (Table 4). DST 4 was the most common DST identified using the scheme A set of loci in *C. dubliniensis*, corresponding to 14 of the 50 isolates examined. The set of seven loci identified 11 DSTs that were unique to single *C. dubliniensis* isolates. When the larger set of loci was used in scheme C, the same 14 isolates from DST 4 in scheme A referred to above also gave an identical sequence type, this time termed DST 7, which correlated with the previously identified DST 4, making this the most common DST in the scheme examined for *C. dubliniensis*. Using this larger set of loci, 19 of the 26 DSTs were unique to single isolates.

TABLE 5. Summary of characteristics of loci used in the MLST analysis

Locus	Length of analyzed sequence	No. (%) of polymorphic nucleotide sites	No. of polymorphic amino acid sites	No. of resulting genotypes
AAT1a	373	1 (0.27)	0	2
AAT1b	341	4 (1.17)	2	5
ACC1	407	3 (0.74)	2	4
ADP1	443	5 (1.13)	1	6
GLN4	404	4 (0.99)	2	2
MPIb	375	2 (0.5)	1	7
RPN2	306	3 (0.98)	0	3
SYA1	391	5 (1.28)	2	5
VPS13	403	2 (0.5)	1	2
(ex)*VPS13*	675	3 (0.44)	2	4
ZWF1b	491	5 (1.02)	1	4
(ex)*ZWF1b*	621	7 (1.13)	1	6

Population analysis of *C. dubliniensis* using MLST. A UPGMA dendrogram was constructed based on the sequence data from all 10 loci (scheme C) examined in *C. dubliniensis* by using the Bionumerics version 4.6 software program (Fig. 1). At a cutoff node of 99.7% sequence homology, the dendrogram revealed the presence of three major clades of isolates termed C1 to C3, which showed a significant degree of correlation with the major clades previously identified by fingerprinting using the Cd25 fingerprint probe (Fig. 1 and Table 1). Clades C1 and C2 corresponded to the previously identified Cd25 fingerprint groups I and II, respectively, whereas clade C3 included strains previously identified as belonging to Cd25 fingerprint groups II and III, respectively (Fig. 1 and Table 1). Furthermore, clade C1 consisted solely of ITS genotype 1 isolates, clade C2 consisted solely of ITS genotype 2 isolates, whereas clade C3 consisted of ITS genotype 3 and 4 isolates.

Linkage disequilibrium and clonality. Genetic diversity and linkage disequilibria were assessed by using statistics implemented in the Multilocus 1.3 software (1) and genotypes obtained for all loci investigated in this study (scheme C). Each of these statistics tested the null hypothesis of a freely recombining population. Highly significant linkage disequilibria (P was $<10^{-5}$ with 100,000 randomizations) were found for both the total collection of 50 *C. dubliniensis* isolates and a reduced collection of 26 isolates that represented each of the 26 DSTs defined by scheme C (i.e., that did not contain repeated genotypes) (data not shown). These results provide evidence that the sample of *C. dubliniensis* isolates analyzed in this study represents a clonal population.

Comparative population analysis of *C. albicans* and *C. dubliniensis* using MLST. *C. albicans* MLST sequences were available for the seven consensus MLST loci (i.e., *AAT1a*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b*) and an extra locus, *RPN2*. The resulting 3,189 nucleotides from the eight loci resulted in a fifth scheme, termed scheme E (Table 4). All locus sequences were concatenated and treated as one sequence for each of 50 *C. albicans* isolates and the corresponding sequences in 50 *C. dubliniensis* isolates in order to allow comparison between the two species using MLST. These concatenated sequences were used to construct a maximum-parsimony tree in order to assess comparative phylogenies between the species (Fig. 2). The maximum-parsimony tree displays the comparative divergence within the two species, as well as the level of relatedness between the two species. The maximum-parsimony tree demonstrated that *C. albicans* isolates belonging to different MLST clades can differ at as many as 90

TABLE 6. Nucleotide polymorphisms and amino acid substitutions in *C. dubliniensis* loci[d]

Gene fragment	Nucleotide position showing polymorphism[a]	Triplet polymorphism[b]	Amino acids[c]	Corresponding codon in *C. albicans*	Corresponding amino acid residue in *C. albicans*
Cd*AAT1a*	2	(T/C)TG	L/L	TTG	L
Cd*AAT1b*	105	A(T/A)T	I/N	ATC	I
	114	TC(A/T)	S/S	TCA	S
	121	(G/A)AA	E/K	GAA	E
	129	TC(T/G)	S/S	TCT	S
Cd*ACC1*	128	TT(T/C)	F/F	TTC	F
	313	T(T/C)G	L/S	CTG	L
	390	AA(A/T)	K/N	AAA	K
Cd*ADP1*	31	AT(C/T)	I/I	ATT	I
	214	AG(C/T)	S/S	AGT	S
	234	G(A/G)G	E/G	GAG	E
	331	TG(T/C)	C/C	TGT	C
	421	TC(G/T)	S/S	TCA	S
Cd*GLN4*	104	GA(A/T)	E/D	GTA	V
	141	AA(T/C)	N/N	AAA	K
	147	GA(T/C)	D/D	GAC	D
	244	(A/G)TT	I/V	ATT	I
Cd*MPIb*	101	A(G/A)A	R/K	AGA	R
	375	GC(A/G)	A/A	GCA	A
Cd*RPN2*	127	TC(C/A)	S/S	TCC	S
	136	GC(C/T)	A/A	GCC	A
	298	GC(T/C)	A/A	GCA	A
Cd*SYA1*	82	AA(C/T)	N/N	AAT	N
	148	TT(A/G)	L/L	TTA	L
	186	**AA(T/C)**	**N/N**	**AAT**	**N**
	203	G(T/C)T	V/A	GTT	V
	383	(T/G)CT	S/A	GCT	A
Cd*VPS13*	397*	AA(A/G)	K/K	AAG	K
	402†	T(G/C)G	W/S	TGG	W
exCd*VPS13*	45	GA(G/C)	E/D	GAC	D
	654*	AA(A/G)	K/K	AAG	K
	659†	T(G/C)G	W/S	TGG	W
Cd*ZWF1b*	1‡	AC(C/T)	T/T	AAA	K
	28§	AC(G/A)	T/T	ACC	T
	265¶	AA(G/A)	K/K	AAA	K
	299‖	(G/T)AT	D/Y	GAT	D
	334#	AA(A/G)	K/K	AAA	K
exCd*ZWF1b*	78‡	AC(C/T)	T/T	AAA	K
	105§	AC(G/A)	T/T	ACC	T
	342¶	AA(G/A)	K/K	AAA	K
	376‖	(G/T)AT	D/Y	GAT	D
	411#	AA(A/G)	K/K	AAA	K
	573	TA(T/C)	Y/Y	TAT	Y
	588	TC(A/T)	S/S	TCA	S

[a] Nucleotide positions refer to the positions in the analyzed fragment only.
[b] Position of the polymorphism relative to its codon.
[c] Amino acid resulting from the codon both before and after polymorphism.
[d] The nucleotide sites shown in bold typeface and underlined denote the only common nucleotide sites that display polymorphism in both *C. albicans* and *C. dubliniensis*. Identical polymorphic sites in both of the extended gene segments analyzed and in the shorter sequences that were based on the *C. albicans* scheme lengths are marked with the same symbols.

nucleotide sites but that isolates belonging to the same MLST clade can differ at as many as 31 nucleotide sites (data not shown). In contrast, our data showed that *C. dubliniensis* consisted of three closely related clades, termed clades C1, C2, and C3 (Fig. 2), that show complete agreement with the clades described by the UPGMA dendrogram (Fig. 1). *C. dubliniensis* isolates from the three MLST clades differ from each other at a minimum of 10 nucleotide sites (range, 10 to 23 nucleotides). Isolates belonging to clade C1 differ at a maximum of 10 nucleotide sites, isolates belonging to clade C2 differ at a maximum of 6 nucleotide sites, and isolates belonging to clade C3 differ at a maximum of 8 nucleotide sites (data not shown).

DISCUSSION

MSLT has previously been shown to be a useful tool in the analysis of the epidemiology and population structure of *C. albicans* (8–10, 35, 52, 54). The purpose of the present study was to determine the usefulness of MSLT in the analysis of *C. dubliniensis*. In addition, we investigated whether the same set of MSLT loci currently used for *C. albicans* could be applied to the analysis of *C. dubliniensis* due to the high levels of nucleotide sequence homology (~90%) shared by the two species, thus allowing comparative sequence data to be used to provide valuable information concerning the evolutionary relatedness of the two species.

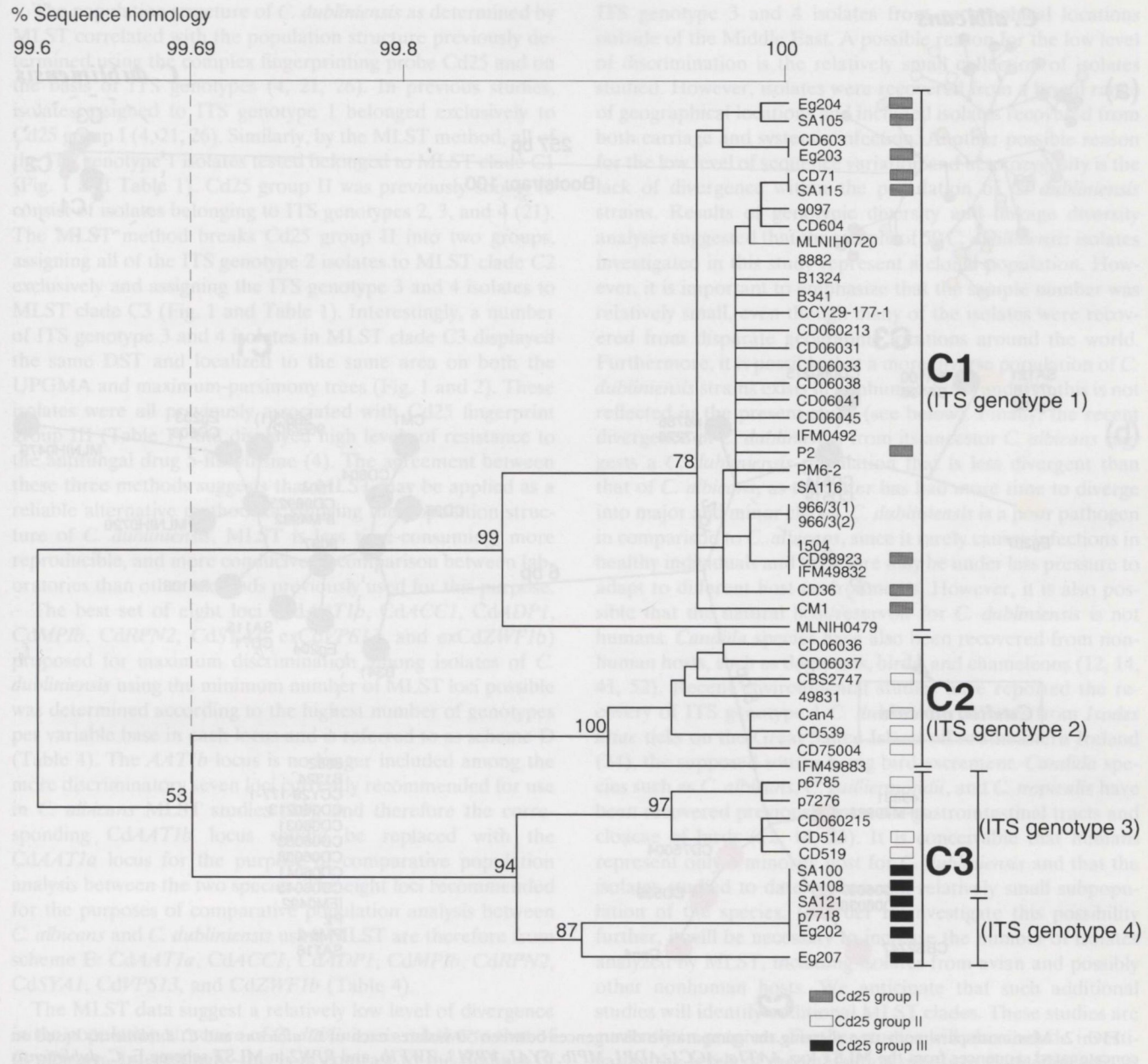


FIG. 1. UPGMA dendrogram based on concatenated sequences from the MLST scheme C loci Cd*AAT1a*, Cd*ACC1*, Cd*ADP1*, Cd*MPIb*, Cd*SYA1*, exCd*VPS13*, exCd*ZWF1b*, Cd*RPN2*, Cd*GLN4*, and Cd*AAT1b* (Table 3), showing percentages of sequence homology for 50 *C. dubliniensis* isolates from a broad range of geographical locations, including isolates from all known ITS genotypes (Table 1). Three distinct major clades termed clades C1 to C3 are evident at 99.69% sequence homology. C1 consists of ITS genotype 1 isolates exclusively, C2 consists of ITS genotype 2 isolates exclusively, and C3 consists of ITS genotype 3 and 4 isolates. Twenty-three of the 50 isolates included in the study were previously fingerprinted with the *C. dubliniensis*-specific probe Cd25 (4, 21), and the Cd25 fingerprint groups of these isolates are indicated by shaded rectangular boxes.

The current MLST scheme in use for isolates for *C. albicans* examines seven loci: *AAT1a*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b* (scheme A in Table 4) and was also applied in the analysis of *C. dubliniensis* isolates. It demonstrated poor levels of discrimination among *C. dubliniensis* isolates, identifying only 20 DSTs from 50 isolates. SNPs have been detected in 172 of these 2,883 nucleotides (6%) in *C. albicans* (5). To date, 1,391 isolates of *C. albicans* have been examined, identifying 1,005 DSTs (35). An identical scheme demonstrated poor levels of discrimination among *C. dubliniensis* isolates, identifying only 20 DSTs from 50 isolates. Discriminatory powers were improved by extension of the Cd*VPS13* and Cd*ZWF1b* loci and by incorporation of an additional three loci, *AAT1b*, *GLN4*, and *RPN2*, which had also previously been examined for use in the MLST analysis of *C. albicans*. The scheme (i.e., scheme C) which examined the largest number of loci (Table 4) identified 26 DSTs from the 50 *C. dubliniensis* isolates. The 10 loci included in scheme C were concatenated

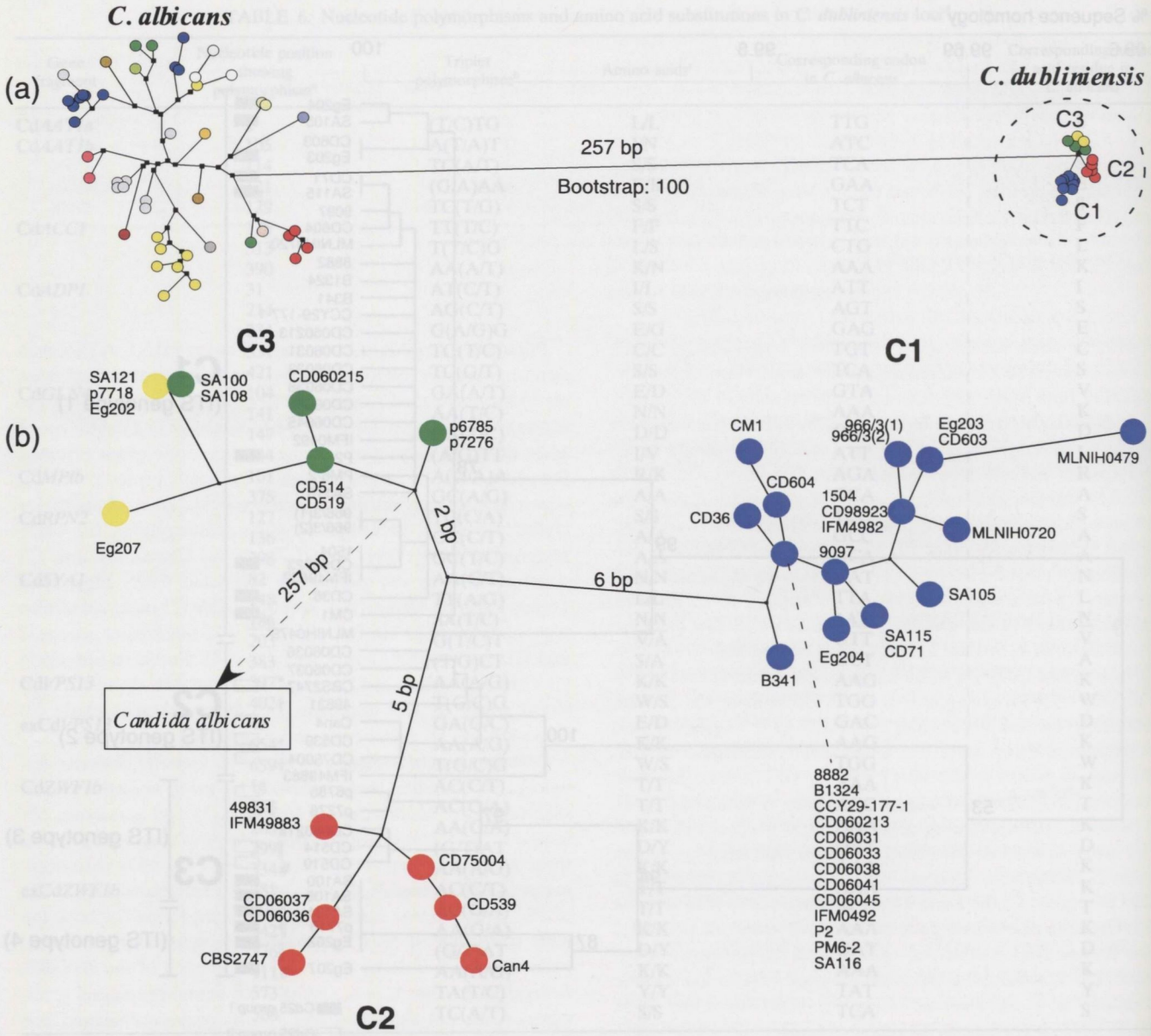


FIG. 2. Maximum-parsimony tree showing the comparative divergences between 50 isolates each of *C. albicans* and *C. dubliniensis* based on concatenated sequences from the MLST loci *AAT1a*, *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, *ZWF1b*, and *RPN2* in MLST scheme E. *C. dubliniensis* isolates were selected from a diverse range of geographic locations and from all four ITS genotypes (Table 1). *C. albicans* isolates were selected as representatives of the MLST clades recently described by Odds et al. (35) (Table 2). (a) Comparative divergence between the *C. albicans* and *C. dubliniensis* isolates tested showing that the two species are separated by 257-bp differences. The *C. dubliniensis* isolates formed three closely related groups of isolates (C1 to C3), which correspond to those identified in the UPGMA dendrogram shown in Fig. 1. Isolates from distinct clades are highlighted by specific colors for each species. (b) Enlarged view of the three *C. dubliniensis* major clades encircled in panel a. Clade C1 consists exclusively of ITS genotype 1 isolates (blue), clade C2 consists exclusively of ITS genotype 2 isolates (red), and clade C3 consists of ITS genotype 3 isolates (green) and ITS genotype 4 isolates (yellow).

and used in the construction of a UPGMA dendrogram (Fig. 1), which clustered the isolates into three distinct major clades at a cutoff sequence homology of 99.7%. A maximum-parsimony tree (Fig. 2) was also constructed using concatenated sequence data, this time with a further scheme, scheme E (Table 4), which was based on the seven loci from scheme A and an extra locus, *RPN2* (Table 4). Sequence data for these eight loci were available for both *C. albicans* and *C. dubliniensis*, thus enabling the comparative study of the population structures for both species. Our study shows that the *C. albicans* clades are more divergent than those observed in *C. dubliniensis* (Fig. 2). The maximum-parsimony tree identified three distinct major clades (C1 to C3) in the population of *C. dubliniensis*, which were identical to those of the UPGMA dendrogram (Fig. 1). Overall, 257 nucleotides of the 3,189 nucleotides analyzed (8%) were identified as being different between the two species, correlating with the level of sequence homology typically exhibited between the two species (33).

The population structure of *C. dubliniensis* as determined by MLST correlated with the population structure previously determined using the complex fingerprinting probe Cd25 and on the basis of ITS genotypes (4, 21, 26). In previous studies, isolates assigned to ITS genotype 1 belonged exclusively to Cd25 group I (4, 21, 26). Similarly, by the MLST method, all of the ITS genotype 1 isolates tested belonged to MLST clade C1 (Fig. 1 and Table 1). Cd25 group II was previously shown to consist of isolates belonging to ITS genotypes 2, 3, and 4 (21). The MLST method breaks Cd25 group II into two groups, assigning all of the ITS genotype 2 isolates to MLST clade C2 exclusively and assigning the ITS genotype 3 and 4 isolates to MLST clade C3 (Fig. 1 and Table 1). Interestingly, a number of ITS genotype 3 and 4 isolates in MLST clade C3 displayed the same DST and localized to the same area on both the UPGMA and maximum-parsimony trees (Fig. 1 and 2). These isolates were all previously associated with Cd25 fingerprint group III (Table 1) and displayed high levels of resistance to the antifungal drug 5-flucytosine (4). The agreement between these three methods suggests that MLST may be applied as a reliable alternative method for studying the population structure of *C. dubliniensis*. MLST is less time-consuming, more reproducible, and more conducive to comparison between laboratories than other methods previously used for this purpose.

The best set of eight loci (Cd*AAT1b*, Cd*ACC1*, Cd*ADP1*, Cd*MPIb*, Cd*RPN2*, Cd*SYA1*, exCd*VPS13*, and exCd*ZWF1b*) proposed for maximum discrimination among isolates of *C. dubliniensis* using the minimum number of MLST loci possible was determined according to the highest number of genotypes per variable base in each locus and is referred to as scheme D (Table 4). The *AAT1b* locus is no longer included among the more discriminatory seven loci currently recommended for use in *C. albicans* MLST studies (10), and therefore the corresponding Cd*AAT1b* locus should be replaced with the Cd*AAT1a* locus for the purpose of comparative population analysis between the two species. The eight loci recommended for the purposes of comparative population analysis between *C. albicans* and *C. dubliniensis* using MLST are therefore from scheme E: Cd*AAT1a*, Cd*ACC1*, Cd*ADP1*, Cd*MPIb*, Cd*RPN2*, Cd*SYA1*, Cd*VPS13*, and Cd*ZWF1b* (Table 4).

The MLST data suggest a relatively low level of divergence in the population structure of *C. dubliniensis* relative to that of *C. albicans*. Interestingly, this is in contrast to the findings of electrophoretic karyotypic analysis showing that isolates of *C. dubliniensis* have significantly different major karyotypic patterns due to the processes of microevolution and chromosomal rearrangement (21). The low level of sequence variation throughout the population of *C. dubliniensis* suggests that MLST may not be ideal for local epidemiological studies (e.g., an outbreak in a hospital); in this instance, karyotypic analysis may be more appropriate. Karyotype analysis was able to distinguish the Cd25 group I and II isolates to a large extent but was unable to distinguish among the ITS genotypes. In comparison, the MLST method is able to distinguish readily among certain ITS genotypes, most notably the ITS genotype 1 and 2 isolates. ITS genotype 3 and 4 isolates could not be distinguished reliably by MLST, due to the lack of sequence variation among these isolates, the majority of which to date have been recovered from the Middle East. However, results may possibly be improved with the inclusion of a larger number of ITS genotype 3 and 4 isolates from geographical locations outside of the Middle East. A possible reason for the low level of discrimination is the relatively small collection of isolates studied. However, isolates were recovered from a broad range of geographical locations and included isolates recovered from both carriage and systemic infection. Another possible reason for the low level of sequence variation and heterozygosity is the lack of divergence within the population of *C. dubliniensis* strains. Results of genotypic diversity and linkage diversity analyses suggested that the sample of 50 *C. dubliniensis* isolates investigated in this study represent a clonal population. However, it is important to emphasize that the sample number was relatively small, even though many of the isolates were recovered from disparate geographic locations around the world. Furthermore, it is possible that a more diverse population of *C. dubliniensis* strains exists in nonhuman hosts and that this is not reflected in the present study (see below). Finally, the recent divergence of *C. dubliniensis* from its ancestor *C. albicans* suggests a *C. dubliniensis* population that is less divergent than that of *C. albicans*, as the latter has had more time to diverge into major and minor clades. *C. dubliniensis* is a poor pathogen in comparison to *C. albicans*, since it rarely causes infections in healthy individuals and therefore may be under less pressure to adapt to different host environments. However, it is also possible that the natural host/reservoir for *C. dubliniensis* is not humans. *Candida* species have also been recovered from nonhuman hosts, such as dogs, cats, birds, and chameleons (12, 14, 41, 52). Recent environmental studies have reported the recovery of ITS genotype 1 *C. dubliniensis* isolates from *Ixodes uriae* ticks on the Great Saltee Island off southeastern Ireland (34), the supposed source being bird excrement. *Candida* species such as *C. albicans*, *C. guilliermondii*, and *C. tropicalis* have been recovered previously from the gastrointestinal tracts and cloacae of birds (12, 14, 23). It is conceivable that humans represent only a minority host for *C. dubliniensis* and that the isolates studied to date represent a relatively small subpopulation of the species. In order to investigate this possibility further, it will be necessary to increase the number of isolates analyzed by MLST, including isolates from avian and possibly other nonhuman hosts. We anticipate that such additional studies will identify additional MLST clades. These studies are currently in progress. Further studies will also include investigating the levels of recombination events in *C. dubliniensis*, as Odds et al. have recently shown that a high frequency of recombination is apparent in *C. albicans* (35).

ACKNOWLEDGMENTS

This study was supported by the Microbiology Research Unit, Dublin Dental School & Hospital. D.D. is the recipient of a Ph.D. fellowship from the Institut National de la Recherche Agronomique.

REFERENCES

1. **Agapow, P. M., and A. Burt.** 2001. Indices of multilocus linkage disequilibrium. Mol. Ecol. Notes **1**:101–102.
2. **Al Mosaid, A., D. Sullivan, I. F. Salkin, D. Shanley, and D. C. Coleman.** 2001. Differentiation of *Candida dubliniensis* from *Candida albicans* on Staib agar and caffeic acid-ferric citrate agar. J. Clin. Microbiol. **39**:323–327.
3. **Al Mosaid, A., D. J. Sullivan, and D. C. Coleman.** 2003. Differentiation of *Candida dubliniensis* from *Candida albicans* on Pal's agar. J. Clin. Microbiol. **41**:4787–4789.
4. **Al Mosaid, A., D. J. Sullivan, I. Polacheck, F. A. Shaheen, O. Soliman, S. Al Hedaithy, S. Al Thawad, M. Kabadaya, and D. C. Coleman.** 2005. Novel 5-flucytosine-resistant clade of *Candida dubliniensis* from Saudi Arabia and Egypt identified by Cd25 fingerprinting. J. Clin. Microbiol. **43**:4026–4036.

5. **Bain, J. M., A. Tavanti, A. D. Davidson, M. D. Jacobsen, D. Shaw, N. A. Gow, and F. C. Odds.** 2007. Multilocus sequence typing of the pathogenic fungus *Aspergillus fumigatus*. J. Clin. Microbiol. **45:**1469–1477.
6. **Blignaut, E., C. Pujol, S. Lockhart, S. Joly, and D. R. Soll.** 2002. Ca3 fingerprinting of *Candida albicans* isolates from human immunodeficiency virus-positive and healthy individuals reveals a new clade in South Africa. J. Clin. Microbiol. **40:**826–836.
7. **Bougnoux, M. E., D. M. Aanensen, S. Morand, M. Theraud, B. G. Spratt, and C. d'Enfert.** 2004. Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. Infect. Genet. Evol. **4:**243–252.
8. **Bougnoux, M. E., D. Diogo, N. Francois, B. Sendid, S. Veirmeire, J. F. Colombel, C. Bouchier, H. Van Kruiningen, C. d'Enfert, and D. Poulain.** 2006. Multilocus sequence typing reveals intrafamilial transmission and microevolutions of *Candida albicans* isolates from the human digestive tract. J. Clin. Microbiol. **44:**1810–1820.
9. **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.
10. **Bougnoux, M. E., A. Tavanti, C. Bouchier, N. A. Gow, A. Magnier, A. D. Davidson, M. C. Maiden, C. D'Enfert, and F. C. Odds.** 2003. Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. J. Clin. Microbiol. **41:**5265–5266.
11. **Boyle, B. M., D. J. Sullivan, C. Forkin, F. Mulcahy, C. T. Keane, and D. C. Coleman.** 2002. *Candida dubliniensis* candidaemia in an HIV-positive patient in Ireland. Int. J. STD AIDS **13:**55–57.
12. **Buck, J. D.** 1990. Isolation of *Candida albicans* and halophilic *Vibrio* spp. from aquatic birds in Connecticut and Florida. Appl. Environ. Microbiol. **56:**826–828.
13. **Bujdakova, H., S. Melkusova, I. Soji, M. Mokras, and Y. Mikami.** 2004. Discrimination between *Candida albicans* and *Candida dubliniensis* isolated from HIV-positive patients by using commercial method in comparison with PCR assay. Folia Microbiol. (Prague) **49:**484–490.
14. **Cafarchia, C., A. Camarda, D. Romito, M. Campolo, N. C. Quaglia, D. Tullio, and D. Otranto.** 2006. Occurrence of yeasts in cloacae of migratory birds. Mycopathologia **161:**229–234.
15. **Chan-Tack, K. M.** 2005. Fatal *Candida dubliniensis* septicemia in a patient with AIDS. Clin. Infect. Dis. **40:**1209–1210.
16. **Chen, K. W., Y. C. Chen, H. J. Lo, F. C. Odds, T. H. Wang, C. Y. Lin, and S. Y. Li.** 2006. Multilocus sequence typing for analyses of clonality of *Candida albicans* strains in Taiwan. J. Clin. Microbiol. **44:**2172–2178.
17. **Dodgson, A. R., C. Pujol, D. W. Denning, D. R. Soll, and A. J. Fox.** 2003. Multilocus sequence typing of *Candida glabrata* reveals geographically enriched clades. J. Clin. Microbiol. **41:**5709–5717.
18. **Donnelly, S. M., D. J. Sullivan, D. B. Shanley, and D. C. Coleman.** 1999. Phylogenetic analysis and rapid identification of *Candida dubliniensis* based on analysis of *ACT1* intron and exon sequences. Microbiology **145:**1871–1882.
19. **Fell, J. W.** 1993. Rapid identification of yeast species using three primers in a polymerase chain reaction. Mol. Mar. Biol. Biotechnol. **2:**174–180.
20. **Gallagher, P. J., D. E. Bennett, M. C. Henman, R. J. Russell, S. R. Flint, D. B. Shanley, and D. C. Coleman.** 1992. Reduced azole susceptibility of oral isolates of *Candida albicans* from HIV-positive patients and a derivative exhibiting colony morphology variation. J. Gen. Microbiol. **138:**1901–1911.
21. **Gee, S. F., S. Joly, D. R. Soll, J. F. Meis, P. E. Verweij, I. Polacheck, D. J. Sullivan, and D. C. Coleman.** 2002. Identification of four distinct genotypes of Candida dubliniensis and detection of microevolution in vitro and in vivo. J. Clin. Microbiol. **40:**556–574.
22. **Gilfillan, G. D., D. J. Sullivan, K. Haynes, T. Parkinson, D. C. Coleman, and N. A. Gow.** 1998. *Candida dubliniensis*: phylogeny and putative virulence factors. Microbiology **144:**829–838.
23. **Hasenclever, H. F., and R. M. Kogan.** 1975. *Candida albicans* associated with the gastrointestinal tract of the common pigeon (*Columbia livia*). Sabouraudia **13:**116–120.
24. **Hunter, P. R.** 1990. Reproducibility and indices of discriminatory power of microbial typing methods. J. Clin. Microbiol. **28:**1903–1905.
25. **Jacobsen, M. D., N. A. Gow, M. C. Maiden, D. J. Shaw, and F. C. Odds.** 2007. Strain typing and determination of population structure of *Candida krusei* by multilocus sequence typing. J. Clin. Microbiol. **45:**317–323.
26. **Joly, S., C. Pujol, M. Rysz, K. Vargas, and D. R. Soll.** 1999. Development and characterization of complex DNA fingerprinting probes for the infectious yeast *Candida dubliniensis*. J. Clin. Microbiol. **37:**1035–1044.
27. **Kibbler, C. C., S. Seaton, R. A. Barnes, W. R. Gransden, R. E. Holliman, E. M. Johnson, J. D. Perry, D. J. Sullivan, and J. A. Wilson.** 2003. Management and outcome of bloodstream infections due to *Candida* species in England and Wales. J. Hosp. Infect. **54:**18–24.
28. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 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.
29. **Marriott, D., M. Laxton, and J. Harkness.** 2001. *Candida dubliniensis* candidemia in Australia. Emerg. Infect. Dis. **7:**479.
30. **McMullan, R., J. Xu, J. E. Moore, B. C. Millar, M. J. Walker, S. T. Irwin, J. Price, J. Barr, and S. Hedderwick.** 2002. *Candida dubliniensis* bloodstream infection in patients with gynaecological malignancy. Eur. J. Clin. Microbiol. Infect. Dis. **21:**635–636.
31. **Meis, J. F., M. Ruhnke, B. E. De Pauw, F. C. Odds, W. Siegert, and P. E. Verweij.** 1999. *Candida dubliniensis* candidemia in patients with chemotherapy-induced neutropenia and bone marrow transplantation. Emerg. Infect. Dis. **5:**150–153.
32. **Melkusova, S., M. Lisalova, P. Pavlik, and H. Bujdakova.** 2005. The first clinical isolates of *Candida dubliniensis* in Slovakia. Mycopathologia **159:**369–371.
33. **Moran, G., C. Stokes, S. Thewes, B. Hube, D. C. Coleman, and D. Sullivan.** 2004. Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence-associated genes in *Candida dubliniensis*. Microbiology **150:**3363–3382.
34. **Nunn, M. A.** 2007. Environmental source of *Candida dubliniensis*. Emerg. Infect. Dis. **13:**747–750.
35. **Odds, F. C., M. E. Bougnoux, D. J. Shaw, J. M. Bain, A. D. Davidson, D. Diogo, M. D. Jacobsen, M. Lecomte, S. Y. Li, A. Tavanti, M. C. Maiden, N. A. Gow, and C. d'Enfert.** 2007. Molecular phylogenetics of *Candida albicans*. Eukaryot. Cell **6:**1041–1052.
36. **Odds, F. C., A. D. Davidson, M. D. Jacobsen, A. Tavanti, J. A. Whyte, C. C. Kibbler, D. H. Ellis, M. C. Maiden, D. J. Shaw, and N. A. Gow.** 2006. *Candida albicans* strain maintenance, replacement, and microvariation demonstrated by multilocus sequence typing. J. Clin. Microbiol. **44:**3647–3658.
37. **Pincus, D. H., D. C. Coleman, W. R. Pruitt, A. A. Padhye, I. F. Salkin, M. Geimer, A. Bassel, D. J. Sullivan, M. Clarke, and V. Hearn.** 1999. Rapid identification of *Candida dubliniensis* with commercial yeast identification systems. J. Clin. Microbiol. **37:**3533–3539.
38. **Pinjon, E., D. Sullivan, I. Salkin, D. Shanley, and D. Coleman.** 1998. Simple, inexpensive, reliable method for differentiation of *Candida dubliniensis* from *Candida albicans*. J. Clin. Microbiol. **36:**2093–2095.
39. **Polacheck, I., J. Strahilevitz, D. Sullivan, S. Donnelly, I. F. Salkin, and D. C. Coleman.** 2000. Recovery of *Candida dubliniensis* from non-human immunodeficiency virus-infected patients in Israel. J. Clin. Microbiol. **38:**170–174.
40. **Ponton, J., R. Ruchel, K. V. Clemons, D. C. Coleman, R. Grillot, J. Guarro, D. Aldebert, P. Ambroise-Thomas, J. Cano, A. J. Carrillo-Munoz, J. Gene, C. Pinel, D. A. Stevens, and D. J. Sullivan.** 2000. Emerging pathogens. Med. Mycol. **38**(Suppl. 1)**:**225–236.
41. **Pressler, B. M., S. L. Vaden, I. F. Lane, L. D. Cowgill, and J. A. Dye.** 2003. *Candida* spp. urinary tract infections in 13 dogs and seven cats: predisposing factors, treatment, and outcome. J. Am. Anim. Hosp. Assoc. **39:**263–270.
42. **Pujol, C., M. Pfaller, and D. R. Soll.** 2002. Ca3 fingerprinting of *Candida albicans* bloodstream isolates from the United States, Canada, South America, and Europe reveals a European clade. J. Clin. Microbiol. **40:**2729–2740.
43. **Robles, J. C., L. Koreen, S. Park, and D. S. Perlin.** 2004. Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of *Candida albicans*. J. Clin. Microbiol. **42:**2480–2488.
44. **Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90:**4384–4388.
45. **Soll, D. R.** 2000. The ins and outs of DNA fingerprinting the infectious fungi. Clin. Microbiol. Rev. **13:**332–370.
46. **Soll, D. R., and C. Pujol.** 2003. *Candida albicans* clades. FEMS Immunol. Med. Microbiol. **39:**1–7.
47. **Stokes, C., G. P. Moran, M. J. Spiering, G. T. Cole, D. C. Coleman, and D. J. Sullivan.** 2007. Lower filamentation rates of *Candida dubliniensis* contribute to its lower virulence in comparison with *Candida albicans*. Fungal Genet. Biol. **44:**920–931.
48. **Sullivan, D., K. Haynes, J. Bille, P. Boerlin, L. Rodero, S. Lloyd, M. Henman, and D. Coleman.** 1997. Widespread geographic distribution of oral *Candida dubliniensis* strains in human immunodeficiency virus-infected individuals. J. Clin. Microbiol. **35:**960–964.
49. **Sullivan, D. J., G. P. Moran, and D. C. Coleman.** 2005. *Candida dubliniensis*: ten years on. FEMS Microbiol. Lett. **253:**9–17.
50. **Sullivan, D. J., G. P. Moran, E. Pinjon, A. Al-Mosaid, C. Stokes, C. Vaughan, and D. C. Coleman.** 2004. Comparison of the epidemiology, drug resistance mechanisms, and virulence of *Candida dubliniensis* and *Candida albicans*. FEMS Yeast Res. **4:**369–376.
51. **Sullivan, D. J., T. J. Westerneng, K. A. Haynes, D. E. Bennett, and D. C. Coleman.** 1995. *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. Microbiology **141:**1507–1521.
52. **Tavanti, A., A. D. Davidson, M. J. Fordyce, N. A. Gow, M. C. Maiden, and F. C. Odds.** 2005. Population structure and properties of *Candida albicans*, as determined by multilocus sequence typing. J. Clin. Microbiol. **43:**5601–5613.
53. **Tavanti, A., A. D. Davidson, E. M. Johnson, M. C. Maiden, D. J. Shaw, N. A. Gow, and F. C. Odds.** 2005. Multilocus sequence typing for differentiation of strains of *Candida tropicalis*. J. Clin. Microbiol. **43:**5593–5600.
54. **Tavanti, A., N. A. Gow, S. Senesi, M. C. Maiden, and F. C. Odds.** 2003.

Optimization and validation of multilocus sequence typing for *Candida albicans*. J. Clin. Microbiol. **41:**3765–3776.

55. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
56. **Vilela, M. M., K. Kamei, A. Sano, R. Tanaka, J. Uno, I. Takahashi, J. Ito, K. Yarita, and M. Miyaji.** 2002. Pathogenicity and virulence of *Candida dubliniensis*: comparison with *C. albicans*. Med. Mycol. **40:**249–257.
57. **White, T. J., T. D. Bruns, S. B. Lee, and J. W. Taylor.** 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p. 315–322. In M. A. Innis, D. H. Gelfand, J. J. Sininsky, and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, Inc., San Diego, CA.

Genetic Differences between Avian and Human Isolates of *Candida dubliniensis*

Brenda A. McManus, Derek J. Sullivan, Gary P. Moran, Christophe d'Enfert, Marie-Elisabeth Bougnoux, Miles A. Nunn, and David C. Coleman

When *Candida dubliniensis* isolates obtained from seabird excrement and from humans in Ireland were compared by using multilocus sequence typing, 13 of 14 avian isolates were genetically distinct from human isolates. The remaining avian isolate was indistinguishable from a human isolate, suggesting that transmission may occur between humans and birds.

Candida dubliniensis is an opportunistic yeast species phenotypically and genetically closely related to *C. albicans*, the most common cause of *Candida* infection. However, *C. dubliniensis* is less pathogenic and is most commonly associated with superficial infection in immunocompromised persons. Although *C. albicans* has frequently been isolated from avian and animal sources (*1–4*), the recent study by Nunn et al. identified *C. dubliniensis* from a nonhuman source (*5*). These isolates were obtained from the surface of *Ixodes uriae* ticks that lived in cracks filled with seabird excrement at 2 locations at a seabird breeding colony on Great Saltee Island off the southeastern coast of Ireland.

Multilocus sequence typing (MLST) is an informative tool for investigating the population structure and epidemiology of many bacterial and fungal species (*6*). We have used MLST to show that *C. dubliniensis* has less genetic diversity than *C. albicans* and that *C. dubliniensis* isolates comprise 3 distinct clades (C1, C2, and C3), which correspond to described internally transcribed spacer (ITS) region genotypes 1–4 (*7*). Two other research groups recently used MLST to show genetic differences between *C. albicans* isolates from humans and animals (*3*,*4*). The purpose of our study was to use MLST, the presence or absence of a previously identified polymorphism in the *CDR1* gene (*8*), and mating type analysis to determine genetic relatedness between avian-associated and human *C. dubliniensis* isolates and whether avian-associated isolates are a source of human opportunistic infections.

Author affiliations: Dublin Dental School and Hospital, Dublin, Ireland (B.A. McManus, D.J. Sullivan, G.P. Moran, D.C. Coleman); Trinity College Dublin, Dublin (B.A. McManus, D.J. Sullivan, G.P. Moran, D.C. Coleman); Institut Pasteur, Paris, France (C. d'Enfert, M.-E. Bougnoux); and National Environmental Research Council Centre for Ecology and Hydrology, Oxford, UK (M.A. Nunn)

DOI: 10.3201/eid1509.081660

The Study

To obtain avian-associated *C. dubliniensis* isolates from a novel geographic site, fresh seabird excrement was sampled from the campus of Trinity College Dublin, ≈150 km north of Great Saltee Island by using nitrogen-gassed VI-PAK sterile swabs (Sarstedt-Drinagh, Wexford, Ireland). Samples were plated within 2 h of collection on CHROMagar *Candida* medium (CHROMagar, Paris, France), incubated at 30°C for 48 h, and identified as described (*7*,*9–12*). Three new *C. dubliniensis* isolates were obtained from 134 fecal samples. Like isolates from Great Saltee Island (*5*), these 3 isolates obtained directly from freshly deposited herring gull (*Larus argentatus*) excrement were ITS genotype 1 (*13*). Because the isolates originally described by Nunn et al. (*5*) were obtained from the surface of ticks living in avian excrement, avian-associated isolates refers to avian excrement–associated isolates. The avian-associated isolates were compared with 31 human *C. dubliniensis* strains belonging to MLST clade C1 as previously reported (*7*), and 5 additional *C. dubliniensis* clade C1 human isolates from Ireland (Table).

Isolates were assigned a diploid sequence type (DST) on the basis of genotype numbers for the 8 loci in the recommended *C. dubliniensis* MLST typing scheme (*7*) (Table). Six new DSTs were identified in 13 of 14 avian-associated isolates because of the identification of 2 new *exZWF1b* alleles that were found exclusively in avian- associated isolates. DST2 was the only previously identified DST (isolate AV7, Table). DST 31 was the most frequently (5/14 isolates) found DST in avian-associated *C. dubliniensis* isolates, all 5 of which were from Great Saltee Island (*5*). Four isolates belonged to DST 27, three from Great Saltee Island and 1 from Dublin (Table).

Polymorphic sites (n = 36) from the 8 MLST loci (*7*) of all 50 clade C1 human and avian-associated *C. dubliniensis* isolates were concatenated and used to construct a neighbor-joining tree (MEGA software program version 3.1 [*14*]), which included all known clade C1 DSTs identified. Thirteen of 14 avian-associated *C. dubliniensis* isolates, 11 from Great Saltee Island (*5*) and 2 from Dublin, formed a distinct subgroup within clade C1 (Figure, panel A). This same subgroup was also identified in trees generated by using the unweighted pair group method with arithmetic mean, maximum parsimony, and maximum likelihood, and based on related sequence types (BURST) analysis. To test

Table. Newly investigated avian-associated and human isolates of *Candida dubliniensis*, Ireland*

Isolate	Source	Year of isolation	Location	DST†	Mating type	TAG	Reference
SL411	*Ixodes uriae* ticks	2007	GSI	27	aa	+	(*5*)
SL422	*I. uriae*	2007	GSI	27	aa	+	(*5*)
SL370	*I. uriae*	2007	GSI	27	aa	+	(*5*)
SL410	*I. uriae*	2007	GSI	29	aa	+	(*5*)
SL375-I	*I. uriae*	2007	GSI	31	aa	+	(*5*)
SL375-II	*I. uriae*	2007	GSI	31	aa	+	(*5*)
SL397	*I. uriae*	2007	GSI	31	aa	+	(*5*)
SL414	*I. uriae*	2007	GSI	31	aa	+	(*5*)
SL495	*I. uriae*	2007	GSI	33	aa	+	(*5*)
SL509	*I. uriae*	2007	GSI	30	aa	+	(*5*)
SL522	*I. uriae*	2007	GSI	31	aa	+	(*5*)
AV5	*Larus argentatus*‡	2008	TCD	29	aa	+	This study
AV6	*L. argentatus*	2008	TCD	27	aa	+	This study
AV7	*L. argentatus*	2008	TCD	2	aα	+	This study
CD06032	Human, oral	2006	Ireland	36	αα	–	This study
CD06027	Human, oral	2006	Ireland	1	aα	+	This study
CD0512	Human, oral	2005	Ireland	37	aα	–	This study
CD524	Human, oral	1997	Ireland	35	aα	–	(*13*)
CD505	Human, oral	1989	Ireland	28	αα	+	(*13*)

*All 19 isolates were internal transcribed spacer genotype 1 and belonged to multilocus sequence typing clade C1. DST, diploid sequence type; TAG, TAG polymorphism; GSI, Great Saltee Island off the southeastern coast of Ireland; TCD, Trinity College Dublin.
†Assigned to each isolate according the recommended multilocus sequence typing scheme for *C. dubliniensis* (scheme D) (*7*). All DSTs, except for DST1 and DST 2, are new.
‡Herring gull.

for genetic separation between human and avian-associated isolates obtained from the same country, a neighbor-joining tree was constructed by using 13 avian-associated and human clade C1 isolates from Ireland, each of which represented unique DSTs. The tree showed the robustness of the avian-associated subgroup within a population of human isolates from the same country, and the distribution of avian-associated and human isolates differed significantly ($p = 0.025$, by Fisher exact test) (www.exactoid.com/fisher/index.php) (Figure, panel B).

The prevalence of a common point mutation, previously identified in the *CDR1* gene of some ITS genotype 1 isolates, was determined for avian and human *C. dubliniensis* isolates as described previously (*8*). All 14 avian-associated isolates had the TAG polymorphism (Table) compared with 19 (53%) of 36 human clade C1 isolates. The mating types of the isolates were determined by multiplex PCR amplification by using 2 pairs of mating type locus (MTL)–specific primers. A 535-bp amplimer was generated with primers MTLa1-F (5′-TGAAAATGAAGACAATGCGA-3′) and MTLa1-R (5′-CATCTTTTTCTGCTATCAATTC-3′) in the presence of MTL type a DNA, and a 615-bp product resulted from primers MTLα1-F (5′-ATGAATTCACATCTGGAGGC-3′) and MTLα1-R (5′-CTGTTAATAGCAAAGCAGCC-3′) in the presence of MTL type α DNA. Amplification reactions contained 10 pmol of each of the forward and reverse primers, 2.5 mmol/L $MgCl_2$, 10 mmol/L Tris-HCl, pH 9.0 at 25°C, 10 mmol/L KCl, 0.1% (vol/vol) Triton X-100, 1.25 U GoTaq polymerase (Promega, Madison, WI, USA), and 25 μL of template DNA in a total volume of 50 μL. Cycling conditions were at 94°C for 10 min; 30 cycles at 94°C for 1 min, 55°C for 2 min, and 72°C for 3 min; and a final step at 72°C for 10 min.

Of the 14 avian excrement–associated isolates, 13 were MTLa homozygous (a/a), (Table). Only 4 (11.1%) of 36 human clade C1 isolates were homozygous for MTLa; 28 (77.7%) of 36 were heterozygous for MTL (a/α). A previous study also reported that 17 (20.7%) of 82 human *C. dubliniensis* isolates were homozygous for MTLa (*15*). The gull isolate AV7 was indistinguishable from human isolates CD71, SA115, and CM1 by MLST and has the same mating type (a/α). We propose that AV7 may be a human isolate that colonized a gull scavenging on the Trinity College Dublin campus. The TAG polymorphism and mating type data from the avian-associated isolates suggest a highly clonal population.

Conclusions

The avian-associated *C. dubliniensis* isolates investigated belong to MLST clade C1, which includes most human isolates. However, most (13/14) of the avian-associated isolates form a distinct subgroup within this clade, which suggests that despite the low level of variation within *C. dubliniensis*, a distinct avian subpopulation may be present. This suggestion is supported by the observation that 2/3 isolates (AV5 and AV6) obtained in Dublin belonged to the same subpopulation (defined by MLST, *CDR1*, and MTL loci) as isolates obtained from Great Saltee Island, which is 150 km from Dublin. Similar data suggesting ge-

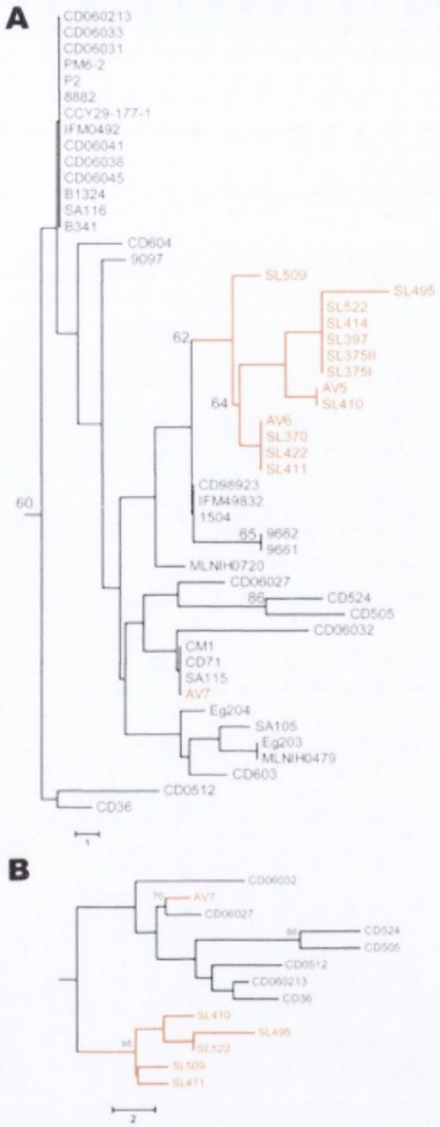


Figure. Neighbor-joining trees based on the polymorphic sites in *Candida dubliniensis* multilocus sequence typing (MLST) sequences. Bootstrap values >60% are indicated at cluster nodes. Avian-associated isolates are indicated in red. Numbers of polymorphic sites in isolates are indicated by scale bars. A) Isolates of MLST clade C1 defined by McManus et al. (*7*) showing location of avian-associated isolates in relation to human isolates in the same clade; human isolates were originally obtained in many countries. B) Neighbor-joining tree based on polymorphic sites in MLST sequences for each of 13 internal transcribed spacer genotype 1 *C. dubliniensis* isolates, 7 of which were obtained from humans in Ireland and 6 from seabird excrement in Ireland. Isolates that had identical diploid sequence types (DSTs) were not included in the tree so that only 1 of each DST is included. Tree displays the robustness of the avian-associated subgroup of isolates within a population of similar human-associated isolates from the same region. The rate of heterozygosity among human and avian-associated clade C1 isolates was 1.6 and 1 heterozygous site per DST, respectively, from 36 polymorphic sites, which indicated that avian-associated isolates were more clonal.

netic separation and differential clade distribution between human and animal populations of *C. albicans* have been reported (*3*,*4*). The presence of the avian-associated subgroup within the most predominant clade (C1), which had previously only been identified in human isolates, and the close genetic relatedness between isolates, in particular gull isolate AV7, suggests that transmission between the 2 hosts can occur. However, in this instance the most likely direction of transfer is from human to bird.

Acknowledgments

We thank Oscar Merne for assistance with collection of avian excrement samples and identification of specific species of birds at sampling sites.

This study was supported by the Microbiology Research Unit, Dublin Dental School and Hospital.

Ms McManus is a doctoral candidate in the Microbiology Research Unit at Dublin Dental School and Hospital, Trinity College Dublin. Her research interests focus on analysis of the population structure of the novel yeast pathogen *C. dubliniensis* by using MLST.

References

1. Buck JD. Isolation of *Candida albicans* and halophilic *Vibrio* spp. from aquatic birds in Connecticut and Florida. Appl Environ Microbiol. 1990;56:826–8.
2. Edelmann A, Kruger M, Schmid J. Genetic relationship between human and animal isolates of *Candida albicans*. J Clin Microbiol. 2005;43:6164–6. DOI: 10.1128/JCM.43.12.6164-6166.2005
3. Wrobel L, Whittington JK, Pujol C, Oh SH, Ruiz MO, Pfaller MA, et al. Molecular phylogenetic analysis of a geographically and temporally matched collection of *Candida albicans* isolates from humans and nonmigratory wildlife in central Illinois. Eukaryot Cell. 2008;7:1475–86. DOI: 10.1128/EC.00162-08
4. Jacobsen MD, Bougnoux ME, d'Enfert C, Odds FC. Multilocus sequence typing of *Candida albicans* isolates from animals. Res Microbiol. 2008;159:436–40. DOI: 10.1016/j.resmic.2008.05.003
5. Nunn MA, Schaefer SM, Petrou MA, Brown JR. Environmental source of *Candida dubliniensis*. Emerg Infect Dis. 2007;13:747–50.
6. Odds FC, Jacobsen MD. Multilocus sequence typing of pathogenic *Candida* species. Eukaryot Cell. 2008;7:1075–84. DOI: 10.1128/EC.00062-08
7. McManus BA, Coleman DC, Moran G, Pinjon E, Diogo D, Bougnoux ME, et al. Multilocus sequence typing reveals that the population structure of *Candida dubliniensis* is significantly less divergent than that of *Candida albicans*. J Clin Microbiol. 2008;46:652–64. DOI: 10.1128/JCM.01574-07
8. Moran G, Sullivan D, Morschhauser J, Coleman D. The *Candida dubliniensis CdCDR1* gene is not essential for fluconazole resistance. Antimicrob Agents Chemother. 2002;46:2829–41. DOI: 10.1128/AAC.46.9.2829-2841.2002
9. Coleman D, Sullivan D, Harrington B, Haynes K, Henman M, Shanley D, et al. Molecular and phenotypic analysis of *Candida dubliniensis*: a recently identified species linked with oral candidosis in HIV-infected and AIDS patients. Oral Dis. 1997;3(Suppl 1):S96–101.

10. Al Mosaid A, Sullivan DJ, Coleman DC. Differentiation of *Candida dubliniensis* from *Candida albicans* on Pal's agar. J Clin Microbiol. 2003;41:4787–9. DOI: 10.1128/JCM.41.10.4787-4789.2003
11. Pincus DH, Coleman DC, Pruitt WR, Padhye AA, Salkin IF, Geimer M, et al. Rapid identification of *Candida dubliniensis* with commercial yeast identification systems. J Clin Microbiol. 1999;37:3533–9.
12. Donnelly SM, Sullivan DJ, Shanley DB, Coleman DC. Phylogenetic analysis and rapid identification of *Candida dubliniensis* based on analysis of *ACT1* intron and exon sequences. Microbiology. 1999;145:1871–82. DOI: 10.1099/13500872-145-8-1871
13. Gee SF, Joly S, Soll DR, Meis JF, Verweij PE, Polacheck I, et al. Identification of four distinct genotypes of *Candida dubliniensis* and detection of microevolution in vitro and in vivo. J Clin Microbiol. 2002;40:556–74. DOI: 10.1128/JCM.40.2.556-574.2002
14. Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform. 2004;5:150–63. DOI: 10.1093/bib/5.2.150
15. Pujol C, Daniels KJ, Lockhart SR, Srikantha T, Radke JB, Geiger J, et al. The closely related species *Candida albicans* and *Candida dubliniensis* can mate. Eukaryot Cell. 2004;3:1015–27. DOI: 10.1128/EC.3.4.1015-1027.2004

Address for correspondence: David C. Coleman, Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental School and Hospital, Trinity College Dublin, Dublin 2, Ireland; email: david.coleman@dental.tcd.ie

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Nov. 2009, p. 4678–4685 Vol. 53, No. 11
0066-4804/09/$12.00 doi:10.1128/AAC.00607-09

A Ser29Leu Substitution in the Cytosine Deaminase Fca1p Is Responsible for Clade-Specific Flucytosine Resistance in *Candida dubliniensis*[▽]

Brenda A. McManus, Gary P. Moran, Judy A. Higgins, Derek J. Sullivan, and David C. Coleman*

Microbiology Research Unit, Division of Oral Biosciences, School of Dental Science, and Dublin Dental Hospital, University of Dublin, Trinity College Dublin, Dublin 2, Ireland

Received 5 May 2009/Returned for modification 16 July 2009/Accepted 12 August 2009

The population structure of the opportunistic yeast pathogen *Candida dubliniensis* is composed of three main multilocus sequence typing clades (clades C1 to C3), and clade C3 predominantly consists of isolates from the Middle East that exhibit high-level resistance ($MIC_{50} \geq 128$ μg/ml) to the fungicidal agent flucytosine (5FC). The close relative of *C. dubliniensis*, *C. albicans*, also exhibits clade-specific resistance to 5FC, and resistance is most commonly mediated by an Arg101Cys substitution in the *FUR1* gene encoding uracil phosphoribosyltransferase. Broth microdilution assays with fluorouracil (5FU), the toxic deaminated form of 5FC, showed that both 5FC-resistant and 5FC-susceptible *C. dubliniensis* isolates exhibited similar 5FU MICs, suggesting that the *C. dubliniensis* cytosine deaminase (Fca1p) encoded by *C. dubliniensis FCA1* (Cd*FCA1*) may play a role in mediating *C. dubliniensis* clade-specific 5FC resistance. Amino acid sequence analysis of the Cd*FCA1* open reading frame (ORF) identified a homozygous Ser29Leu substitution in all 12 5FC-resistant isolates investigated which was not present in any of the 9 5FC-susceptible isolates examined. The tetracycline-inducible expression of the Cd*FCA1* ORF from a 5FC-susceptible *C. dubliniensis* isolate in two separate 5FC-resistant clade C3 isolates restored susceptibility to 5FC, demonstrating that the Ser29Leu substitution was responsible for the clade-specific 5FC resistance and that the 5FC resistance encoded by *FCA1* genes with the Ser29Leu transition is recessive. Quantitative real-time PCR analysis showed no significant difference in Cd*FCA1* expression between 5FC-susceptible and 5FC-resistant isolates in either the presence or the absence of subinhibitory concentrations of 5FC, suggesting that the Ser29Leu substitution in the Cd*FCA1* ORF is the sole cause of 5FC resistance in clade C3 *C. dubliniensis* isolates.

Candida dubliniensis is an opportunistic yeast pathogen that was first described in 1995 in human immunodeficiency virus-infected patients in Ireland (39). Since then the organism has been shown to have a worldwide distribution and has been recovered from other groups of immunocompromised individuals and from patients with severe underlying disease (2–4, 11, 29, 30, 36–38, 44). The population structure of *C. dubliniensis* has previously been investigated by using the species-specific complex DNA fingerprinting probe Cd25 and multilocus sequence typing (MLST) (4, 11, 15, 18). Early Cd25 fingerprinting analyses demonstrated that *C. dubliniensis* consists of two fingerprinting groups, termed Cd25 group I and Cd25 group II (15). Group I isolates comprise the majority of isolates investigated from many countries around the world and are very closely related, with an average similarity coefficient value (S_{AB}) of 0.8. Group II isolates are less closely related and have an average S_{AB} value of 0.47 (15). These results were later confirmed with a larger collection of isolates by Gee et al. (11), who also showed that Cd25 group I isolates comprised a single genotype (genotype 1) on the basis of sequence analysis of the internal transcribed spacer (ITS) region of the ribosomal DNA operon. Furthermore, Cd25 group II isolates were found to belong to three ITS genotypes (genotypes 2 to 4). In 2005, a study by Al Mosaid et al. (4) identified a third Cd25 fingerprinting group, termed Cd25 group III, which exhibited an average S_{AB} value of 0.35, among *C. dubliniensis* isolates recovered exclusively in Egypt, Saudi Arabia, and Israel, all of which belonged to ITS genotypes 3 or 4. All isolates belonging to Cd25 group III examined exhibited high-level intrinsic resistance to the antifungal drug flucytosine (5FC), apart from one Israeli isolate that was 5FC susceptible. This phenotype did not occur in isolates belonging to either Cd25 group I or Cd25 group II, including isolates from Cd25 groups I and II recovered from Egypt, Saudi Arabia, and Israel (4). Recent studies that have used MLST analysis to investigate the population structure of *C. dubliniensis* revealed the presence of three distinct clades, termed clades C1 to C3 (18). All 5FC-resistant isolates belonging to Cd25 fingerprint group III were found to cluster exclusively in MLST clade C3 (18). More recently, MLST was used to show that clade C1 *C. dubliniensis* isolates recovered from avian excrement-associated samples were genetically distinct from other clade C1 isolates that were recovered from humans (19).

The closest relative of *C. dubliniensis*, *Candida albicans*, also exhibits clade-specific resistance to 5FC, with 72.7% of isolates in MLST clade C1 (Ca3 fingerprinting clade I) exhibiting reduced susceptibility to this antifungal agent (23, 32). In *C. albicans*, the 5FC resistance patterns vary among isolates and range from reduced susceptibility (MICs, 0.5 to 2 μg/ml) to intermediate resistance (MICs, 2 to 8 μg/ml) or high-level

* Corresponding author. Mailing address: Microbiology Research Unit, Division of Oral Biosciences, School of Dental Science and Dublin Dental Hospital, University of Dublin, Trinity College Dublin, Dublin 2, Ireland. Phone: 353-1-6127276. Fax: 353-1-6127295. E-mail: david.coleman@dental.tcd.ie.

▽ Published ahead of print on 24 August 2009.

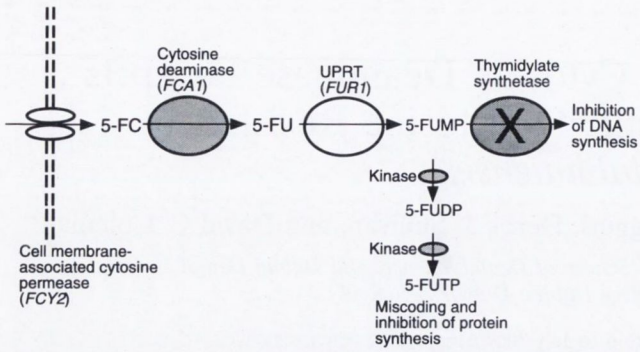


FIG. 1. Metabolic pathway and mode of action of 5FC in *Candida* yeasts. 5FC and 5FU are transported into the cell by cell membrane-associated cytosine-purine permeases. In *Candida* spp., these are encoded by two genes that display amino acid sequence homology with the *FCY2* gene of *S. cerevisiae* (13). Upon entry into the cell, 5FC is then deaminated to 5FU by Fca1p, encoded by *FCA1*. 5FU is then phosphorylated by UPRT, encoded by *FUR1*, yielding 5FUMP. 5FUMP inhibits thymidylate synthetase, which leads to thymidine depletion in the cell and which ultimately interrupts DNA synthesis. 5FUMP is also metabolized by two kinases, yielding fluorouridine diphosphate (5FUDP) and, subsequently, fluorouridine triphosphate (5FUTP), the latter of which is incorporated into RNA in the place of UTP, which leads to miscoding and the inhibition of protein synthesis.

resistance (MICs, ≥8 μg/ml); and a wide range of 5FC MICs for this drug have been reported among isolates (range, 0.06 μg/ml to ≥128 μg/ml) (7, 13). In *C. dubliniensis*, the resistance patterns are more clearly defined, with 5FC-susceptible isolates exhibiting 5FC MICs of ≤0.125 μg/ml and 5FC-resistant isolates exhibiting 5FC MICs of ≥128 μg/ml (4). To date, 5FC resistance in *C. dubliniensis* has been reported only in isolates from the Middle East, all of which that have been tested belong to MLST clade C3 (1, 4, 18, 29).

The antifungal action of 5FC relies on the intracellular conversion of 5FC to fluorouracil (5FU) by cytosine deaminases upon entry into fungal cells (Fig. 1). Cytosine deaminase (Fca1p) is encoded by *FCA1* in *C. albicans* and *C. dubliniensis* (Cd*FCA1*) (4, 9), and the *FCA1* genes in these two species are homologues of the *FCY1* gene in *Saccharomyces cerevisiae* (9) and in other *Candida* species, such as *Candida lusitaniae* (26). The absence of cytosine deaminases in mammalian cells prevents 5FC toxicity in humans, as the 5FC prodrug itself is nontoxic. After the conversion of 5FC to 5FU, the *FUR1*-encoded uracil phosphoribosyltransferase (UPRT) catalyzes the phosphorylation of 5FU to fluorouridine monophosphate (5FUMP) (Fig. 1). Two specific kinases catalyze the further phosphorylation of 5FUMP, eventually converting it to fluorouridine triphosphate. Fluorouridine triphosphate in turn gets incorporated into RNA, which causes miscoding, leading to the inhibition of fungal protein synthesis (Fig. 1). As a secondary method of inhibition, 5FUMP inhibits thymidylate synthetase (Fig. 1), leading to the depletion of dTTP and the misincorporation of dUTP into newly synthesized DNA, causing irreversible DNA damage and cell cycle arrest (14, 31, 41).

In haploid *C. lusitaniae* isolates, a missense T26C nucleotide mutation in the *FCY1* gene has been reported in four clinical isolates demonstrating 5FC resistance, although 5FC and 5FC-fluconazole cross-resistance has more commonly been attributed to defects in the purine cytosine permease-encoded *FCY2* gene in this species (10). In *C. albicans*, resistance to 5FC is mediated by a reduction in the activity of either the Fca1p encoded by *FCA1* or the UPRT encoded by *FUR1* (13, 31, 43). Two different research groups reported that in the majority of 5FC-resistant *C. albicans* isolates, resistance is associated with a homozygous single amino acid substitution, Arg101Cys, in UPRT (7, 13). However, other 5FC-resistant *C. albicans* isolates lack this substitution (13). One such isolate (5FC MIC, >64 μg/ml) was reported to contain a homozygous Gly28Asp substitution in the cytosine deaminase gene, and a Ser29Leu amino acid substitution was also observed in the same gene of another *C. albicans* isolate displaying intermediate 5FC resistance (5FC MIC, 4 μg/ml) (13). In *C. dubliniensis*, the DNA sequences of the *FUR1* genes encoding the UPRTs of four 5FC-resistant and five 5FC-susceptible isolates from the Middle East were determined previously, and while several single nucleotide polymorphisms (SNPs) were identified, no amino acid substitutions were observed between the isolates (4).

The purpose of the present study was to investigate the role of Fca1p in *C. dubliniensis* clade-specific 5FC resistance by the use of broth microdilution assays with 5FC and 5FU, analysis of DNA and amino acid sequences, and analysis of Cd*FCA1* expression. A tetracycline-inducible expression plasmid was used to incorporate the Cd*FCA1* gene from a 5FC-susceptible isolate (hereafter called Cd*FCA1*^{s}) into the *ADH1* locus of a 5FC-resistant isolate and the Cd*FCA1* gene from a 5FC-resistant isolate (hereafter called Cd*FCA1*^{r}) into the *ADH1* locus of a 5FC-susceptible isolate. These strains were used to determine if 5FC susceptibility or resistance could be induced in isolates upon the acquisition and expression of the respective Cd*FCA1* gene.

MATERIALS AND METHODS

Isolates and culture conditions. Twenty-one epidemiologically unrelated human *C. dubliniensis* isolates were included in the present study, including 9 5FC-susceptible isolates and 12 5FC-resistant isolates (Table 1), as reported previously (2, 4, 11, 30, 39). Previously, 5FC resistance in *C. dubliniensis* has ever been reported only in isolates from Saudi Arabia, Egypt, Israel, and Kuwait (1, 4, 29); and all but two of these have previously been investigated by MLST analysis and/or Cd25 fingerprint analysis and were shown to belong to *C. dubliniensis* MLST clade C3 and Cd25 fingerprint group III (4, 18). For these reasons, 20 of the 21 isolates chosen for study (12 5FC-resistant isolates and 8 5FC-susceptible isolates) were originally recovered in Egypt, Saudi Arabia, or Israel (Table 1). The 5FC-susceptible isolates belonged to MLST clade C1 or C2 and Cd25 fingerprint group I or II (4, 18). The *C. dubliniensis* type strain CD36, originally isolated from the oral cavity of a human immunodeficiency virus-infected individual in Ireland, was also included as a reference isolate because the complete genome sequence of this organism has been determined (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/). The *C. dubliniensis* isolates were routinely cultured on yeast extract-peptone-dextrose (YPD) agar, pH 5.6, at 37°C. For liquid culturing, isolates were grown in YPD broth (with the following per liter: 10 g yeast extract [Sigma-Aldrich Ireland Ltd., Wicklow, Ireland], 20 g peptone [Oxoid, Basingstoke, Hampshire, England], and 20 g D-glucose, pH 5.5) at 37°C in an orbital incubator (Gallenkamp, Leicester, United Kingdom) at 200 rpm. *Escherichia coli* strain DH5α (12) [F$^-$ ϕ80d*lacZ*Δm15 Δ(*lacZYA-argF*)*U169 endA1 recA1 hsdR17*(r_K^- m_K^+) *supE44 thi-1 d*$^-$ *gyrA96 relA1*] was routinely grown on Luria-Bertani agar, pH 7.4, at 37°C or in Luria-Bertani broth, pH 7.4, at 37°C with shaking at 200 rpm for liquid culture.

Susceptibility testing. The MICs for 5FC and 5FU (Sigma-Aldrich) were determined by the method described in document M27-A2 of the Clinical and Laboratory Standards Institute (CLSI; formerly the National Committee for Clinical Laboratory Standards) by using RPMI 1640 medium (6). Both drugs were titrated from a concentration of 128 μg/ml to one of 0.25 μg/ml for the preliminary analysis of all 21 clinical isolates, as well as for analysis of the

TABLE 1. *Candida dubliniensis* isolates used to investigate the molecular mechanism of 5FC resistance and their susceptibilities to 5FC and 5FU[a]

Isolate	Country of origin	Yr of isolation	Sample	Cd25 fingerprint group	ITS genotype	MIC_{50} (μg/ml)		Reference
						5FC	5FU	
CD36[b]	Ireland	1988	Oral	I	1	≤0.25	32	4, 11, 39
SA101	S. Arabia	2002	Oral	I	1	≤0.25	16	4
SA105	S. Arabia	2002	Oral	I	1	≤0.25	32	4
SA115	S. Arabia	2002	Oral	I	1	≤0.25	32	4
Eg203	Egypt	2002	Oral	I	1	≤0.25	16	4
Eg206	Egypt	2002	Oral	I	1	≤0.25	32	4
p7276	Israel	1999	RT	II	3	≤0.25	8	4, 11
p6785	Israel	1999	Urine	II	3	≤0.25	16	4, 11, 30
p7718	Israel	1999	Wound	III	4	≤0.25	16	4, 11
Eg200	Egypt	2002	Oral	III	4	≥128	8	4
Eg201	Egypt	2002	Oral	III	4	≥128	32	4
Eg202	Egypt	2002	Oral	III	4	≥128	32	4
Eg207	Egypt	2002	Oral	III	4	≥128	32	4
SA100	S. Arabia	2002	Oral	III	3	≥128	32	4
SA103	S. Arabia	2002	BAL	III	3	≥128	32	4
SA107	S. Arabia	2002	Oral	III	3	≥128	32	4
SA108	S. Arabia	2002	Oral	III	3	≥128	32	4
SA109	S. Arabia	2002	Oral	III	3	≥128	8	4
SA113	S. Arabia	2002	Oral	III	4	≥128	32	4
SA118	S. Arabia	2002	Oral	III	3	≥128	32	4
SA121	S. Arabia	2002	Oral	III	4	≥128	32	4

[a] Abbreviations: S. Arabia, Saudi Arabia; RT, respiratory tract; BAL, broncheoalveolar lavage fluid.
[b] *C. dubliniensis* type strain.

recipient isolates and transformant derivatives from cloning experiments. These were determined as the lowest concentrations of the drug that reduced the turbidity by 50% relative to the turbidity of the growth of the drug-free controls. All isolates were tested in duplicate and on two separate occasions.

DNA extraction. Isolates were grown overnight in 5 ml of YPD broth as described above. Cells were harvested from 1.5 ml of culture by centrifugation at 14,000 × *g*, and the DNA was extracted from the resulting pellet by using a DNeasy blood and tissue kit, according to the manufacturer's instructions (Qiagen Science, Crawley, West Sussex, United Kingdom) and resuspended in a final volume of 200 μl. Nucleic acids were ethanol precipitated and resuspended in 50 μl of molecular-grade Milli-Q Biocel-purified water (resistivity, 18.2 MΩ/cm; Millipore, Carrigtwohill, Cork, Ireland).

PCR amplification, sequencing, and sequence analysis of Cd*FCA1*. The complete open reading frame (ORF) of the *C. dubliniensis* Cd*FCA1* gene was amplified from 12 5FC-resistant isolates (Table 1) and 9 5FC-susceptible isolates (Table 1) by using oligonucleotide primers FCA1F and FCA1R, which incorporated SalI and BglII restriction endonuclease recognition sites, respectively (Table 2). The reaction mixtures contained 100 ng of purified template DNA, 1× Expand high-fidelity buffer, 2.5 U of Expand high-fidelity PCR system enzyme mixture (Roche Applied Science, Mannheim, Germany), 0.2 mM concentrations of each deoxynucleoside triphosphate (Promega Corporation, Madison, WI), and 0.2 μM concentrations of each oligonucleotide (Sigma Genosys Biotechnologies Europe Ltd., Pampisford, Cambridgeshire, United Kingdom). The reaction mixtures underwent an initial denaturation step of 94°C for 10 min, followed by 35 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 1 min and a final elongation step of 72°C for 7 min. The 550-bp products were purified with a GenElute PCR cleanup kit (Sigma-Aldrich), and both strands were sequenced by using the same primers that had been used for amplification. DNA sequencing reactions were performed commercially by Cogenics (Essex, United Kingdom) with an ABI 3730xl DNA analyzer. Multiple DNA and amino acid sequence alignments of Cd*FCA1* genes and their encoded proteins from 5FC-susceptible and 5FC-resistant *C. dubliniensis* isolates were carried out by using the CLUSTAL W sequence alignment computer program (40), available at the EMBL-EBI website (http://www.ebi.ac.uk/).

Tetracycline-inducible Cd*FCA1* expression in *C. dubliniensis*. The tetracycline-inducible gene expression plasmid pNIM1, developed by Park and Morschhauser (27), was adapted to investigate the inducible expression of the *C. dubliniensis* Cd*FCA1*[s] and Cd*FCA1*[r] genes in both 5FC-susceptible and 5FC-resistant *C. dubliniensis* isolates. The pNIM1 cassette was originally designed to integrate into the *C. albicans* alcohol dehydrogenase encoding gene *ADH1*. The DNA

TABLE 2. Oligonucleotide primers used in this study

Oligonucleotide	Sequence (5′–3′)[b]	Function
ADH1F	ATGCAAGCAAGCTTATTCA	PCR screening of transformants[a]
cartTA	CGGCATACTATCAGTAGTAG	
SAT	CAATGCCGCCGAGAGTAAAG	
ADH1R	CCCAAGATCTTACCTTCTTCCATT	
FCA1F	GACGCGTCGACGATATCAACGATGACATTT	Cd*FCA1* cloning and PCR screening
FCA1R	CGGGATCCAGATCTTTATTCTCCAATATCTTC	
RTFCA1F	AAACGCAGGAAGATTGCCAG	Gene expression analysis
RTFCA1R	TGGCCCCTGTACACATACTACATG	
RTACT1F	AGCTCCAGAAGCTTTGTTCAGACC	
RTACT1R	TGCATACGTTCAGCAATACCTGGG	

[a] The regions of the pNIM1-Cd*FCA1* cassette and background *ADH1* locus amplified during PCR screening of transformants are displayed in Fig. 2.
[b] The SalI and BglII restriction endonuclease recognition sites incorporated into the FCA1F and FCA1R primer sequences, respectively, are underlined.

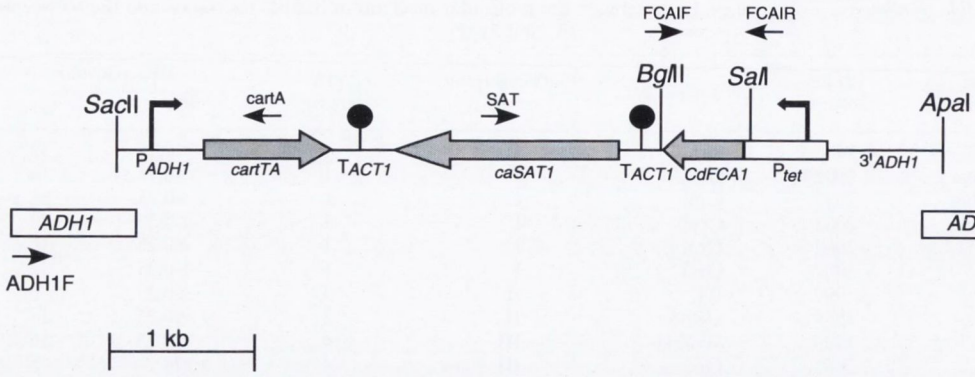


FIG. 2. Structure and *ADH1* integration site of the pNIM1-Cd*FCA1* cassette used in the tetracycline-inducible expression transformation studies. The restriction sites used for the excision of the *C. albicans GFP1* gene (replaced by the Cd*FCA1* gene) and the excision of the entire pNIM1-Cd*FCA1* cassette from the pNIM1 plasmid (27) are indicated. Transcription start sites and the directions of transcription are displayed by right-angled arrows. Terminator sequences are displayed as black hairpin loops and function in the termination of transcription of the *Candida*-adapted reverse tetracycline-dependent transactivator (*cartTA*) and the Cd*FCA1* target gene. The primers used in the PCR screening of the pNIM1-Cd*FCA1* transformants are indicated by labeled arrows and are listed in Table 2.

sequence for the *C. albicans ADH1* gene (GenBank accession no. CaO19.3997) was used in a BLAST search against the *C. dubliniensis* genome sequence database (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/) in order to identify a homologue in *C. dubliniensis*. A high level of sequence homology (94%) is shared by the *ADH1* ORFs of *C. albicans* and *C. dubliniensis*. The tetracycline-inducible promoter P_{tet}, included in pNIM1 (27), was used to drive expression of the Cd*FCA1*^{s} and Cd*FCA1*^{r} genes individually in the *ADH1* locus of a *C. dubliniensis* isolate with the opposite 5FC phenotype. The complete coding regions of the Cd*FCA1* gene from 5FC-susceptible *C. dubliniensis* isolate p7276 and 5FC-resistant *C. dubliniensis* isolates SA113 and SA109 were amplified from genomic DNA with the FCA1F-FCA1R primer pair (Table 2). The amplimers were ligated into pGEM T-Easy vector I (Promega) vector DNA and transformed into *E. coli* strain DH5α. Plasmids were recovered from the transformants by using a GenElute plasmid miniprep kit (Sigma-Aldrich), and the cloned DNA was sequenced. The complete Cd*FCA1* ORF was digested from the pGEM T-Easy plasmids by using the SalI and BglII restriction endonuclease recognition sites, which were introduced upstream and downstream of the ORF, respectively, and the FCA1F-FCA1R primer pair. The Cd*FCA1* fragments were then gel purified with a Wizard SV Gel and PCR cleanup system (Promega) and separately cloned between P_{tet} and T_{ACT1} in SalI-BglII-digested pNIM1, incorporating the Cd*FCA1* ORFs into the pNIM1 cassette in the place of the *GFP1* gene (27), (Fig. 2). The resulting plasmids, pNIM1-Cd*FCA1*^{s} and pNIM1-Cd*FCA1*^{r}, were transformed into *E. coli* strain DH5α for replication of the plasmid, prior to purification with the GenElute plasmid miniprep kit (Sigma-Aldrich), SacII-ApaI linearization of the cassette, gel purification, and transformation into *C. dubliniensis* as described previously (35). 5FC-resistant isolates SA113 and SA109 were transformed with pNIM1-Cd*FCA1*^{s}, and 5FC-susceptible isolate p7276 was transformed with pNIM1-Cd*FCA1*^{r} (Table 1). All three of these *C. dubliniensis* isolates were also transformed with the pNIM1 cassette containing the *GFP1* gene instead of the Cd*FCA1* gene as a control for the disruption of the *ADH1* locus. In order to confirm the correct integration of the complete pNIM1-Cd*FCA1* cassette into the *ADH1* locus in transformant derivatives, a number of PCR amplifications were carried out. The primer pairs used in these PCR amplifications were ADH1F and cartTA, SAT and FCA1R, FCA1F and ADH1R, and SAT and ADH1R (Table 2). These stepwise amplifications revealed the presence and the correct integration of the full pNIM1-Cd*FCA1* cassette in transformant derivatives. This was further confirmed by Southern hybridization analysis with two separate probes: a Cd*FCA1*-directed probe and a pNIM1-directed probe which was directed toward the *cartTA* transactivator region and the background *ADH1* locus. The probes were labeled with digoxigenin (DIG) by using a DIG DNA labeling and detecting kit (Roche). The prehybridization, hybridization, wash, and detection steps were carried out according to the manufacturer's instructions.

Cd*FCA1* expression analysis. To monitor the relative gene expression of the Cd*FCA1* gene in 5FC-resistant wild-type *C. dubliniensis* isolates (isolates SA113, SA109, and Eg202) and 5FC-susceptible wild-type *C. dubliniensis* isolates (isolates Eg204 and p7276), as well as in the doxycycline (DOX)-inducible transformant derivatives (strains SA113T1, SA109T1, and SA109T2), quantitative real-time PCR was carried out according to standard protocols. In brief, RNA was extracted from isolates and transformant derivatives that were grown in YPD broth in the presence of DOX (15 μg/ml) only or YPD broth in the presence of DOX (15 μg/ml) and subinhibitory concentrations of 5FC (6.4 ng/ml). The RNAs were extracted with an RNeasy minikit (Qiagen) and were treated with DNeasy (Ambion; Applied Biosystems, Warrington, United Kingdom), according to the manufacturer's instructions. The RNA samples were then reverse transcribed to cDNA by using a Superscript II reverse transcriptase kit (Invitrogen, Biosciences Ltd., Dun Laoghaire, Dublin, Ireland). Quantitative real-time PCR was carried out with two pairs of reverse transcription-PCR primers; one pair amplified the Cd*FCA1* gene, and the second pair amplified the *ACT1* gene as an internal expression control (Table 2). The comparative amplification efficiencies of these primers were assessed, prior to reverse transcription-PCR, by using primer amplification efficiency plot analysis, as described previously (28). Quantitative real-time PCRs were carried out with 0.3 μM of each primer and SYBR green master mixture in an ABI 7500 real-time PCR system (Applied Biosystems), according to the manufacturer's recommended protocols. Data analysis was carried out as described by Schmittgen and Livak (34), and the $2^{-\Delta CT}$ values were calculated from the average threshold cycle (C_T) values acquired from three replicates for both the Cd*FCA1* and the *ACT1* genes.

RESULTS

In vitro susceptibility testing. Nine *C. dubliniensis* isolates previously reported to be 5FC susceptible and 12 *C. dubliniensis* isolates previously reported to be 5FC resistant by Al Mosaid et al. (4) (Table 1) were tested for their susceptibilities to 5FC by broth microdilution assays. All nine isolates previously reported to be 5FC susceptible were confirmed as such and exhibited 5FC MICs of ≤0.25 μg/ml. Similarly, the 12 previously reported 5FC-resistant isolates exhibited 5FC MICs of ≥128 μg/ml. In an attempt to localize potential blocks or lesions in the 5FC metabolic pathway in 5FC-resistant isolates which may contribute to resistance, broth microdilution assays were also carried out with 5FU instead of 5FC (Table 1). We hypothesized that if a block in the 5FC metabolic pathway occurred at the level of Fca1p, then 5FC-resistant isolates should be susceptible to 5FU (Fig. 1). All 9 5FC-susceptible isolates and 12 5FC-resistant isolates used in the study (Table 1) were tested for their susceptibilities to 5FU by using a range of concentrations (0.25 to 128 μg/ml). All 21 isolates exhibited 5FU MICs in the range of 8 to 32 μg/ml (Table 1). There was no correlation between the 5FU MIC and susceptibility to

5FC. These findings indicated that a block(s) in the 5FC metabolic pathway occurred in resistant isolates at the level of cytosine deaminase or upstream of this enzyme (Fig. 1).

Sequence analysis of Cd*FCA1* in *C. dubliniensis*. The DNA sequence of the *C. albicans FCA1* gene, which encodes Fca1p (GenBank accession no. U55194), was used in a BLAST search against the *C. dubliniensis* genome sequence (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/) in order to identify its homologue in *C. dubliniensis*. In *C. dubliniensis*, Fca1p is encoded by the Cd*FCA1* gene, which shares 89% sequence identity with *FCA1* from *C. albicans*, and both contain an internal intron of 81 bp. In order to investigate whether a mutation(s) or a deletion(s) was present in the Cd*FCA1* gene encoding Fca1p, the DNA sequences of the Cd*FCA1* ORF were determined for all 12 5FC-resistant isolates (Table 1) and all 9 5FC-susceptible isolates (Table 1) investigated in the study. For each isolate, the Cd*FCA1* sequence was compared with that of the Cd*FCA1* consensus sequence of 5FC-susceptible *C. dubliniensis* type strain CD36 (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/). Three separate SNPs were identified in the Cd*FCA1*-coding sequences of the 21 *C. dubliniensis* isolates investigated. Two synonymous SNPs were observed; one of these (position 264, A → T transition) occurred in all isolates sequenced, with the exception of six of the nine 5FC-susceptible isolates (isolates CD36, Eg203, Eg206, SA101, SA105, and SA115), and the second (position 390, T → C transition) occurred in only one of the 5FC-susceptible isolates (isolate p7276). The third SNP was nonsynonymous (position 86, C → T transition), resulting in an amino acid substitution (Ser29Leu) in all 12 of the 5FC-resistant isolates tested, but was not present in the Cd*FCA1* gene of the 9 5FC-susceptible isolates sequenced.

Tetracycline-inducible expression of Cd*FCA1* in *C. dubliniensis*. Transformation of the pNIM1-Cd*FCA1*^{r} cassette into 5FC-susceptible isolate p7276 yielded several transformants with the correct integration of the pNIM1-Cd*FCA1*^{r} cassette into the *ADH1* locus, as determined by Southern hybridization and PCR analysis. Broth microdilution assays were carried out with these transformants by using serial dilutions of 5FC from a concentration of 0.25 to one of 128 μg/ml. All of the transformants were 5FC susceptible (5FC MICs ≤ 0.25 μg/ml). This experiment was replicated with the addition of the tetracycline derivative DOX to the RPMI 1640 broth microdilution medium at a final concentration of 15 μg/ml in order to induce the expression of the pNIM1-Cd*FCA1*^{r} cassette. All of the DOX-induced transformants remained 5FC susceptible (5FC MICs ≤ 0.25 μg/ml). Transformation of pNIM1-Cd*FCA1*^{s} DNA into 5FC-resistant isolates SA113 and SA109 yielded several transformants with the correct integration of the pNIM1-Cd*FCA1*^{s} cassette into the *ADH1* locus, as determined by Southern hybridization and PCR analysis (Table 3). All of these transformants were 5FC resistant (5FC MICs ≥ 128 μg/ml). In contrast, in a parallel series of experiments, replicate broth microdilution assays were carried out with the addition of DOX (15 μg/ml) to the RPMI assay medium. This resulted in a dramatic change in the 5FC resistance phenotype of these transformants (isolates SA109T1, SA109T2, and SA113T1) from being 5FC resistant (5FC MICs ≥ 128 μg/ml) to being 5FC susceptible (5FC MICs ≤ 0.25 μg/ml) (Table 3). Similar results were obtained in separate broth microdilution experiments with DOX at a final concentration of 30 μg/ml (Table 3). As transformation controls, 5FC-resistant parental isolates SA109 and SA113 and 5FC-susceptible parental isolate p7276 were also transformed with the pNIM1 cassette containing the *GFP1* gene instead of the Cd*FCA1* gene (Table 3). These control pNIM1-*GFP1* transformants were also examined by broth microdilution assays in the presence and absence of DOX (15 μg/ml). The presence or absence of DOX in the broth microdilution medium had no effect on the 5FC MICs, and all the transformant derivatives tested exhibited 5FC MICs similar to those of their respective parental isolates (Table 3). These results strongly suggest that expression of the Cd*FCA1*^{s} gene by the transformants harboring the entire pNIM1-Cd*FCA1*^{s} cassette (i.e., isolates SA109T1, SA109T2, and SA113T1; Table 3) is responsible for the DOX-inducible 5FC susceptibility exhibited by these transformants.

TABLE 3. Susceptibilities of *C. dubliniensis* isolates and pNIM1-*FCA1/GFP1* transformant derivatives in the presence or absence of DOX

Isolate or transformant	5FC MIC_{50} (μg/ml) with DOX at:		
	None	15 μg/ml	30 μg/ml
C. dubliniensis clinical isolates			
SA113	≥128	≥128	≥128
SA109	≥128	≥128	≥128
p7276	≤0.25	≤0.25	≤0.25
C. dubliniensis pNIM1-Cd*FCA1*^{r} transformants			
p7276T1	≤0.25	≤0.25	≤0.25
p7276T2	≤0.25	≤0.25	≤0.25
p7276T3	≤0.25	≤0.25	≤0.25
C. dubliniensis pNIM1-Cd*FCA1*^{s} transformants			
SA109T1	≥128	≤0.25	≤0.25
SA109T2	≥128	≤0.25	≤0.25
SA113T1	≥128	≤0.25	≤0.25
C. dubliniensis pNIM1-*GFP1* transformants			
SA113-GFP1T1	≥128	≥128	≥128
SA109-GFP1T1	≥128	≥128	≥128
SA109-GFP1T2	≥128	≥128	≥128
p7276-GFP1T1	≤0.25	≤0.25	≤0.25

Cd*FCA1* expression analysis. A comparison of Cd*FCA1* gene expression by the 5FC-susceptible isolates (isolates Eg204 and p7276) and 5FC-resistant isolates (isolates SA113, SA109, and Eg202) and the pNIM1-Cd*FCA1*^{s} transformant derivatives (isolates SA113T1, SA109T1, and SA109T2) was undertaken following exposure to DOX (15 μg/ml) or following exposure to DOX (15 μg/ml) and a subinhibitory concentration (6.4 ng/ml) of 5FC. The expression of Cd*FCA1* was analyzed by quantitative real-time PCR, and the data were normalized to the level of *ACT1* expression, which was used as an internal control. Two-tailed Student's *t* tests were carried out on the Cd*FCA1* expression values obtained from 5FC-susceptible and 5FC-resistant clinical isolates, and no significant differences in Cd*FCA1* expression were observed between the two groups of isolates in either the presence ($P = 0.47$) or the absence ($P = 0.16$) of 5FC in the growth medium. This suggests that alter-

ations in Cd*FCA1* expression do not play a significant role in mediating 5FC resistance. Upon exposure to 5FC (6.4 ng/ml), the level of Cd*FCA1* expression increased in both the 5FC-susceptible isolates (range, 14.5- to 25-fold) and the 5FC-resistant isolates (range, 4- to 18-fold). Two-tailed Student's *t* tests confirmed the significance of these Cd*FCA1* expression increases upon addition of 5FC to the growth medium ($P < 0.001$). On exposure to DOX, transformant derivatives SA109T1, SA109T2, and SA113T1 all showed significant ($P < 0.001$) increases in their levels of Cd*FCA1* expression (range, 5- to 26-fold) in comparison to those of parental isolates SA109 and SA113. These transformant derivatives also showed significant ($P < 0.001$) increases in their levels of Cd*FCA1* expression (range, 5- to 22-fold) in comparison to those of their parental isolates, isolates SA109 and SA113, in the presence of a subinhibitory concentration of 5FC as well as DOX.

DISCUSSION

Studies of pyrimidine salvage pathways (Fig. 1) and 5FC resistance mechanisms in yeast species have previously been undertaken with *S. cerevisiae* (9, 17, 25, 42), *C. albicans* (9, 13, 43), and *C. lusitaniae* (5, 22, 26). Investigations with *S. cerevisiae* have shown that the disruption of the *FCY2* or the *FUR1* gene can play a role in 5FC resistance, but only the *FCA1* gene is absolutely required for the mediation of 5FC susceptibility (24, 25). In *C. lusitaniae*, Papon et al. (26) reported that inactivation of either the *FCA1* or the *FCY2* gene mediates 5FC resistance (5FC MICs, 128 μg/ml and 64 μg/ml, respectively) and promotes cross-resistance to 5FC and fluconazole (MICs, 4 μg/ml to 32 μg/ml), and the authors suggested that this was due to the competitive inhibition of fluconazole uptake by extracellular 5FC (5, 22, 26). Further analysis has identified a nonsense mutation in the *FCY2* gene that resulted in a truncated purine cytosine permease in seven such isolates. In addition to this finding, a missense mutation (Met9Thr transition) has been identified in the *FCY1* genes of four clinical *C. lusitaniae* isolates that also exhibited 5FC and 5FC-fluconazole cross-resistance (10). In *C. albicans*, two different research groups (7, 13) identified a homozygous Arg101Cys amino acid substitution in the *FUR1*-encoded UPRT to be the most common cause of high-level 5FC resistance (5FC MICs, 8 to >64 μg/ml). Isolates that were heterozygous for this transition exhibited reduced 5FC susceptibility (5FC MICs, 0.5 to 1 μg/ml) (13). Furthermore, a homozygous Gly28Asp substitution in Fca1p, encoded by *FCA1*, was suggested by Hope et al. to be an alternative method of resistance in a *C. albicans* isolate that did not harbor the UPRT-associated Arg101Cys substitution (13). Finally, Hope et al. (13) also described a *C. albicans* isolate with a Ser29Leu substitution in Fca1p which exhibited an intermediate level of 5FC resistance (5FC MIC, 4 μg/ml). In the light of the findings of the previous studies, we hypothesized that the *C. dubliniensis* pyrimidine salvage pathway very likely retains structural and functional homology with the pyrimidine salvage pathways in other *Candida* species, as *C. dubliniensis* is the closest relative to *C. albicans* in the genus *Candida*. Therefore, we investigated the Cd*FCA1*-encoded Fca1p (Fig. 1) as a possible cause of *C. dubliniensis* clade-specific 5FC resistance. Initially, broth microdilution assays were carried out with both 5FC and 5FU to determine whether the deamination step in the 5FC metabolic pathway was responsible for 5FC resistance. If the deamination step was responsible for 5FC resistance, bypassing its requirement in the metabolic pathway by exposing 5FC-resistant cells to 5FU should result in susceptibility to 5FU (see Fig. 1). Both 5FC-susceptible and 5FC-resistant isolates exhibited similar 5FU MICs (Table 1), indicating that Fca1p is very likely responsible for the clade-specific 5FC resistance in *C. dubliniensis*. In order to investigate this possibility further, DNA sequence analysis of the Cd*FCA1* genes from 12 5FC-resistant and 9 5FC-susceptible *C. dubliniensis* isolates was undertaken. This identified a homozygous Ser29Leu substitution that occurred exclusively among 5FC-resistant isolates. This radical substitution results in the replacement of a hydrophilic polar amino acid (serine) with a hydrophobic nonpolar residue (leucine) in the β1 strand of the cytosine deaminase enzyme and is closely linked to an active-site residue, according to the yeast cytosine deaminase structure defined by Ko et al. (16). This amino acid substitution may disrupt the quaternary structure of the enzyme, distorting the active site and inhibiting the conversion of the 5FC prodrug to its toxic form, 5FU. As mentioned above, a similar amino acid substitution was reported by Hope et al. in a *C. albicans* isolate; however, that isolate exhibited intermediate resistance (5FC MIC, 4 μg/ml) to 5FC (13), in comparison to the high levels of 5FC resistance observed in the *C. dubliniensis* isolates displaying the Ser29Leu substitution in the present study. The differences in the levels of resistance to 5FC exhibited by the *C. albicans* isolate reported by Hope et al. and the 5FC-resistant *C. dubliniensis* isolates reported here, all of which harbored the same Ser29Leu substitution in Fca1p, may be due to the fact that different 5FC MIC determination methods were used in the two studies: the EUCAST method (33) was used in the previous study for the *C. albicans* isolates, and the CLSI method (6) was used in the present study for the *C. dubliniensis* isolates. Alternatively, differences in Cd*FCA1* and *FCA1* gene expression or posttranscriptional or posttranslational modifications may be responsible for the differences between the two species in the levels of resistance to 5FC exhibited by isolates harboring the Fca1p Ser29Leu substitution.

In order to obtain direct evidence that the Ser29Leu substitution present in Fca1p from 5FC-resistant *C. dubliniensis* isolates was responsible for the 5FC-resistant phenotype in these isolates, the gene encoding cytosine deaminase from 5FC-susceptible *C. dubliniensis* isolate p7276 (Cd*FCA1*^{s}), which was originally recovered in Israel and which lacked the Ser29Leu substitution, was introduced into the *ADH1* locus of two separate 5FC-resistant Saudi Arabian isolates, SA109 (ITS genotype 3) and SA113 (ITS genotype 4), by the tetracycline-inducible cassette pNIM1 (27). Three transformant derivatives (isolates SA109T1, SA109T2, and SA113T1) harboring the complete pNIM1-Cd*FCA1*^{s} cassette integrated into the *ADH1* locus were tested and exhibited DOX-inducible 5FC susceptibility on the acquisition and expression of the Cd*FCA1*^{s} gene (Table 3). In contrast, transformant derivatives of 5FC-susceptible *C. dubliniensis* isolate p7276 harboring the complete pNIM1-Cd*FCA1*^{r} cassette encoding the *FCA1* gene with the Ser29Leu substitution from 5FC-resistant *C. dubliniensis* isolate SA113 integrated into the *ADH1* locus remained 5FC susceptible following DOX induction (Table 3). These findings provided convincing evidence that the Ser29Leu substitution in

Fca1p from 5FC-resistant isolates was responsible for 5FC resistance in these isolates, but they also showed that the 5FC resistance mutation is recessive and, thus, that 5FC resistance is not expressed in the presence of a wild-type, functional *FCA1* allele.

No significant difference in Cd*FCA1* expression was detected between the 5FC-susceptible and the 5FC-resistant *C. dubliniensis* isolates tested by quantitative real-time PCR analysis in either the presence or the absence of subinhibitory concentrations of 5FC in the growth medium. These results indicate that the lack of expression or the reduced expression of Cd*FCA1* in 5FC-resistant *C. dubliniensis* isolates following exposure to subinhibitory concentrations of 5FC was not responsible for the 5FC resistance in 5FC-resistant isolates and that the Ser29Leu substitution is very likely the sole method of Cd*FCA1*-mediated 5FC resistance in the *C. dubliniensis* isolates investigated. Previous studies have shown that free pyrimidines often present in peptones present in some brands of culture media can antagonize the activity of 5FC (8). Antagonism was not observed in the present study with the YPD-grown cultures used in the expression studies. Following the addition of a subinhibitory concentration of 5FC (i.e., 6.4 ng/ml) to YPD-grown cultures, the quantitative real-time PCR experiments consistently showed that both 5FC-susceptible and 5FC-resistant *C. dubliniensis* isolates exhibited significant upregulation of *FCA1* expression: 14.5- to 20-fold and 4- to 18-fold, respectively.

All Cd25 fingerprint group III *C. dubliniensis* isolates, apart from one (isolate p7718; Table 1), tested so far exhibit high-level 5FC resistance; all were originally recovered in Saudi Arabia or Egypt; and all belong to MLST clade C3 (4, 18). The close genetic relationship shared by these isolates is reflected by their identical Cd*FCA1* DNA sequences and high-level resistance to 5FC. It is highly likely that an identical mechanism is used to mediate 5FC resistance in all of these isolates. MLST C3 clade (Cd25 fingerprint group III) *C. dubliniensis* isolates can be subdivided into ITS genotypes 3 and 4 on the basis of the nucleotides sequence of the ITS region of the ribosomal DNA operon (4), although clade C3 isolates of both ITS genotypes exhibit high-level 5FC resistance. In the present study, clade C3 *C. dubliniensis* isolates SA109 (ITS genotype 3) and SA113 (genotype 4) were both transformed with the pNIM1-Cd*FCA1*^{s} cassette, and both yielded transformant derivatives (isolates SA109T1, SA109T2, and SA113T1) that exhibited DOX-inducible 5FC susceptibility. These findings support our view that clade-specific 5FC resistance in *C. dubliniensis* is mediated by a common molecular mechanism, i.e., the presence of the Ser29Leu substitution in Fca1p.

This is not the first report of a clade-specific SNP that has resulted in the alteration of a protein involved in antifungal drug resistance in *C. dubliniensis*. In 2002, Moran et al. (20) reported that 58% of ITS genotype 1 *C. dubliniensis* isolates (Cd25 group I, MLST clade C1) harbored a TAG nonsense mutation in the *CDR1* gene encoding an ABC transporter. In *C. albicans*, the upregulation of *CDR1* is the most common mechanism of fluconazole resistance, whereas in *C. dubliniensis*, the most common mechanism of fluconazole resistance involves the overexpression of the *MDR1* gene encoding a multidrug transporter (21). These studies highlight the fact that despite the close phylogenetic relationship between *C. dubliniensis* and *C. albicans*, resistance to particular antifungal drugs can be due to different mechanisms in the two species.

In conclusion, the results of this study demonstrate that the presence of a Ser29Leu substitution in Fca1p in *C. dubliniensis* isolates is responsible for clade-specific resistance to 5FC. Isolates belonging to *C. dubliniensis* clade C3 have been recovered only from individuals of Arab ethnicity in Saudi Arabia, Egypt, and Israel (4, 18). Resistance to 5FC has not yet been reported in *C. dubliniensis* isolates from other countries around the world, apart from Kuwait. In 2004, Ahmad et al. reported the recovery of two 5FC-resistant isolates of *C. dubliniensis* from Kuwait (1). Because of Kuwait's close proximity to Saudi Arabia, it is likely that these isolates also belong to *C. dubliniensis* MLST clade C3.

ACKNOWLEDGMENT

This study was supported by the Microbiology Research Unit, Dublin Dental School and Hospital.

REFERENCES

1. **Ahmad, S., Z. Khan, E. Mokaddas, and Z. U. Khan.** 2004. Isolation and molecular identification of *Candida dubliniensis* from non-human immunodeficiency virus-infected patients in Kuwait. J. Med. Microbiol. **53:**633–637.
2. **Al Mosaid, A., D. Sullivan, I. F. Salkin, D. Shanley, and D. C. Coleman.** 2001. Differentiation of *Candida dubliniensis* from *Candida albicans* on Staib agar and caffeic acid-ferric citrate agar. J. Clin. Microbiol. **39:**323–327.
3. **Al Mosaid, A., D. J. Sullivan, and D. C. Coleman.** 2003. Differentiation of *Candida dubliniensis* from *Candida albicans* on Pal's agar. J. Clin. Microbiol. **41:**4787–4789.
4. **Al Mosaid, A., D. J. Sullivan, I. Polacheck, F. A. Shaheen, O. Soliman, S. Al Hedaithy, S. Al Thawad, M. Kabadaya, and D. C. Coleman.** 2005. Novel 5-flucytosine-resistant clade of *Candida dubliniensis* from Saudi Arabia and Egypt identified by Cd25 fingerprinting. J. Clin. Microbiol. **43:**4026–4036.
5. **Chapeland-Leclerc, F., J. Bouchoux, A. Goumar, C. Chastin, J. Villard, and T. Noel.** 2005. Inactivation of the *FCY2* gene encoding purine-cytosine permease promotes cross-resistance to flucytosine and fluconazole in *Candida lusitaniae*. Antimicrob. Agents Chemother. **49:**3101–3108.
6. **Clinical and Laboratory Standards Institute.** 2002. Reference method for broth dilution antifungal susceptibility testing of yeasts; approved standard; M27-A2. Clinical and Laboratory Standards Institute, Wayne, PA.
7. **Dodgson, A. R., K. J. Dodgson, C. Pujol, M. A. Pfaller, and D. R. Soll.** 2004. Clade-specific flucytosine resistance is due to a single nucleotide change in the *FUR1* gene of *Candida albicans*. Antimicrob. Agents Chemother. **48:**2223–2227.
8. **Doern, G. V., T. A. Tubert, K. Chapin, and M. G. Rinaldi.** 1986. Effect of medium composition on results of macrobroth dilution antifungal susceptibility testing of yeasts. J. Clin. Microbiol. **24:**507–511.
9. **Erbs, P., F. Exinger, and R. Jund.** 1997. Characterization of the *Saccharomyces cerevisiae FCY1* gene encoding cytosine deaminase and its homologue *FCA1* of *Candida albicans*. Curr. Genet. **31:**1–6.
10. **Florent, M., T. Noel, G. Ruprich-Robert, B. Da Silva, V. Fitton-Ouhabi, C. Chastin, N. Papon, and F. Chapeland-Leclerc.** 2009. Nonsense and missense mutations in *FCY2* and *FCY1* genes are responsible for flucytosine resistance and flucytosine-fluconazole cross-resistance in clinical isolates of *Candida lusitaniae*. Antimicrob. Agents Chemother. **53:**2982–2990.
11. **Gee, S. F., S. Joly, D. R. Soll, J. F. Meis, P. E. Verweij, I. Polacheck, D. J. Sullivan, and D. C. Coleman.** 2002. Identification of four distinct genotypes of *Candida dubliniensis* and detection of microevolution in vitro and in vivo. J. Clin. Microbiol. **40:**556–574.
12. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557–580.
13. **Hope, W. W., L. Tabernero, D. W. Denning, and M. J. Anderson.** 2004. Molecular mechanisms of primary resistance to flucytosine in *Candida albicans*. Antimicrob. Agents Chemother. **48:**4377–4386.
14. **Hoskins, J., and J. S. Butler.** 2008. RNA-based 5-fluorouracil toxicity requires the pseudouridylation activity of Cbf5p. Genetics **179:**323–330.
15. **Joly, S., C. Pujol, M. Rysz, K. Vargas, and D. R. Soll.** 1999. Development and characterization of complex DNA fingerprinting probes for the infectious yeast *Candida dubliniensis*. J. Clin. Microbiol. **37:**1035–1044.
16. **Ko, T. P., J. J. Lin, C. Y. Hu, Y. H. Hsu, A. H. Wang, and S. H. Liaw.** 2003. Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. J. Biol. Chem. **278:**19111–19117.
17. **Kurtz, J. E., F. Exinger, P. Erbs, and R. Jund.** 1999. New insights into the pyrimidine salvage pathway of *Saccharomyces cerevisiae*: requirement of six genes for cytidine metabolism. Curr. Genet. **36:**130–136.

18. **McManus, B. A., D. C. Coleman, G. Moran, E. Pinjon, D. Diogo, M. E. Bougnoux, S. Borecka-Melkusova, H. Bujdakova, P. Murphy, C. d'Enfert, and D. J. Sullivan.** 2008. Multilocus sequence typing reveals that the population structure of *Candida dubliniensis* is significantly less divergent than that of *Candida albicans*. J. Clin. Microbiol. **46:**652–664.
19. **McManus, B. A., D. J. Sullivan, G. P. Moran, C. D'Enfert, M. E. Bougnoux, M. A. Nunn, and D. C. Coleman.** Genetic differences between avian and human isolates of *Candida dubliniensis*. Emerg. Infect. Dis. **15:**1467–1470.
20. **Moran, G., D. Sullivan, J. Morschhauser, and D. Coleman.** 2002. The *Candida dubliniensis* Cd*CDR1* gene is not essential for fluconazole resistance. Antimicrob. Agents Chemother. **46:**2829–2841.
21. **Moran, G. P., D. Sanglard, S. M. Donnelly, D. B. Shanley, D. J. Sullivan, and D. C. Coleman.** 1998. Identification and expression of multidrug transporters responsible for fluconazole resistance in *Candida dubliniensis*. Antimicrob. Agents Chemother. **42:**1819–1830.
22. **Noel, T., F. Francois, P. Paumard, C. Chastin, D. Brethes, and J. Villard.** 2003. Flucytosine-fluconazole cross-resistance in purine-cytosine permease-deficient *Candida lusitaniae* clinical isolates: indirect evidence of a fluconazole uptake transporter. Antimicrob. Agents Chemother. **47:**1275–1284.
23. **Odds, F. C., M. E. Bougnoux, D. J. Shaw, J. M. Bain, A. D. Davidson, D. Diogo, M. D. Jacobsen, M. Lecomte, S. Y. Li, A. Tavanti, M. C. Maiden, N. A. Gow, and C. d'Enfert.** 2007. Molecular phylogenetics of *Candida albicans*. Eukaryot. Cell **6:**1041–1052.
24. **Paluszynski, J. P., R. Klassen, and F. Meinhardt.** 2008. Genetic prerequisites for additive or synergistic actions of 5-fluorocytosine and fluconazole in baker's yeast. Microbiology **154:**3154–3164.
25. **Paluszynski, J. P., R. Klassen, M. Rohe, and F. Meinhardt.** 2006. Various cytosine/adenine permease homologues are involved in the toxicity of 5-fluorocytosine in *Saccharomyces cerevisiae*. Yeast **23:**707–715.
26. **Papon, N., T. Noel, M. Florent, S. Gibot-Leclerc, D. Jean, C. Chastin, J. Villard, and F. Chapeland-Leclerc.** 2007. Molecular mechanism of flucytosine resistance in *Candida lusitaniae*: contribution of the *FCY2*, *FCY1*, and *FUR1* genes to 5-fluorouracil and fluconazole cross-resistance. Antimicrob. Agents Chemother. **51:**369–371.
27. **Park, Y. N., and J. Morschhauser.** 2005. Tetracycline-inducible gene expression and gene deletion in *Candida albicans*. Eukaryot. Cell **4:**1328–1342.
28. **Pfaffl, M. W.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29:**e45.
29. **Pfaller, M. A., S. A. Messer, S. Gee, S. Joly, C. Pujol, D. J. Sullivan, D. C. Coleman, and D. R. Soll.** 1999. In vitro susceptibilities of *Candida dubliniensis* isolates tested against the new triazole and echinocandin antifungal agents. J. Clin. Microbiol. **37:**870–872.
30. **Polacheck, I., J. Strahilevitz, D. Sullivan, S. Donnelly, I. F. Salkin, and D. C. Coleman.** 2000. Recovery of *Candida dubliniensis* from non-human immunodeficiency virus-infected patients in Israel. J. Clin. Microbiol. **38:**170–174.
31. **Polak, A., and H. J. Scholer.** 1975. Mode of action of 5-fluorocytosine and mechanisms of resistance. Chemotherapy **21:**113–130.
32. **Pujol, C., M. A. Pfaller, and D. R. Soll.** 2004. Flucytosine resistance is restricted to a single genetic clade of *Candida albicans*. Antimicrob. Agents Chemother. **48:**262–266.
33. **Rodríguez-Tudela, J. L., F. Barchiesi, F. Bille, E. Chryssanthou, M. Cuenca-Estrella, D. Denning, J. P. Donnelly, B. Dupont, W. Fegeler, C. Moore, M. Richardson, and P. E. Verweij.** 2003. Method for the determination of minimum inhibitory concentration (MIC) by broth dilution of fermentative yeasts. Clin. Microbiol. Infect. **9:**i–viii.
34. **Schmittgen, T. D., and K. J. Livak.** 2008. Analyzing real-time PCR data by the comparative C_T method. Nat. Protoc. **3:**1101–1108.
35. **Staib, P., G. P. Moran, D. J. Sullivan, D. C. Coleman, and J. Morschhauser.** 2001. Isogenic strain construction and gene targeting in *Candida dubliniensis*. J. Bacteriol. **183:**2859–2865.
36. **Sullivan, D., K. Haynes, J. Bille, P. Boerlin, L. Rodero, S. Lloyd, M. Henman, and D. Coleman.** 1997. Widespread geographic distribution of oral *Candida dubliniensis* strains in human immunodeficiency virus-infected individuals. J. Clin. Microbiol. **35:**960–964.
37. **Sullivan, D. J., G. P. Moran, and D. C. Coleman.** 2005. *Candida dubliniensis*: ten years on. FEMS Microbiol. Lett. **253:**9–17.
38. **Sullivan, D. J., G. P. Moran, E. Pinjon, A. Al-Mosaid, C. Stokes, C. Vaughan, and D. C. Coleman.** 2004. Comparison of the epidemiology, drug resistance mechanisms, and virulence of *Candida dubliniensis* and *Candida albicans*. FEMS Yeast Res. **4:**369–376.
39. **Sullivan, D. J., T. J. Westerneng, K. A. Haynes, D. E. Bennett, and D. C. Coleman.** 1995. *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. Microbiology **141:**1507–1521.
40. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
41. **Waldorf, A. R., and A. Polak.** 1983. Mechanisms of action of 5-fluorocytosine. Antimicrob. Agents Chemother. **23:**79–85.
42. **Weber, E., C. Rodriguez, M. R. Chevallier, and R. Jund.** 1990. The purine-cytosine permease gene of *Saccharomyces cerevisiae*: primary structure and deduced protein sequence of the *FCY2* gene product. Mol. Microbiol. **4:**585–596.
43. **Whelan, W. L., and D. Kerridge.** 1984. Decreased activity of UMP pyrophosphorylase associated with resistance to 5-fluorocytosine in *Candida albicans*. Antimicrob. Agents Chemother. **26:**570–574.
44. **Willis, A. M., W. A. Coulter, D. J. Sullivan, D. C. Coleman, J. R. Hayes, P. M. Bell, and P. J. Lamey.** 2000. Isolation of *C. dubliniensis* from insulin-using diabetes mellitus patients. J. Oral Pathol. Med. **29:**86–90.

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

Chapter 2
Molecular Epidemiology of *Candida* Species

Gary P. Moran, Brenda A. McManus, David C. Coleman, and Derek J. Sullivan

Abstract *Candida* species have increased in importance as opportunistic pathogens over the last 25 years. *C. albicans* is still the major fungal pathogen of humans, however during this period, other previously obscure *Candida* species have emerged as significant pathogens. This increase in infections has created the need for reliable, informative and discriminatory techniques for strain typing in *Candida* species and several molecular techniques have been evaluated for this purpose. In the post-genome era, analysis of sequence polymorphisms has become the method of choice for strain typing in *C. albicans* and multi-locus sequence typing (MLST) has become the standard tool for this purpose. This chapter summarises the main developments in this area in recent years, describing the impact of MLST on our understanding of the epidemiology and population structure of *Candida* species. The potential impact of high throughput, post-Sanger sequencing technologies on the field is also discussed.

2.1 Introduction

Candida albicans is the major fungal pathogen of humans. A normal resident of the oral-gastrointestinal tract, *C. albicans* is an opportunistic pathogen and infection is generally restricted to those with impaired defences or specific immunodeficiencies (Wenzel 1995). Although *C. albicans* is responsible for the majority of yeast infections in humans, several other *Candida* species have also been associated with disease, including *C. dubliniensis*, *C. tropicalis*, *C. parapsilosis*, *C. glabrata* and *C. krusei* (Moran et al. 2002). These species are recovered less frequently from the oral-gastrointestinal tract in healthy individuals and are generally considered less pathogenic than *C. albicans*. However, in the compromised host, these species

G.P. Moran (✉), B.A. McManus, D.C. Coleman, and D.J. Sullivan
Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental School & Hospital, Trinity College Dublin, University of Dublin, Dublin 2, Republic of Ireland
e-mail: gpmoran@dental.tcd.ie

R. Ashbee and E.M. Bignell (eds.), *Pathogenic Yeasts*, The Yeast Handbook,
DOI 10.1007/978-3-642-03150-2_2,

have become significant pathogens and distinct differences in their epidemiology have been shown due to their unique biology, host specificities and anti-fungal drug susceptibilities (Moran et al. 2002).

Candida infections are generally endogenous in origin, and prior colonisation with the organism is often regarded as one of the major risk factors for candidiasis (Pfaller 1995; Pfaller and Diekema 2007). Colonisation rates are higher in individuals whose mucosal immunity is impaired due to old age, diabetes mellitus or smoking (Lockhart et al. 1998; Manfredi et al. 2002). The balance between colonisation and overt infection is delicate and even discreet changes in the host's normal commensal flora can lead to mucosal infection (Lockhart et al. 1998; Vargas and Joly 2002). Mucosal infection, in the form of pseudomembraneous candidosis (thrush) also commonly occurs when oral or vaginal immunity fails to keep the endogenous yeast population in check. More severe, invasive forms of infection can occur when neutrophil function is impaired or counts are lowered due to immunosuppressive therapy or cancer (Pfaller 1995).

2.1.1 Epidemiology of Candidosis

Candidal carriage is ubiquitous in the human population and superficial infection of the oral mucosa is a common sequela of underlying immunodeficiencies. Oral candidosis is a frequent complication of cancer chemotherapy, diabetes, broad-spectrum antibiotic use and HIV-infection (Manfredi et al. 2002; Sangeorzan et al. 1994; Vargas and Joly 2002). The HIV pandemic resulted in a huge increase in the incidence of mucosal *Candida* infection during the 1990s. However, since the introduction of highly active anti-retroviral therapy (HAART) in the late 1990s, the incidence of oral candidosis in the HIV-infected population has dropped dramatically (Cauda et al. 1999). *C. albicans* is by far the most common cause of mucosal yeast infection, being the sole species recovered from up to 70% of HIV-infected individuals and up to 90% of cases of *Candida* vaginitis (Coleman et al. 1993; Sobel 2007). Other *Candida* species can be recovered alone or co-isolated with *C. albicans* from sites of mucosal infection (Coleman et al. 1995). The significance of non-*C. albicans Candida* species in oral specimens is disputed by some researchers, who have associated their isolation with asymptomatic carriage (Ruhnke 2006). However, some studies have directly implicated non-*C. albicans Candida* species, such as *C. dubliniensis* and *C. glabrata,* with overt symptoms of oral candidosis and have also associated these species with alternative clinical presentations such as erythematous candidosis (Fidel et al. 1999; Sullivan et al. 1993). The emergence of some non-*C. albicans Candida* species in the HIV-infected population in the 1990s may have been a direct result of the widespread use of fluconazole, as *C. glabrata* and *C. krusei* tend to exhibit intrinsic resistance to this agent (Warnock et al. 1988; Wingard et al. 1991).

The epidemiology of invasive *Candida* infection has changed dramatically in the last 30 years (Pfaller 1995). The incidence of these infections has steadily increased

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

since the 1980s, largely due to the increasing population of immunocompromised patients in our hospitals (Banerjee et al. 1991; Martin et al. 2003). Widespread use of cytotoxic therapies to treat cancer and the use of immunosuppressive drugs in organ transplantation have greatly increased the number of neutropenic patients in intensive care units. Risk factors for infection include cancer, extremes of age, prior colonisation and the presence of intravenous catheters (Pittet et al. 1994; Wenzel 1995). Recently, Martin et al. (2003) analysed the rate of sepsis in hospitals in the USA from 1979 to 2000 and found that the rate of sepsis due to fungal organisms increased by 207%. More recent data based on figures compiled from National Hospital Discharge Survey (NHDS) statistics in the USA indicate a levelling off in the incidence of nosocomial fungal infection, with an incidence of 22–29 infections per 100,000 population in the period 1996–2003 (Pfaller and Diekema 2007). Similar incidences have been reported in Europe and Canada (Pfaller and Diekema 2007). Data indicate that the distribution of species responsible for invasive infection has also shifted during this period. Most reports indicate that the recovery of non-*C. albicans Candida* species from blood cultures has increased relative to *C. albicans* (Nguyen et al. 1996; Pfaller and Diekema 2004). *C. albicans* now only accounts for 50–60% of all species recovered from blood cultures, with *C. glabrata, C. parapsilosis* and *C. tropicalis* making up for the majority of the remaining species (Pfaller and Diekema 2007). The reasons for this shift are unclear, but this may be partly due to the reduced susceptibility of these species to fluconazole, commonly used throughout the 1990s, or to the increase in the numbers of immunocompromised patients susceptible to infection with less virulent species of *Candida* (Moran et al. 2002). *C. glabrata* has a high propensity to develop resistance to azole anti-fungals, whereas *C. krusei* is inherently resistant to fluconazole (Fidel et al. 1999; Samaranayake and Samaranayake 1994). However, increased reporting of infection caused by non-*C. albicans Candida* species may also be the result of recent improvements in isolation and identification methods for *Candida* species.

The distribution of species recovered from blood culture also changes with geography, particularly with regard to *C. parapsilosis*, which is reported as the second most commonly isolated *Candida* species in Latin America and Europe, whereas in North America, *C. glabrata* is the second most significant species (Table 2.1). *C. parapsilosis* is the species most commonly recovered from the hands of health care workers and can often produce a mucoid biofilm, features

Table 2.1 Geographic variations in the recovery of *Candida* species from blood culture

Species	Location[a]			
	USA (%)	Europe (%)	Latin America (%)	Asia-Pacific (%)
C. albicans	51	60	50	56
C. glabrata	22	10	7	10
C. parapsilosis	14	12	16	16
C. tropicalis	7	9	20	14
C. krusei	2	5	2	2

[a]Data taken from Pfaller et al. (2006)

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

that may account for its high prevalence in catheter-related infections (Levin et al. 1998). *C. parapsilosis* is also particularly associated with infection in neonatal intensive care units (Levy et al. 1998).

2.2 Molecular Epidemiology

One of the goals of molecular epidemiology is to devise reliable, reproducible and informative methods to differentiate between unrelated isolates of the same species for purposes of epidemiological surveillance (Soll 2000). By distinguishing isolates based on phenotypic or molecular properties, one can identify those isolates that are highly likely to be epidemiologically related. These data can allow microbiologists to locate the source of infecting isolates in nosocomial outbreaks, in recurrent infections and inform us on the population structure of the organism in question. Prior to the widespread use of molecular techniques, mycologists relied on phenotypic properties such as morphology, carbohydrate utilisation patterns and serotyping to distinguish between isolates of *C. albicans* (Pfaller et al. 1990). The use of these techniques was hampered by their poor discriminatory power and the inherent phenotypic instability of *C. albicans*. During the 1990s, molecular techniques began to take precedence over phenotypic tests due to their greater discriminatory power. Several molecular typing methods have been applied to *C. albicans*, including multi-locus enzyme electrophoresis (MLEE), Restriction enzyme analysis (REA), karyotype analysis and randomly amplified polymorphic DNA (RAPD) analysis (Sullivan and Coleman 2002). However, of all of the techniques used during this period, Southern hybridisation of genomic DNA with sequences corresponding to dispersed, repetitive elements in the *Candida* genome proved to be the most discriminatory and reliable.

2.2.1 DNA Fingerprinting with Dispersed, Repetitive Elements

Southern hybridisation with the dispersed repetitive element Ca3 has proven to be one of the most informative typing methods available for epidemiological analysis of *C. albicans* (Schmid et al. 1990; Soll 2000). One of the advantages of Ca3 fingerprinting is the ability to digitally compare fingerprint patterns, which allows quantitative analysis of the genetic relationships between isolates (Schmid et al. 1990). However, the drawbacks of DNA fingerprint analysis include the laborious nature of generating the fingerprints and the difficulty of comparing fingerprint data between laboratories. Population studies with the Ca3 probe have identified five major genetic groups, referred to as 'clades', in the *C. albicans* population (Pujol et al. 2002). These clades have been termed I, II, III, SA and E and exhibit different geographic specificities and phenotypic traits. Isolates from clades SA and E are recovered predominately from South Africa and Europe, respectively. However,

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

strains from clade I predominate in all geographical areas (Pujol et al. 2002). Isolates from clade I also exhibit reduced susceptibility to the anti-fungal agent 5-fluorocytosine (5FC). Pujol et al. (2004) found that 73% of clade I isolates were resistant or less susceptible to flucytosine (MIC $\geq 0.5\,\mu g\ ml^{-1}$) whereas only 2% of non-clade I isolates exhibit reduced susceptibility (Pujol et al. 2004). The mechanism of 5FC resistance in this population has been linked to a point mutation (C301T) in the *FUR1* gene encoding phosphoribosyltransfersase (Dodgson et al. 2004). Isolates heterozygous at this locus exhibit reduced susceptibility while those exhibiting high-level resistance are homozygous for this substitution. In parallel to the work of Pujol et al. (2004), Schmid et al. (1999) also identified a group of closely related *C. albicans* isolates that predominate in all geographic areas and could be associated with all forms of disease. Schmid et al. (1999) argued that this group represents a general purpose genotype (GPG) of *C. albicans* that are especially successful at colonising the human host (Schmid et al. 1999). Evidence for how these genetic differences could contribute to virulence was provided by examination of *ALS7* allelic variation (Zhang et al. 2003). *ALS7* is a member of a gene family encoding a group of cell wall proteins called the agglutinin-like sequences (Als) with roles in adhesion (see Chap. 4). The majority of isolates within the GPG cluster had between 14 and 17 copies of a tandem repeat located within the open reading frame, and that these alleles were much less common in strains outside of the cluster. Variation in the number of tandem repeat copies has been associated with changes in Als protein adhesive properties (Oh et al. 2005).

Ca3 fingerprinting has also been used to resolve questions about the source and spread of infecting *C. albicans* isolates. Most individuals harbour their own unique strain of *C. albicans* and commensal isolates and infecting isolates are often genetically indistinguishable (Schmid et al. 1990; Schroppel et al. 1994; Vargas and Joly 2002). Some individuals, particularly HIV-infected patients, may be colonised by more than one strain of *C. albicans* (Vargas and Joly 2002). In cases of recurrent oral or vaginal candidosis, Ca3 fingerprinting has shown that the same strain often persists through different episodes of infection, however replacement of the original strain or the emergence of a closely related genetic variant of the original strain is not uncommon (Lockhart et al. 1996; Schroppel et al. 1994). The latter phenomenon has been termed 'substrain shuffling' or 'microevolution' and was initially identified in isolates recovered from recurrent vaginal infections (Lockhart et al. 1996). In this study, a fragment of the Ca3 probe, termed C1, was shown to be useful in distinguishing between closely related isolates. DNA fingerprinting has also provided evidence for transmission of *C. albicans* strains between sexual partners (Schroppel et al. 1994). Nosocomial transmission of *C. albicans* strains between patients in intensive care units has also been investigated by Ca3 fingerprinting (Marco et al. 1999; Taylor et al. 2003). However, evidence suggests that most infections are endogenous and that transmission of strains from healthcare workers to patients is less common.

Similar repetitive elements have been isolated from other *Candida* species and have been used to generate fingerprint patterns. A *C. dubliniensis* specific probe, Cd25, was described by Joly et al. (1999) that could discriminate two distinct

groups of *C. dubliniensis* isolates, termed Cd25 group I and II. The majority of Cd25 group I isolates (67.6%) were recovered from human immunodeficiency virus (HIV)-infected individuals, whereas the majority of Cd25 group II isolates (70.4%) were from HIV-negative individuals. Subsequent analysis identified a third distinct clade of *C. dubliniensis* isolates (Cd25 group III) recovered from patients in Saudi Arabia and Egypt (Al Mosaid et al. 2005). Interestingly, this clade of *C. dubliniensis* isolates were found to be resistant to 5FC, although the mechanism is so far unknown. David Soll and colleagues have also developed fingerprinting probes for *C. glabrata* (Lockhart et al. 1997) *C. tropicalis* (Joly et al. 1996) and *C. parapsilosis* (Enger et al. 2001).

2.2.2 *Multi-Locus Sequence Typing (MLST) of C. albicans*

In recent years, DNA sequencing has become more affordable and widely available, which has made typing methods that involve characterising DNA sequence polymorphisms more accessible. The most widely used of these techniques is MLST, initially developed for typing pathogenic bacteria (Maiden et al. 1998). The chief advantage of this technology is that data can be stored in databases and is readily accessible by researchers in other locations (Bougnoux et al. 2004). The nature of DNA sequence analysis means that the data are reproducible and unambiguous. MLST is a highly discriminatory method that relies on the analysis of nucleotide sequence polymorphisms within the sequences of PCR-generated fragments (400–500 bp) of 6–8 housekeeping genes (loci) (Odds and Jacobsen, 2008). An outline of the procedure is shown in Fig. 2.1. In haploid organisms, the sequences obtained at each locus are assigned as discreet alleles, and for each isolate the combination of alleles define an allelic profile, or sequence type (ST). While there are four possible variations at each polymorphic locus in haploid species, diploidy presents 10 possible combinations of the bases ATG and C due to potential heterozygosity, thus increasing the potential number of alleles at each locus. In the current schemes available for diploid *Candida* species, heterozygous genotypes are handled by superimposing the IUPAC one letter code on heterozygous bases (e.g. A or G = R, C or T = Y etc). To reflect this heterozygosity, allelic profiles in diploid species are assigned a diploid sequence type (DST). In *C. albicans*, two independent MLST schemes were initially proposed based on the sequences of six (Bougnoux et al. 2002) or eight (Tavanti et al. 2003) loci. Since then, a consensus scheme has been agreed consisting of seven loci for optimised MLST of *C. albicans*; Table 2.2). The choice of genes for MLST analysis is generally restricted to those with housekeeping functions that are subject to stabilising selection, that is the ratio of non-synononmous to synonymous or silent substitutions (dN/dS ratio) in their nucleotide sequence is less than 1.0 (Odds and Jacobsen 2008). However, the choice of loci must obviously provide sufficient sequence diversity to allow high levels of allelic discrimination.

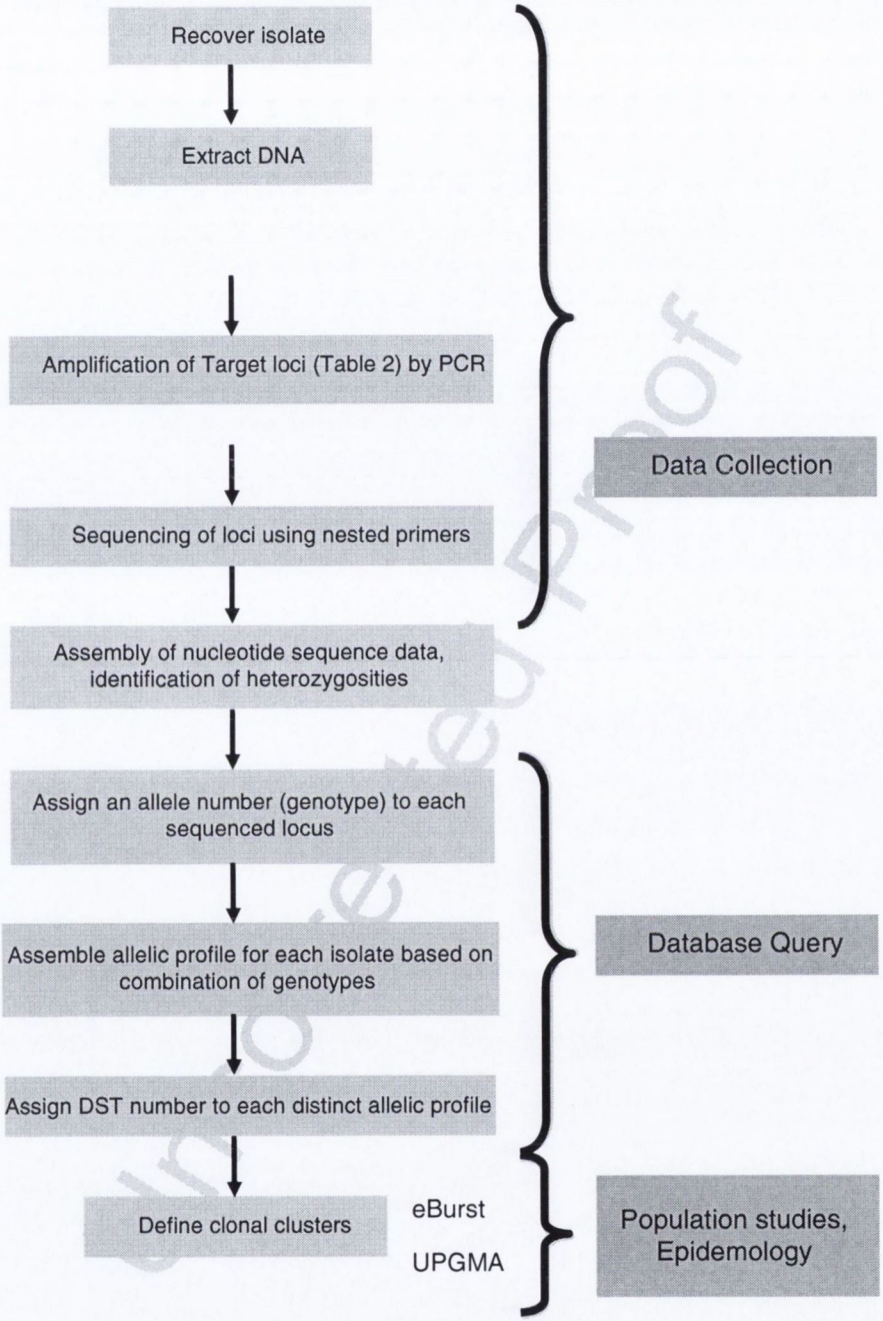


Fig. 2.1 Flow diagram outlining the critical steps in MLST of *Candida* species

t2.1 **Table 2.2** Loci used in the MLST schemes available for analysis of *Candida* species

t2.2 Species	DSTs/isolates[a]	Gene	Genotypes
t2.3 *C. albicans*	1,404/1,771	*AAT1a*	113
t2.4 http://test1.mlst.net/		*ACC1*	79
t2.5		*ADP1*	93
t2.6		*PMI1b*	85
t2.7		*SYA1*	136
t2.8		*VPS13*	194
t2.9		*ZWF1b*	198
t2.10 *C. dubliniensis*[b]	26/50	*AAT1b*	5
t2.11		*ACC1*	4
t2.12		*ADP1*	6
t2.13		*PMI1b*	7
t2.14		*RPN2*	3
t2.15		*SYA1*	5
t2.16		*exVPS13*	4
t2.17		*exZWF1b*	6
t2.18 *C. tropicalis*	205/260	*ICL1*	23
t2.19 http://pubmlst.org/ctropicalis/		*MDR1*	65
t2.20		*SAPT2*	25
t2.21		*SAPT4*	40
t2.22		*XYR1*	74
t2.23		*ZWF1a*	25
t2.24 *C. krusei*	99/134	*ADE2*	21
t2.25 http://pubmlst.org/ckrusei/		*HIS3*	14
t2.26		*LEU2*	17
t2.27		*LYS2*	20
t2.28		*NMT1*	24
t2.29		*TRP1*	24
t2.30 *C. glabrata*	70/212	*FKS*	25
t2.31 http://cglabrata.mlst.net/#		*LEU2*	18
t2.32		*NMT1*	34
t2.33		*TRP1*	23
t2.34		*UGP1*	13
t2.35		*URA3*	20

t2.36 [a]Data obtained from the Internet MLST database (http://calbicans.mlst.net/) for the relevant species (02/06/09) with the exception of *C. dubliniensis* data, which were obtained from McManus et al. (2008)
[b]Online database not yet available for *C. dubliniensis*

226 As mentioned above, the great advantage of MLST is that databases of
227 sequences and allelic profiles can be assembled, allowing multiple users to compare
228 data. The consensus *C. albicans* scheme can be queried at http://test1.mlst.net.
229 Here, using a web-based interface, users can assign allele numbers (referred to as
230 'genotypes') to their sequenced loci using a 'locus query interface' tool. A second
231 tool, the 'profile query interface' can then be used to compare the assembled allelic
232 profile of an isolate to those in the database and to identify isolates with an identical
233 or closely related profile (referred to as the ST in haploid species, or DST in
234 diploids). Users can send details of novel genotypes or sequence types to a database
235 curator for inclusion, permitting rapid expansion of the database.

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

2.2.2.1 Defining Clonal Clusters by MLST

For epidemiological purposes, different analytical methods can be applied to MLST data for the purpose of defining the relationships between microbes in a population. Traditionally, dendograms based on the unweighted-pair group method with arithmetic mean (UPGMA) technique is widely used for strain typing analysis (Bougnoux et al. 2004). UPGMA analyses MLST data at the level of the individual single nucleotide polymorphisms (SNPs). When applied to *C. albicans* MLST data sets, UPGMA analysis generates dendograms with a clade structure that closely matches that generated by Ca3 fingerprinting (Fig. 2.2) (Tavanti et al. 2005a). Tavanti et al. (2005a) identified four major clades by MLST, referred to as clades 1 to 4, which correspond to clades I, II, III and SA defined by Ca3 fingerprinting. However, isolates from Ca3 clade E were dispersed throughout the MLST clades.

Although useful for visualising population structure, the bifurcating method of lineage splitting implied in dendograms is not a true representation of the way in which microbial lineages emerge and diversify. An algorithm called eBURST was developed, which does not impose a tree-like pattern of descent on population structure (Feil et al. 2004). eBURST compares (D)STs of isolates and gathers isolates that differ at only one of the set of genes sequences (single locus variants, SLVs) into clonal clusters (Fig. 2.3). The model assumes that certain (D)STs will become established in a population and will then diversify by recombination or the accumulation of point mutations resulting in slight variations on the founding genotype. By this model, the microbial population will consist of a series of clonal complexes that can be recognised by the allelic profiles of the strain within the database (Fig. 2.3). Although useful for clustering isolates, BURST analysis of *C. albicans* MLST data highlights large numbers of singletons, that is, isolates that cannot be assigned to a clonal cluster (Odds and Jacobsen 2008). This may be due to the high levels of mitotic recombination in diploid *Candida* species as eBURST analysis is ideally suited to inferring relationships in populations where mutation is the main source of variation.

2.2.2.2 Insights on Population Structure of *C. albicans* from MLST

To date, the largest published MLST study of the population structure of *C. albicans* consisted of 1,391 isolates, most of which (96.7%) could be assigned to one of 17 clades (Odds et al. 2007). Different clades exhibited significant variation in the geographic origins of isolates. However, no association with anatomical source could be identified. As noted in previous studies, reduced susceptibility to azole anti-fungals was associated with homozygosity at the mating type locus (MTL) (Tavanti et al. 2005a). The reason for the association between azole resistance and MTL homozygosity involves the *TAC1* gene, which is located close to the MTL on chromosome 5 (Coste et al. 2006). Mutations in the *TAC1* gene have been identified, that when homozygous, can result in azole resistance. Loss of heterozygosity at the *TAC1* locus is often associated with MTL homozygosity due

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

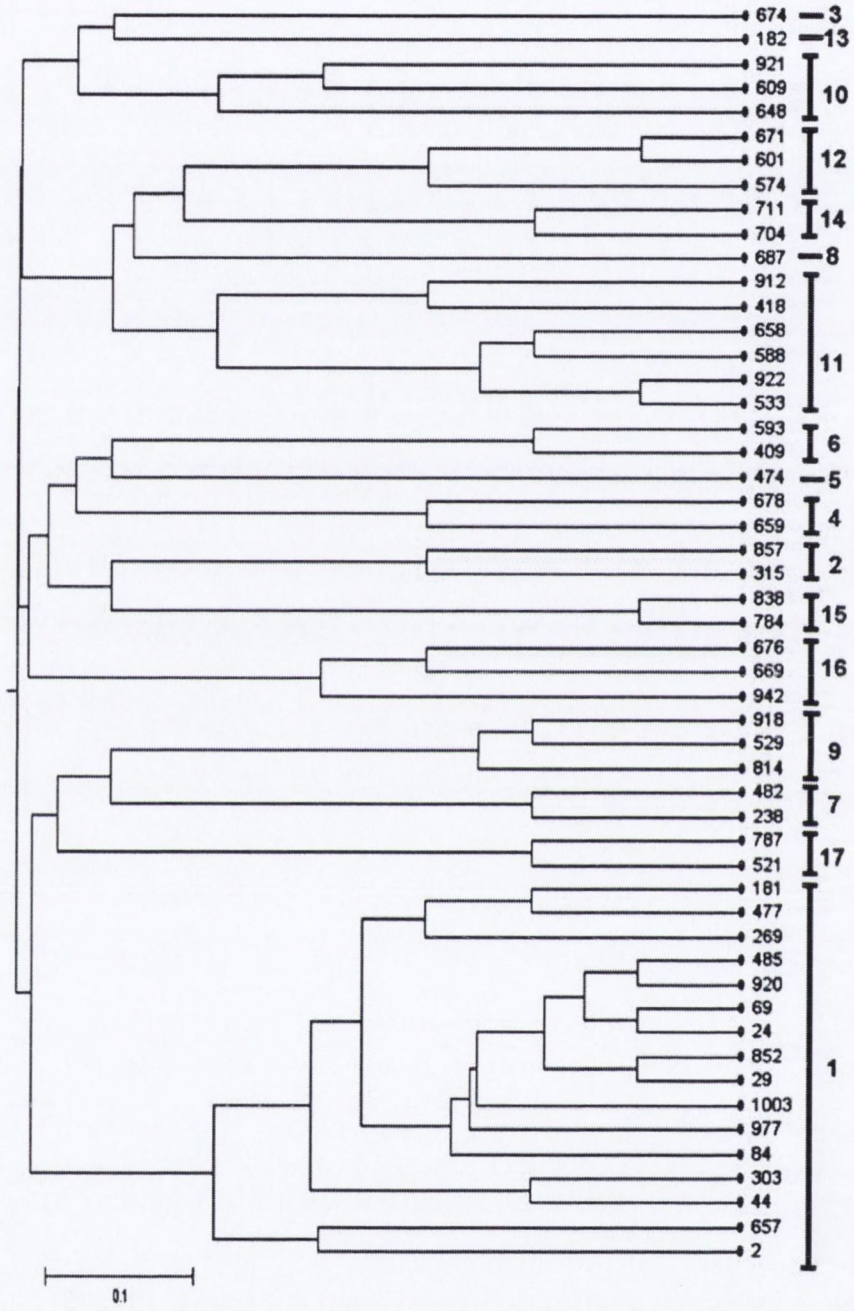


Fig. 2.2 UPGMA dendrogram based on 52 *C. albicans* MLST allelic profiles and their resulting DST numbers. Each of 17 previously defined MLST clades (Odds et al. 2007) are represented and display the genetic relatedness between DSTs in different MLST clades. MLST clades numbers are displayed in bold typeface adjacent to corresponding DST numbers. This dendrogram was generated using START2 software (http://www.ncbi.nlm.nih.gov/pubmed/11751234?dopt=Abstract)

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

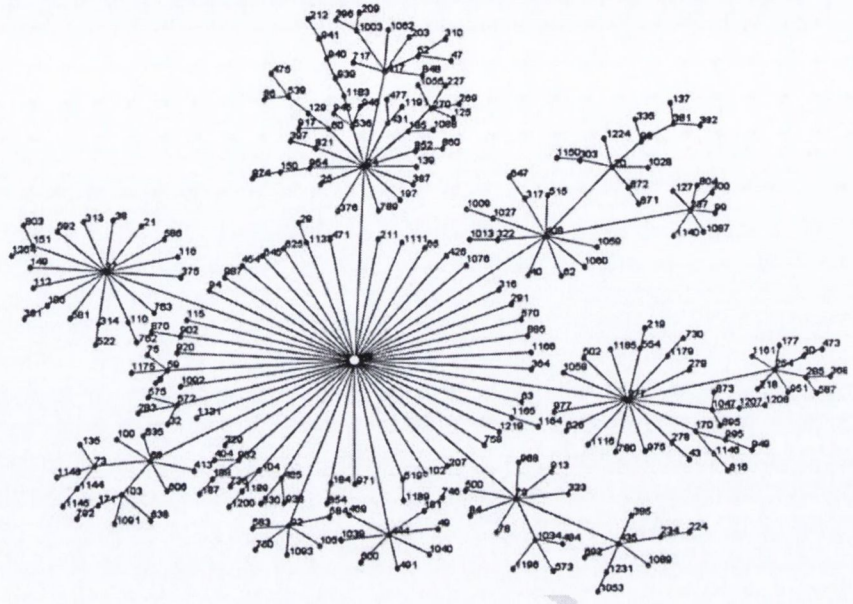


Fig. 2.3 Example of a *C. albicans* clonal cluster generated with eBURST software version 3.0. The primary founder of the clonal cluster (DST 69) is displayed in white, and DSTs that differ by one of the seven MLST loci (i.e. SLVs) are linked to the primary founder. The lengths of the linkages are not significant. Subgroup founders are also SLVs of DST 69, and are further linked to double locus variants (DLVs) of the primary founder DST

to their close proximity. One of the most striking findings of MLST analysis of *C. albicans* is the tendency of isolates from similar geographic locations to cluster within the same clade (Odds et al. 2007). Clades enriched with isolates from the UK, continental Europe and Asia can be discriminated. However, geographic delineations were not absolute, as would be expected due to movement and migration of human and animal populations. The most common, globally distributed *C. albicans* strain types are those of MLST clade 1. More interesting associations between clade structure and isolate source could be inferred when European isolates were analysed in isolation, thus removing geographical bias from the analysis (Odds et al. 2007). This analysis found that the majority of European clade 1 isolates were commonly associated with commensalism and with superficial infection rather than systemic disease. The ubiquity of these isolates in the human population suggests that they may have evolved characteristics that make them highly efficient colonisers of human mucosal surfaces, perhaps analogous to the 'general purpose genotype' proposed by Schmid et al. (1999). In contrast, clade 4 isolates were significantly enriched with isolates recovered from blood culture (Odds et al. 2007).

C. albicans cells of opposite mating types (i.e. homozygous at the mating locus) have been demonstrated to undergo a process similar to mating in *Saccharomyces cerevisiae* (Hull et al. 2000; Magee and Magee 2000). However, although tetraploid progeny have been generated in vitro and during in vivo experiments, meiosis or

reductive cell division has not been described. The debate on whether natural populations of *C. albicans* undergo mating continues, and MLST has provided significant evidence that mating is, at the very least, extremely rare. MLST can provide data useful for investigating the mating structure of a population. The DST generated by MLST allows one to generate a sequence type for each individual diploid allele, termed the haplotype (Tavanti et al. 2004). Haplotypes allow investigation of allele frequencies in a population and may provide evidence for sexual reproduction. For example, in a sexually reproducing population with random mating, the frequencies of these haplotypes should be in Hardy–Weinberg (H–W) equilibrium due to the random assortment of pairs of haplotypes in diploid cells (Tavanti et al. 2004). Initial analysis of haplotypes generated from *C. albicans* MLST data suggested that the *C. albicans* population may undergo sexual or parasexual reproduction (Odds et al. 2007; Tavanti et al. 2004). Odds et al. (2007) found that some combinations of haplotypes were in H–W equilibrium, providing evidence of chromosomal segregation or intrachromosomal recombination and concluded that although largely clonal, *C. albicans* populations may rarely undergo sexual reproduction (Odds et al. 2007). More recently, Bougnoux et al. (2008) analysed the haplotypes of a larger group of *C. albicans* isolates. This larger group of isolates allowed them to test the hypothesis that mating may only occur in closely related isolates, i.e. between isolates of the same clade. In contrast to other studies, polymorphic nucleotide sites were found to be in H–W disequilibrium with an excess of heterozygotes. The authors concluded that mating within clades in *C. albicans* must be extremely rare. Previous studies have analysed allele frequencies in disparate isolates (i.e. isolates from multiple MLST clades) from different clonal lineages and this may have given the appearance of high levels of recombination, and therefore, mating in the *C. albicans* population. The study of haplotypes also revealed that loss of heterozygosity was a common phenomenon in *C. albicans*, however, selective pressure maintained an excess of heterozygosity. The authors suggest that the excess of heterozygosity is globally maintained as it may mask deleterious alleles and that the maintenance of alternative alleles may confer a selective advantage (Bougnoux et al. 2008).

2.2.2.3 Epidemiological Investigations with MLST

MLST has confirmed much of the existing data regarding strain carriage in *C. albicans*, confirming that strain maintenance, strain replacement and microevolution can occur within an individual (Bougnoux et al. 2006; Odds et al. 2006). Recently, MLST revealed a high incidence of multiple strains of *C. albicans* in samples from healthy individuals (Jacobsen et al. 2008b). Studies have also provided evidence for microevolution through frequent loss of heterozygosity by either chromosome loss or mitotic recombination (Odds et al. 2006).

MLST has recently been applied to analyse nosocomial transmission of *C. albicans* in an intensive care unit in a large UK teaching hospital (Cliff et al. 2008). This study provided evidence for an endemic strain corresponding to DST69,

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

which was recovered from patients and from health-care workers. However, DST69 is the most common *C. albicans* DST in the MLST database, making it difficult to determine whether transmission of this strain between individuals has occurred in the study, or whether the individuals were coincidentally colonised by the dominant *C. albicans* DST (Cliff et al. 2008). Perhaps a combination of MLST and fingerprinting with the highly discriminatory C1 probe (a fragment of Ca3) could be applied for greater discriminatory power in characterising nosocomial outbreaks.

Recently, Wrobel et al. (2008) applied MLST to examine the genetic relationships between *C. albicans* isolates recovered from humans and non-migratory wildlife in the same geographic locale (central Illinois, U.S.A.). This study found that the clade distribution of human and wildlife isolates was significantly different, indicating limited strain transfer between the two populations (Wrobel et al. 2008). A similar conclusion was reached by Jacobsen et al. (2008a), who found that strains of *C. albicans* from animal hosts were genetically distinct from those recovered from humans (Jacobsen et al. 2008a).

2.3 MLST of Non-*C. albicans Candida* Species

MLST schemes have now been published for all of the major pathogenic *Candida* species, including *C. dubliniensis* (McManus et al. 2008), *C. tropicalis* (Tavanti et al. 2005b), *C. parapsilosis* (Odds et al. 2007), *C. krusei* (Jacobsen et al. 2007b) and *C. glabrata* (Dodgson et al. 2003) (Table 2.2). As *C. dubliniensis* and *C. albicans* are so closely related, the same loci used in the *C. albicans* scheme can be applied to *C. dubliniensis*. However, the level of sequence polymorphism at these loci was found to be significantly less in *C. dubliniensis*, suggesting that the *C. dubliniensis* population is significantly less divergent (McManus et al. 2008). Fewer than 1% of the bases sequenced in *C. dubliniensis* to date exhibit SNPs, compared to 6% of *C. albicans* bases. A scheme consisting of 8 loci, including two loci of extended length (prefixed 'ex'), have been recommended for maximum discrimination of *C. dubliniensis* isolates (Table 2.2). In an initial study, 50 *C. dubliniensis* isolates were examined and the population structure revealed by MLST confirmed previous findings with the Cd25 fingerprinting probe (Fig. 2.4). However, unlike DNA fingerprint analysis, MLST typing allows one to quantify the level of divergence between *C. albicans* and *C. dubliniensis* populations. By comparing the concatenated sequences for the 8 loci (*AAT1a, ACC1, ADP1, MPIb, SYA1, VPS13, ZWF1b* and *RPN2*) in the recommended *C. dubliniensis* MLST typing scheme, McManus et al. (2008) showed that the two species are separated by 257 bp differences (Fig. 2.4).

Analysis of *C. tropicalis* has revealed a similar clonal population structure to *C. albicans* (Tavanti et al. 2005b). A study that analysed a group of 52 *C. tropicalis* isolates from hospitals in Taiwan identified a clonal cluster consisting of 20 isolates with a high prevalence of reduced fluconazole susceptibility (70%)

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

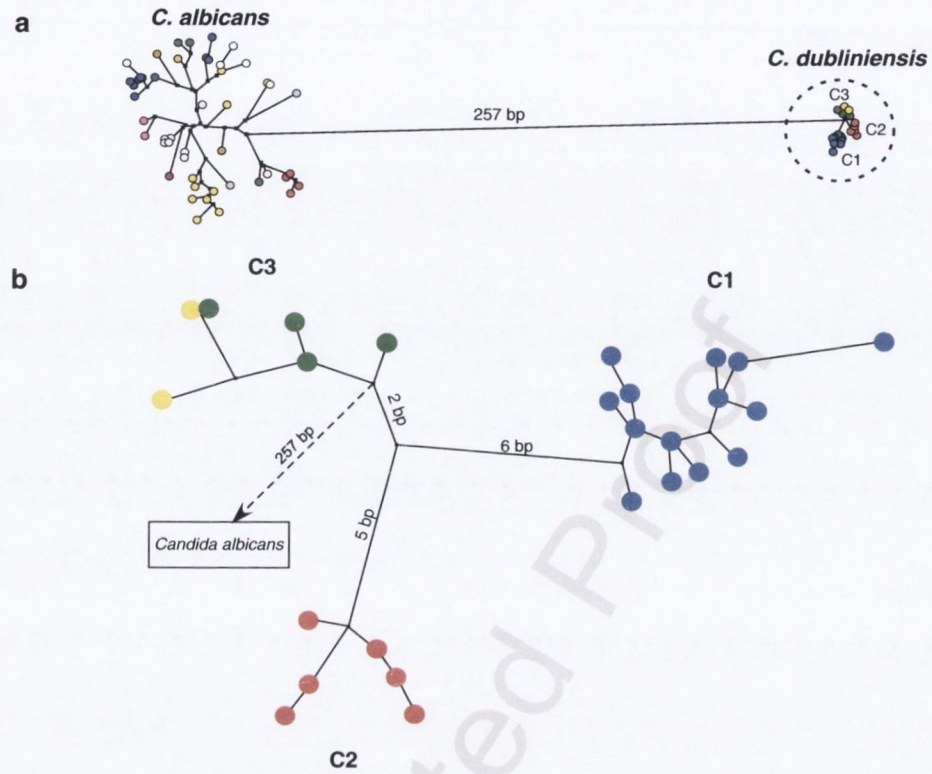


Fig. 2.4 Maximum parsimony tree showing the comparative divergence between 50 isolates each of *C. albicans* and *C. dubliniensis* based on concatenated sequences for the 8 loci (*AAT1a, ACC1, ADP1, MPIb, SYA1, VPS13, ZWF1b* and *RPN2*) in the recommended *C. dubliniensis* MLST typing scheme described by McManus et al. 2008. *C. dubliniensis* isolates were selected from a diverse range of geographic locations and from all four ITS genotypes. *C. albicans* isolates were selected as representatives of the MLST clades described by Odds et al. 2007. Panel (*A*) Comparative divergence between the *C. albicans* and *C. dubliniensis* isolates tested showing that the two species are separated by 257 bp differences. The *C. dubliniensis* isolates formed three closely related groups of isolates or clades (C1–C3). Isolates from distinct clades are highlighted by specific colours for each species. Panel (*B*) shows an enlarged view of the three *C. dubliniensis* major clades encircled in panel (*A*). Clade C1 consists exclusively of ITS genotype 1 isolates (blue), clade C2 consists exclusively of ITS genotype 2 isolates (*red*), and clade C3 consists of ITS genotype 3 isolates (*green*) and ITS genotype 4 isolates (*yellow*). Figure adapted from McManus et al. 2008

(Chou et al. 2007). Attempts to generate an MLST typing scheme for *C. parapsilosis* revealed almost a complete absence of DNA polymorphims within isolates of this species, perhaps indicating a very recent evolutionary divergence for this species (Odds et al. 2007). A scheme for the haploid *C. glabrata* has been developed consisting of 6 loci (Dodgson et al. 2003). Analysis of 109 isolates with this scheme revealed a clonal population structure with several clades exhibiting different geographic specificities. In contrast, analysis of six loci in

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

122 *C. krusei* isolates revealed no evidence for geographical associations with particular subtypes (Jacobsen et al. 2007b).

Analysis of isolates designated as *C. stellatoidea* has revealed interesting findings regarding their position in the current *C. albicans* clade structure (Jacobsen et al. 2007a). *C. stellatoidea* isolates were traditionally identified on the basis of their inability to assimilate sucrose. Type II *C. stellatoidea* are merely sucrose assimilation-negative variants of *C. albicans*. However, the relationship of type I *C. stellatoidea* isolates to *C. albicans* is less clear. Application of the *C. albicans* MLST scheme to four isolates identified as *C. stellatoidea* type I revealed that they clustered with two sucrose negative isolates designated *C. africana* in a group of strains highly distinct from the majority of *C. albicans* strains. These data suggest that *C. stellatoidea* type I may represent a genetically distinct subgroup of *C. albicans* strains (Jacobsen et al. 2007a).

2.4 Future Directions for Typing of *Candida* Species

The *C. albicans* MLST database currently contains data from over 1,500 strains of *C. albicans*. Odds et al. (2007) recently commented that the addition of further *C. albicans* strains to this database is unlikely to reveal anything novel about *C. albicans* population structure. However, the respective databases for the non-*C. albicans Candida* species contain comparatively few strains and continued typing of isolates could reveal new information about the population structures of these species.

At present, MLST is unlikely to be applied to routine screening of clinical isolates as the process is time-consuming and is unlikely to provide data useful to a diagnostic laboratory. Implementation of routine typing for *C. albicans* will depend not only upon the development of cost-effective high-throughput platfoms for SNP analysis but also on the identification of SNPs associated with clinically relevant phenotypic traits (e.g. drug resistance). Microarray technology has the potential for development as a platform for high-through SNP analysis. Lott and Scarborough (2008) recently described an MLST-based SNP microarray for *C. albicans*. The array consisted of oligonucleotide probes specific for 79 SNPs present in 19 discrete loci. One advantage of an array-based platform is the ability to include large numbers of loci without an increase in workload or significant increase in overall cost. The array contains sequences from 12 loci in addition to those in the consensus MLST scheme and includes loci from all 8 *C. albicans* chromosomes. As four of the loci in the consensus MLST scheme are linked (*ADP1* and *ZWF1* on Chr1 and *AAT1* and *PMI1* on Chr2), this leads to a bias in the detection of polymorphisms on these chromosomes. To date, a pilot study analysing 5 isolates has been published and further studies are required in order to determine if this system can offer any new insights into the population structure of *C. albicans* (Lott and Scarborough, 2008). In the future, microarray technology could potentially be useful in the clinical diagnostic laboratory for rapid identification of fungi

from clinical specimens. At present, arrays have been developed for rapid identification of yeasts in clinical specimens using species-specific probes. In order for microarrays to become a routine tool for rapid discrimination of strains in a clinical laboratory environment, they will have to yield clinically relevant data. Arrays have been extensively used to detect anti-microbial resistance genes in prokaryotic organisms and to characterise SNPs associated with anti-malarial drug resistance in *Plasmodium falciparum* (Crameri et al. 2007; Frye et al. 2006.) At present, apart from predicting an isolates susceptibility to 5FC, routine MLST of clinical isolates is of little value to a diagnostic laboratory. However, as our knowledge of anti-fungal resistance mechanisms improves in *C. albicans*, SNP arrays may have potential as diagnostic tools in the mycology laboratory. Certain mutations have been identified in *C. albicans* that are associated with azole resistance, including mutations in the transcriptional activators *TAC1* and *MRR1* and in the gene encoding the target of azole anti-fungal drugs, *ERG11*. An array format could potentially be used to screen for SNPs in these genes-associated azole resistance. The predictive value of such an array is currently unknown as the full range of potential mutations that can result in azole resistance in *C. albicans* are probably not yet known.

As high-throughput, post-Sanger sequencing technologies improve and become more readily available, the possibility of comparing whole genomes of different strains is becoming feasible. Within a matter of years, epidemiological analysis of *C. albicans* populations will involve whole genome comparisons between strains. At present, next-generation sequencing technologies such as Illumina's Solexa system and the 454 Life Sciences GS FLX system have made the goal of 'a genome in a day' achievable (Medini et al. 2008). However, some technical challenges remain before genome sequencing becomes routine. At present, assembly of genome sequences from such short reads (200-400 bp for the GS FLX) is technically challenging and genomes may have large regions that may be difficult to sequence using these technologies (Medini et al. 2008). In addition, genome comparison software tools are at present unable to efficiently compare a large number of genome sequences simultaneously. Once these technical barriers have been overcome, next generation technologies will allow researchers to compare the genomes of isolates from different clades and to identify specific polymorphisms associated with particular phenotypes, such as drug resistance. Association of polymorphisms with meaningful phenotypes such as virulence or drug resistance may make the task of routine detection of polymorphisms by microarray analysis worthwhile.

2.5 Conclusions

Both DNA fingerprint analysis and MLST have provided useful information regarding the epidemiology and population structure of *C. albicans*. Our knowledge of the epidemiology of the non-*C. albicans Candida* species has also improved. At present, MLST of *C. albicans* has probably yielded as much practical information

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

as 'low density' genome sequence analysis will allow. Within the next 10 years, our understanding of *Candida* epidemiology will be revolutionised by high-throughput DNA sequencing technologies. Current data suggests that *C. albicans* isolates belonging to different MLST clades may possess different biological properties, with MLST clade 1 isolates more often associated with mucosal colonisation and superficial infection and clade 4 isolates mostly associated with systemic infection. Generation of whole genome sequence data for multiple isolates in different clades may help to explain some of the biological differences between *C. albicans* isolates that MLST analysis can only suggest.

References

Al Mosaid A, Sullivan DJ, Polacheck I, Shaheen FA, Soliman O, Al Hedaithy S, Al Thawad S, Kabadaya M, Coleman DC (2005) Novel 5-Flucytosine-Resistant Clade of *Candida dubliniensis* from Saudi Arabia and Egypt Identified by Cd25 Fingerprinting. J Clin Microbiol 43:4026–4036

Banerjee SN, Emori TG, Culver DH, Gaynes RP, Jarvis WR, Horan T, Edwards JR, Tolson J, Henderson T, Martone WJ (1991) Secular trends in the epidemiology of nosocomial primary bloodstream infections in the United States, 1980–1989. Am J Med 91(Suppl. 3B):86S–89S

Bougnoux ME, 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

Bougnoux ME, Tavanti A, Bouchier C, Gow NA, Magnier A, Davidson AD, Maiden MC, d'Enfert C, Odds FC (2003) Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. J Clin Microbiol 41:5265–5266

Bougnoux ME, Aanensen DM, Morand S, Theraud M, Spratt BG, d'Enfert C (2004) Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. Infect Genet Evol 4:243–252

Bougnoux ME, Diogo D, Francois N, Sendid B, Veirmeire S, Colombel JF, Bouchier C, Van Kruiningen H, d'Enfert C, Poulain D (2006) Multilocus sequence typing reveals intrafamilial transmission and microevolutions of *Candida albicans* isolates from the human digestive tract. J Clin Microbiol 44:1810–1820

Bougnoux ME, Pujol C, Diogo D, Bouchier C, Soll DR, d'Enfert C (2008) Mating is rare within as well as between clades of the human pathogen *Candida albicans*. Fungal Genet Biol 45:221–231

Cauda R, Tacconelli E, Tumbarello M, Morace G, De Bernardis F, Torosantucci A, Cassone A (1999) Role of protease inhibitors in preventing recurrent oral candidosis in patients with HIV infection: a prospective case-control study. J Acquir Immune Defic Syndr 21:20–25

Chou HH, Lo HJ, Chen KW, Liao MH, Li SY (2007) Multilocus sequence typing of *Candida tropicalis* shows clonal cluster enriched in isolates with resistance or trailing growth of fluconazole. Diagn Microbiol Infect Dis 58:427–433

Cliff PR, Sandoe JA, Heritage J, Brton RC (2008) Use of multilocus sequence typing for the investigation of colonisation by *Candida albicans* in intensive care unit patients. J Hosp Infect 69:24–32

Coleman DC, Bennett DE, Sullivan DJ, Gallagher PJ, Henman MC, Shanley DB, Russell RJ (1993) Oral *Candida* in HIV infection and AIDS: new perspectives/new approaches. Crit Rev Microbiol 19:61–82

Coleman DC, Bennett DE, Gallagher PJ, Flint SR, Nolan A, Mulcahy FM, Sullivan DJ, Henman MC, Russell RJ, Shanley DB (1995) Oral candidiasis and HIV infection: anti-fungal drug resistance and changes in *Candida* population dynamics. In: Greenspan JS, Greenspan D, Greenspan JS, Greenspan D (eds) Oral manifestations of HIV infection. Quintessence publishing Company, Chicago, IL

Coste A, Turner V, Ischer F, Morschhauser J, Forche A, Selmecki A, Berman J, Bille J, Sanglard D (2006) A mutation in Tac1p, a transcription factor regulating CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate anti-fungal resistance in *Candida albicans*. Genetics 172:2139–2156

Crameri A, Marfurt J, Mugittu K, Maire N, Regos A, Coppee JY, Sismeiro O, Burki R, Huber E, Laubscher D, Puijalon O, Genton B, Felger I, Beck HP (2007) Rapid microarray-based method for monitoring of all currently known single-nucleotide polymorphisms associated with parasite resistance to anti-malaria drugs. J Clin Microbiol 45:3685–3691

Dodgson AR, Pujol C, Denning DW, Soll DR, Fox AJ (2003) Multilocus sequence typing of *Candida glabrata* reveals geographically enriched clades. J Clin Microbiol 41:5709–5717

Dodgson AR, Dodgson KJ, Pujol C, Pfaller MA, Soll DR (2004) Clade-specific flucytosine resistance is due to a single nucleotide change in the *FUR1* gene of *Candida albicans*. Antimicrob Agents Chemother 48:2223–2227

Enger L, Joly S, Pujol C, Simonson P, Pfaller M, Soll DR (2001) Cloning and characterization of a complex DNA fingerprinting probe for *Candida parapsilosis*. J Clin Micro 39:658–669

Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186:1518–1530

Fidel PL, Vazquez JA, Sobel JD (1999) *Candida glabrata*: review of epidemiology, pathogenesis, and clinical disease with comparison to *C albicans*. Clin Microbiol Rev 12:80–96

Frye JG, Jesse T, Long F, Rondeau G, Porwollik S, McClelland M, Jackson CR, Englen M, Fedorka-Cray PJ (2006) DNA microarray detection of antimicrobial resistance genes in diverse bacteria. Int J Antimicrob Agents 27:138–151

Hull CM, Raisner RM, Johnson AD (2000) Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. Science 289:307–310

Jacobsen MD, Boekhout T, Odds FC (2007a) Multilocus sequence typing confirms synonymy but highlights differences between *Candida albicans* and Candida stellatoidea. FEMS Yeast Res 5:764–770

Jacobsen MD, Gow NA, Maiden MC, Shaw DJ, Odds FC (2007b) Strain typing and determination of population structure of *Candida krusei* by multilocus sequence typing. J Clin Microbiol 45:317–323

Jacobsen MD, Bougnoux ME, d'Enfert C, Odds FC (2008a) Multilocus sequence typing of *Candida albicans* isolates from animals. Res Microbiol 159:436–440

Jacobsen MD, Duncan AD, Bain J, Johnson EM, Naglik JR, Shaw DJ, Gow NA, Odds FC (2008b) Mixed *Candida albicans* strain populations in colonized and infected mucosal tissues. FEMS Yeast Res 15:15

Joly S, Pujol C, Schroppel K, Soll DR (1996) Development and verification of two species fingerprinting probes for *Candida tropicalis* amenable for computer analysis. J Clin Microbiol 34:3063–3071

Joly S, Pujol C, Rysz M, Vargas K, Soll DR (1999) Development and characterization of complex DNA fingerprinting probes for the infectious yeast *Candida dubliniensis*. J Clin Microbiol 37:1035–1044

Levin AS, Costa SF, Mussi NS, Basso M, Sinto SI, Machado C, Geiger DC, Villares MCB, Schreiber AZ, Barone AA, Branchini MLM (1998) *Candida parapsilosis* fungemia associated with implantable and semi-implatable central venous catheters and the hands of healthcare workers. Diagn Microbiol Infect Dis 30:243–249

Levy I, Rubin LG, Vasishtha S, Tucci V, Sood SK (1998) Emergence of *Candida parapsilosis* as the predominant species causing candidemia in children. Clin Infect Dis 26:1086–1088

Lockhart SR, Reed BD, Pierson CL, Soll DR (1996) Most frequent scenario for recurrent *Candida* vaginitis is strain maintenance with "substrain shuffling": demonstration by sequential DNA fingerprinting with probes Ca3, C1, and CARE2. J Clin Microbiol 34:767–777

Lockhart SR, Joly S, Pujol C, Sobel JD, Pfaller MA, Soll DR (1997) Development and verification of fingerprinting probes for Candida glabrata. Microbiology 143:3733–3746

Lockhart S, Joly S, Vargas K, Swails-Wenger J, Enger L, Soll DR (1998) Natural defenses against *Candida* colonization break down in the oral cavity of the elderly. J Dent Res 78:857–868

Lott TJ, Scarborough RT (2008) Development of a MLST-biased SNP microarray for *Candida albicans*. Fungal Genet Biol 45:803–811

Magee BB, Magee PT (2000) Induction of mating in *Candida albicans* by construction of MTLa and MTLalpha strains. Science 289:310–313

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (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

Manfredi M, McCullough MJ, Al-Karaawi ZM, Hurel SJ, Porter SR (2002) The isolation, identification and molecular analysis of *Candida spp*. isolated from the oral cavities of patients with diabetes mellitus. Oral Microbiol Immunol 17:181–185

Marco F, Lockhart SR, Pfaller MA, Pujol C, Rangel-Frausto MS, Wiblin T, Blumberg HM, Edwards JE, Jarvis W, Saiman L, Patterson JE, Rinaldi MG, Wenzel RP, Soll DR (1999) Elucidating the origins of nosocomial infections with *Candida albicans* by DNA fingerprinting with the complex probe Ca3. J Clin Microbiol 37:2817–2828

Martin GS, Mannino DM, Eaton S, Moss M (2003) The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med 348:1546–1554

McManus BA, Coleman DC, Moran G, Pinjon E, Diogo D, Bougnoux ME, Borecka-Melkusova S, Bujdakova H, Murphy P, d'Enfert C, Sullivan DJ (2008) Multilocus sequence typing reveals that the population structure of Candida dubliniensis is significantly less divergent than that of *Candida albicans*. J Clin Microbiol 46:652–664

Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R (2008) Microbiology in the post-genomic era. Nat Rev Microbiol 6:419–430

Moran GP, Sullivan DJ, Coleman DC (2002) Emergence of non-*Candida albicans Candida* species as pathogens. In: Calderone RA (ed) Candida and Candidiasis. ASM Press, Washington, DC, pp 37–53

Nguyen MH, Peacock JE, Morris AJ, Tanner DC, Nguyen ML, Snydman DR, Wagener MM, Rinaldi MG, Yu LY (1996) The changing face of candidemia: emergence of non-*Candida albicans* species and antifungal resistance. Am J Med 100:617–623

Odds FC, Jacobsen MD (2008) Mltilocus sequence typing of pathogenic *Candida* species. Eukaryot Cell 7:1075–1084

Odds FC, Davidson AD, Jacobsen MD, Tavanti A, Whyte JA, Kibbler CC, Ellis DH, Maiden MC, Shaw DJ, Gow NA (2006) *Candida albicans* strain maintenance, replacement, and microvariation demonstrated by multilocus sequence typing. J Clin Microbiol 44:3647–3658

Odds FC, Bougnoux ME, Shaw DJ, Bain JM, Davidson AD, Diogo D, Jacobsen MD, Lecomte M, Li SY, Tavanti A, Maiden MC, Gow NA, d'Enfert C (2007) Molecular phylogenetics of *Candida albicans*. Eukaryot Cell 6:1041–1052

Oh SH, Cheng G, Nuessen JA, Jajko R, Yeater KM, Zhao X, Pujol C, Soll DR, Hoyer LL (2005) Functional specificity of *Candida albicans* Als3p proteins and clade specificity of *ALS3* alleles discriminated by the number of copies of the tandem repeat sequence in the central domain. Microbiology 151:673–681

Pfaller MA (1995) Epidemiology of candidiasis. J Hosp Infect 30:329–338

Pfaller MA, Diekema DJ (2004) Twelve years of fluconazole in clinical practice: global trends in species distribution and fluconazole susceptibility of bloodstream isolates of *Candida*. Clin Microbiol Infect 10(Suppl 1):11–23

Pfaller MA, Diekema DJ (2007) Epidemiology of invasive candidiasis: a persistent public health problem. Clin Microbiol Rev 20:133–163

Pfaller MA, Cabezudo I, Hollis R, Huston B, Wenzel RP (1990) The use of biotyping and DNA fingerprinting in typing *Candida albicans* from hospitalised patients *Diag*. Microbiol Infect Dis 13:481–489

Pfaller MA, Boyken L, Hollis RJ, Messer SA, Tendolkar S, Diekema DJ (2006) In vitro susceptibilities of *Candida* spp. to caspofungin: four years of global surveillance. J Clin Microbiol 44:760–763

Pittet D, Monod M, Suter PM, Frenk E, Auckenthaler R (1994) *Candida* colonization and subsequent infections in critically ill surgical patients. Ann Surg 220:751–758

Pujol C, Pfaller M, Soll DR (2002) Ca3 fingerprinting of *Candida albicans* bloodstream isolates from the United States, Canada, South America, and Europe reveals a European clade. J Clin Microbiol 40:2729–2740

Pujol C, Pfaller MA, Soll DR (2004) Flucytosine resistance is restricted to a single genetic clade of *Candida albicans*. Antimicrob Agents Chemother 48:262–266

Ruhnke M (2006) Epidemiology of *Candida albicans* infections and role of non-Candida-albicans yeasts. Curr Drug Targets 7:495–504

Samaranayake YH, Samaranayake LP (1994) *Candida krusei*: biology, epidemiology, pathogenicity and clinical manifestations of an emerging pathogen. J Med Microbiol 41:295–310

Sangeorzan JA, Bradley SF, He X, Zarins LT, Ridenour GL, Tiballi RN, Kauffman CA (1994) Epidemiology of oral candidiasis in HIV-infected patients: colonisation, infection, treatment, and emergence of fluconazole-resistance. Am J Med 97:339–346

Schmid J, Voss E, Soll DE (1990) Computer-assisted methods for assessing strain relatedness in *Candida albicans* by fingerprinting with the moderately repetitive sequence Ca3. J Clin Microbiol 28:1236–1243

Schmid J, Herd S, Hunter PR, Cannon RD, Yasin MS, Samad S, Carr M, Parr D, McKinney W, Schousboe M, Harris B, Ikram R, Harris M, Restrepo A, Hoyos G, Singh KP (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

Schroppel K, Rotman M, Galask R, Mac K, Soll DR (1994) Evolution and replacement of *Candida albicans* strains during recurrent vaginitis demonstrated by DNA fingerprinting. J Clin Microbiol 32:2646–2654

Sobel JD (2007) Vulvovaginal candidosis. Lancet 369:1961–1971

Soll DR (2000) The ins and outs of DNA fingerprinting the infectious fungi. Clin Microbiol Rev 13:332–370

Sullivan DJ, Coleman DC (2002) Molecular approaches to identification and typing of *Candida* species. In: Caldeone RA (ed) Candida and candidiasis. ASM Press, Washington, DC

Sullivan D, Bennett D, Henman M, Harwood P, Flint S, Mulcahy F, Shanley D, Coleman D (1993) Oligonucleotide fingerprinting of isolates of Candida species other than *C. albicans* and of atypical *Candida* species from human immunodeficiency virus-positive and AIDS patients. J Clin Microbiol 31:2124–2133

Tavanti A, Gow NA, Senesi S, Maiden MC, Odds FC (2003) Optimization and validation of multilocus sequence typing for *Candida albicans*. J Clin Microbiol 41:3765–3776

Tavanti A, Gow NA, Maiden MC, Odds FC, Shaw DJ (2004) Genetic evidence for recombination in *Candida albicans* based on haplotype analysis. Fungal Genet Biol 41:553–562

Tavanti A, Davidson AD, Fordyce MJ, Gow NA, Maiden MC, Odds FC (2005a) Population structure and properties of *Candida albicans*, as determined by multilocus sequence typing. J Clin Microbiol 43:5601–5613

Tavanti A, Davidson AD, Johnson EM, Maiden MC, Shaw DJ, Gow NA, Odds FC (2005b) Multilocus sequence typing for differentiation of strains of *Candida tropicalis*. J Clin Microbiol 43:5593–5600

Taylor BN, Harrer T, Pscheidl E, Schweizer A, Rollinghoff M, Schroppel K (2003) Surveillance of nosocomial transmission of *Candida albicans* in an intensive care unit by DNA fingerprinting. J Hosp Infect 5:283–289

Vargas KG, Joly S (2002) Carriage frequency, intensity of carriage, and strains of oral yeast species vary in the progression to oral candidiasis in human immunodeficiency virus-positive individuals. J Clin Microbiol 40:341–350

BookID 112284_Ashbee_ChapID 2_Proof# 1 - 30/7/09

Warnock DW, Burke J, Cope NJ, Johnson EM, Von Fraunhoffer NA, Williams EW (1988) Fluconazole resistance in Candida glabrata. Lancet 2:1310

Wenzel RP (1995) Nosocomial candidemia: risk factors and attributable mortality. Clin Infect Dis 20:1531–1534

Wingard JR, Merz WG, Rinaldi MG, Johnson TR, Karp JE, Saral R (1991) Increase in *Candida krusei* infection among patients with bone marrow transplantation and neutropenia treated prophylactically with fluconazole. New Eng J Med 325:1274–1277

Wrobel L, Whittington JK, Pujol C, Oh SH, Ruiz MO, Pfaller MA, Diekema DJ, Soll DR, Hoyer LL (2008) Molecular phylogenetic analysis of a geographically and temporally matched set of *Candida albicans* isolates from humans and nonmigratory wildlife in central Illinois. Eukaryot Cell 7:1475–1486

Zhang N, Harrex AL, Holland BR, Fenton LE, Cannon RD, Schmid J (2003) Sixty alleles of the ALS7 open reading frame in *Candida albicans*: *ALS7* is a hypermutable contingency locus. Genome Res 13:2005–2017

Appendices

Appendix A. Allelic profiles and DST numbers for MLST Scheme A

Isolate	*AAT1a*	*ACC1*	*ADP1*	*PMIb*	*ALA1*	*VPS13*	*ZWF1b*	DST
CD36	1	1	1	4	5	1	1	6
CD06033	1	1	1	4	4	1	1	4
CD06031	1	1	1	4	4	1	1	4
CD060213	1	1	1	4	4	1	1	4
CD06041	1	1	1	4	4	1	1	4
CD06038	1	1	1	4	4	1	1	4
CD06045	1	1	1	4	4	1	1	4
CD06037	1	3	3	7	2	1	2	19
CD06036	1	3	1	7	2	1	2	14
CD060215	1	2	1	1	2	2	3	12
CD604	1	1	1	7	4	1	1	9
CD603	1	1	1	1	4	1	1	2
CM1	1	1	1	1	4	1	1	2
SA105	1	1	1	6	4	1	1	8
SA115	1	1	1	1	4	1	1	2
SA116	1	1	1	4	4	1	1	4
SA108	1	2	6	2	2	2	4	13
SA100	1	2	6	2	2	2	4	13
SA121	1	2	6	2	2	2	4	13
Eg202	1	2	6	2	2	2	4	13
Eg203	1	1	1	3	4	1	1	5
Eg204	1	1	1	5	4	1	1	7
Eg207	2	3	1	2	2	2	5	21
p6785	1	2	1	1	2	1	3	11
p7276	1	2	1	1	2	1	3	11
p7718	1	2	6	2	2	2	4	13
CD71	1	1	1	1	4	1	1	2
CD98923	1	1	1	3	4	1	1	5
B1324	1	1	1	4	4	1	1	4
B341	1	1	1	4	4	1	1	4
PM6-2	1	1	1	4	4	1	1	4
P2	1	1	1	4	4	1	1	4
1504	1	1	1	3	4	1	1	5
8882	1	1	1	4	4	1	1	4
9097	1	1	1	5	4	1	1	7
966(3)/1	1	1	5	3	4	1	1	10
966(3)/2	1	1	5	3	4	1	1	10

Continued overleaf

Appendix A *continued*

Isolate	*AAT1a*	*ACC1*	*ADP1*	*PMIb*	*ALA1*	*VPS13*	*ZWF1b*	DST
CCY29-177-1	1	1	1	4	4	1	1	4
MLNIH0479	1	1	1	3	4	1	1	5
MLNIH0720	1	1	1	2	4	1	1	3
49831	1	4	4	7	2	1	2	20
IFM49883	1	4	4	7	2	1	2	20
IFM0492	1	1	1	4	4	1	1	4
IFM49832	1	1	1	3	4	1	1	5
Can4	1	3	2	2	3	1	2	16
CD539	1	3	2	7	3	1	2	17
CBS2747	1	3	3	2	2	1	2	18
CD514	1	2	1	1	2	2	3	12
CD519	1	2	1	1	2	2	3	12
CD75004	1	3	2	2	2	1	2	15

Appendix B. Allelic profiles and DST numbers for MLST Scheme B

Isolate	*AAT1a*	*ACC1*	*ADP1*	*PMIb*	*ALA1*	*exVPS13*	*exZWF1b*	DST
CD36	1	1	1	4	5	1	1	6
CD06033	1	1	1	4	4	1	1	5
CD06031	1	1	1	4	4	1	1	5
CD060213	1	1	1	4	4	1	1	5
CD06041	1	1	1	4	4	1	1	5
CD06038	1	1	1	4	4	1	1	5
CD06045	1	1	1	4	4	1	1	5
CD06037	1	3	3	7	2	2	2	21
CD06036	1	3	1	7	2	2	2	16
CD060215	1	2	1	1	2	4	3	13
CD604	1	1	1	7	4	1	1	9
CD603	1	1	1	1	4	1	1	2
CM1	1	1	1	1	4	1	1	2
SA105	1	1	1	6	4	1	1	8
SA115	1	1	1	1	4	1	1	2
SA116	1	1	1	4	4	1	1	5
SA108	1	2	6	2	2	3	4	15
SA100	1	2	6	2	2	3	4	15
SA121	1	2	6	2	2	3	4	15
Eg202	1	2	6	2	2	3	4	15
Eg203	1	1	1	3	4	1	1	4
Eg204	1	1	1	5	4	1	1	7
Eg207	2	3	1	2	2	3	5	23
p6785	1	2	1	1	2	1	3	11
p7276	1	2	1	1	2	1	3	11
p7718	1	2	6	2	2	3	4	15
CD71	1	1	1	1	4	1	1	2
CD98923	1	1	1	3	4	1	1	4
B1324	1	1	1	4	4	1	1	5
B341	1	1	1	4	4	1	1	5
PM6-2	1	1	1	4	4	1	1	5
P2	1	1	1	4	4	1	1	5
1504	1	1	1	3	4	1	1	4
8882	1	1	1	4	4	1	1	5
9097	1	1	1	5	4	1	1	7
966(3)/1	1	1	5	3	4	1	1	10
966(3)/2	1	1	5	3	4	1	1	10

Continued overleaf

Appendix B *continued*

Isolate	*AAT1a*	*ACC1*	*ADP1*	*PMIb*	*ALA1*	*exVPS13*	*exZWF1b*	DST
CCY29-177-1	1	1	1	4	4	1	1	5
MLNIH0479	1	1	1	3	4	1	1	4
MLNIH0720	1	1	1	2	4	1	1	3
49831	1	4	4	7	2	2	2	22
IFM49883	1	4	4	7	2	2	2	22
IFM0492	1	1	1	4	4	1	1	5
IFM49832	1	1	1	3	4	1	1	4
Can4	1	3	2	2	3	2	2	18
CD539	1	3	2	7	3	2	2	19
CBS2747	1	3	3	2	2	2	2	20
CD514	1	2	1	1	2	4	6	14
CD519	1	2	1	1	2	2	6	12
CD75004	1	3	2	2	2	2	2	17

Appendix C. Allelic profiles and DST numbers for MLST Scheme C

Isolate	*AAT1a*	*AAT1b*	*ACC1*	*ADP1*	*GLN4*	*PMIb*	*RPN2*	*ALA1*	*exVPS13*	*exZWF1b*	DST
CD36	1	1	1	1	1	4	1	5	1	1	8
CD06033	1	1	1	1	1	4	1	4	1	1	7
CD06031	1	1	1	1	1	4	1	4	1	1	7
CD060213	1	1	1	1	1	4	1	4	1	1	7
CD06041	1	1	1	1	1	4	1	4	1	1	7
CD06038	1	1	1	1	1	4	1	4	1	1	7
CD06045	1	1	1	1	1	4	1	4	1	1	7
CD06037	1	2	3	3	2	7	2	2	2	2	20
CD06036	1	2	3	1	2	7	2	2	2	2	15
CD060215	1	4	2	1	2	1	1	2	4	3	24
CD604	1	1	1	1	1	7	1	4	1	1	12
CD603	1	1	1	1	1	1	3	4	1	1	3
CM1	1	1	1	1	1	1	1	4	1	1	2
SA105	1	1	1	1	1	6	3	4	1	1	11
SA115	1	1	1	1	1	1	1	4	1	1	2
SA116	1	1	1	1	1	4	1	4	1	1	7
SA108	1	1	2	6	2	2	1	2	3	4	14
SA100	1	1	2	6	2	2	1	2	3	4	14
SA121	1	1	2	6	2	2	1	2	3	4	14
Eg202	1	1	2	6	2	2	1	2	3	4	14

Continued overleaf

Appendix C *continued*

Isolate	*AAT1a*	*AAT1b*	*ACC1*	*ADP1*	*GLN4*	*PMIb*	*RPN2*	*ALA1*	*exVPS13*	*exZWF1b*	DST
Eg203	1	1	1	1	1	3	3	4	1	1	6
Eg204	1	1	1	1	1	5	3	4	1	1	10
Eg207	2	3	3	1	2	2	1	2	3	5	27
p6785	1	4	2	1	2	1	1	2	1	3	22
p7276	1	4	2	1	2	1	1	2	1	3	22
p7718	1	1	2	6	2	2	1	2	3	4	14
CD71	1	1	1	1	1	1	1	4	1	1	2
CD98923	1	1	1	1	1	3	1	4	1	1	5
B1324	1	1	1	1	1	4	1	4	1	1	7
B341	1	1	1	1	1	4	1	4	1	1	7
PM6-2	1	1	1	1	1	4	1	4	1	1	7
P2	1	1	1	1	1	4	1	4	1	1	7
1504	1	1	1	1	1	3	1	4	1	1	5
8882	1	1	1	1	1	4	1	4	1	1	7
9097	1	1	1	1	1	5	1	4	1	1	9
966/3 (1)	1	1	1	5	1	3	1	4	1	1	13
966/3 (2)	1	1	1	5	1	3	1	4	1	1	13
CCY29-177-1	1	1	1	1	1	4	1	4	1	1	7
MLNIH0479	1	1	1	1	1	3	3	4	1	1	6
MLNIH0720	1	1	1	1	1	2	1	4	1	1	4
49831	1	2	4	4	2	7	2	2	2	2	21

Continued overleaf

Appendix C *continued*

Isolate	*AAT1a*	*AAT1b*	*ACC1*	*ADP1*	*GLN4*	*PMIb*	*RPN2*	*ALA1*	*exVPS13*	*exZWF1b*	DST
IFM49883	1	5	4	4	2	7	2	2	2	2	26
IFM0492	1	1	1	1	1	4	1	4	1	1	7
IFM49832	1	1	1	1	1	3	1	4	1	1	5
Can4	1	2	3	2	2	2	2	3	2	2	17
CD539	1	2	3	2	2	7	2	3	2	2	18
CBS2747	1	2	3	3	2	2	2	2	2	2	19
CD514	1	4	2	1	2	1	1	2	4	6	25
CD519	1	4	2	1	2	1	1	2	2	6	23
CD75004	1	2	3	2	2	2	2	2	2	2	16

Appendix D. Allelic profiles and DST numbers for MLST Scheme D

Isolate	*AAT1b*	*ACC1*	*ADP1*	*PMIb*	*RPN2*	*ALA1*	*exVPS13*	*exZWF1b*	DST
CD36	1	1	1	4	1	5	1	1	8
CD06033	1	1	1	4	1	4	1	1	7
CD06031	1	1	1	4	1	4	1	1	7
CD060213	1	1	1	4	1	4	1	1	7
CD06041	1	1	1	4	1	4	1	1	7
CD06038	1	1	1	4	1	4	1	1	7
CD06045	1	1	1	4	1	4	1	1	7
CD06037	2	3	3	7	2	2	2	2	19
CD06036	2	3	3	7	2	2	2	2	19
CD060215	4	2	1	1	1	2	4	3	24
CD604	1	1	1	7	1	4	1	1	12
CD603	1	1	1	1	3	4	1	1	3
CM1	1	1	1	1	1	4	1	1	2
SA105	1	1	1	6	3	4	1	1	11
SA115	1	1	1	1	1	4	1	1	2
SA116	1	1	1	4	1	4	1	1	7
SA108	1	2	6	2	1	2	3	4	14
SA100	1	2	6	2	1	2	3	4	14
SA121	1	2	6	2	1	2	3	4	14
Eg202	1	2	6	2	1	2	3	4	14
Eg203	1	1	1	3	3	4	1	1	6
Eg204	1	1	1	5	3	4	1	1	10
Eg207	3	3	1	2	1	2	3	5	21
p6785	4	2	1	1	1	2	1	3	22
p7276	4	2	1	1	1	2	1	3	22
p7718	1	2	6	2	1	2	3	4	14
CD71	1	1	1	1	1	4	1	1	2
CD98923	1	1	1	3	1	4	1	1	5
B1324	1	1	1	4	1	4	1	1	7
B341	1	1	1	4	1	4	1	1	7
PM6-2	1	1	1	4	1	4	1	1	7
P2	1	1	1	4	1	4	1	1	7
1504	1	1	1	3	1	4	1	1	5
8882	1	1	1	4	1	4	1	1	7
9097	1	1	1	5	1	4	1	1	9
966/3 (1)	1	1	5	3	1	4	1	1	13

Continued overleaf

Appendix D *continued*

Isolate	*AAT1b*	*ACC1*	*ADP1*	*PMIb*	*RPN2*	*ALA1*	*exVPS13*	*exZWF1b*	DST
966/3 (2)	1	1	5	3	1	4	1	1	13
CCY29-177-1	1	1	1	4	1	4	1	1	7
MLNIH0479	1	1	1	3	3	4	1	1	6
MLNIH0720	1	1	1	2	1	4	1	1	4
49831	2	4	4	7	2	2	2	2	20
IFM49883	5	4	4	7	2	2	2	2	26
IFM0492	1	1	1	4	1	4	1	1	7
IFM49832	1	1	1	3	1	4	1	1	5
Can4	2	3	2	2	2	3	2	2	16
CD539	2	3	2	7	2	3	2	2	17
CBS2747	2	3	3	2	2	2	2	2	18
CD514	4	2	1	1	1	2	4	6	25
CD519	4	2	1	1	1	2	2	6	23
CD75004	2	3	2	2	2	2	2	2	15
98-277	1	1	1	1	1	4	1	1	2
95-677	1	1	1	4	1	4	1	1	7
94-234	6	1	1	4	1	4	1	1	32
CD505	2	1	1	1	1	1	1	1	28
CD06032	6	1	1	1	1	1	6	1	36
CD06027	1	1	1	1	1	1	1	1	1
CD0512	6	1	1	4	1	1	1	1	37
CD524	2	1	1	4	1	1	1	1	35
17P	2	3	3	7	2	6	2	2	34
18P	2	3	3	7	2	6	2	2	34
22BP	2	3	3	7	2	6	2	2	34
25P	2	3	3	7	2	6	2	2	34
110P	2	3	3	7	2	6	2	2	34
140P	2	3	3	7	2	6	2	2	34
SL411	1	1	1	3	1	4	1	7	27
SL422	1	1	1	3	1	4	1	7	27
SL370	1	1	1	3	1	4	1	7	27
SL410	6	1	1	3	1	4	1	7	29
SL375-I	6	1	1	3	1	4	1	8	31
SL375-II	6	1	1	3	1	4	1	8	31
SL397	6	1	1	3	1	4	1	8	31
SL414	6	1	1	3	1	4	1	8	31

Continued overleaf

Appendix D *continued*

Isolate	*AAT1b*	*ACC1*	*ADP1*	*PMIb*	*RPN2*	*ALA1*	*exVPS13*	*exZWF1b*	DST
SL509	1	1	1	3	1	4	1	8	30
SL495	6	1	7	3	1	4	1	8	33
SL522	6	1	1	3	1	4	1	8	31
AV5	6	1	1	3	1	4	1	7	29
AV6	1	1	1	3	1	4	1	7	27
AV7	1	1	1	1	1	4	1	1	2

Appendix E. Allelic profiles and DST numbers for MLST Scheme E

Isolate	*AAT1a*	*ACC1*	*ADP1*	*PMIb*	*RPN2*	*ALA1*	*VPS13*	*ZWF1b*	DST
CD36	1	1	1	4	1	5	1	1	8
CD06033	1	1	1	4	1	4	1	1	7
CD06031	1	1	1	4	1	4	1	1	7
CD060213	1	1	1	4	1	4	1	1	7
CD06041	1	1	1	4	1	4	1	1	7
CD06038	1	1	1	4	1	4	1	1	7
CD06045	1	1	1	4	1	4	1	1	7
CD06037	1	3	3	7	2	2	1	2	21
CD06036	1	3	3	7	2	2	1	2	21
CD060215	1	2	1	1	1	2	2	3	15
CD604	1	1	1	7	1	4	1	1	12
CD603	1	1	1	1	3	4	1	1	3
CM1	1	1	1	1	1	4	1	1	2
SA105	1	1	1	6	3	4	1	1	11
SA115	1	1	1	1	1	4	1	1	2
SA116	1	1	1	4	1	4	1	1	7
SA108	1	2	6	2	1	2	2	4	16
SA100	1	2	6	2	1	2	2	4	16
SA121	1	2	6	2	1	2	2	4	16
Eg202	1	2	6	2	1	2	2	4	16
Eg203	1	1	1	3	3	4	1	1	6
Eg204	1	1	1	5	3	4	1	1	10
Eg207	2	3	1	2	1	2	2	5	23
p6785	1	2	1	1	1	2	1	3	14
p7276	1	2	1	1	1	2	1	3	14
p7718	1	2	6	2	1	2	2	4	16
CD71	1	1	1	1	1	4	1	1	2
CD98923	1	1	1	3	1	4	1	1	5
B1324	1	1	1	4	1	4	1	1	7
B341	1	1	1	4	1	4	1	1	7
PM6-2	1	1	1	4	1	4	1	1	7
P2	1	1	1	4	1	4	1	1	7
1504	1	1	1	3	1	4	1	1	5
8882	1	1	1	4	1	4	1	1	7
9097	1	1	1	5	1	4	1	1	9
966/3 (1)	1	1	5	3	1	4	1	1	13
966/3 (2)	1	1	5	3	1	4	1	1	13

Continued overleaf

Appendix E *continued*

Isolate	*AAT1a*	*ACC1*	*ADP1*	*PMIb*	*RPN2*	*SYA1*	*VPS13*	*ZWF1b*	DST
CCY29-177-1	1	1	1	4	1	4	1	1	7
MLNIH0479	1	1	1	3	3	4	1	1	6
MLNIH0720	1	1	1	2	1	4	1	1	4
49831	1	4	4	7	2	2	1	2	22
IFM49883	1	4	4	7	2	2	1	2	22
IFM0492	1	1	1	4	1	4	1	1	7
IFM49832	1	1	1	3	1	4	1	1	5
Can4	1	3	2	2	2	3	1	2	18
CD539	1	3	2	7	2	3	1	2	19
CBS2747	1	3	3	2	2	2	1	2	20
CD514	1	2	1	1	1	2	2	3	15
CD519	1	2	1	1	1	2	2	3	15
CD75004	1	3	2	2	2	2	1	2	17

In each MLST Scheme, DST 1 refers to that of the reference *C. dubliniensis* genome sequence available at http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/.

Appendix F. MLST genotypes

> *AAT1a*-1
ATTGAAATGGAATGAAATTCCATTAGCTCCACCAGATAAAATTTTGGGTATT
TCTGAAGCTTATAATAATGATTCTAATCCACAAAAAATTAATTTAGGAGTTG
GTGCTTATAGAGATAATTCTGGTAAACCAATTATTTTCCCATCAGTTAAAAA
AGCTGAAGAAATTTTATTAAATAAAGAAACTGAAAAAGAATATACTGCTAT
TATTGGTTCAAAAAATTTCCAATCAATTGTGAAAAATTTTATTTTTAATAATT
CTAATAAAGATATTAATGGTAAACAATTAATTGATGATAATAGAATTGTTAC
TGCTCAAACCATTTCTGGTACTGGATCACTTAGAGTTATTGCTGATTTTTTAA
ATCGATTT

> *AAT1a*-2
AYTGAAATGGAATGAAATTCCATTAGCTCCACCAGATAAAATTTTGGGTATT
TCTGAAGCTTATAATAATGATTCTAATCCACAAAAAATTAATTTAGGAGTTG
GTGCTTATAGAGATAATTCTGGTAAACCAATTATTTTCCCATCAGTTAAAAA
AGCTGAAGAAATTTTATTAAATAAAGAAACTGAAAAAGAATATACTGCTAT
TATTGGTTCAAAAAATTTCCAATCAATTGTGAAAAATTTTATTTTTAATAATT
CTAATAAAGATATTAATGGTAAACAATTAATTGATGATAATAGAATTGTTAC
TGCTCAAACCATTTCTGGTACTGGATCACTTAGAGTTATTGCTGATTTTTTAA
ATCGATTT

> *AAT1a*-3
ATTGAAATGGAATGAAATTCCATTAGCTCCACCMGATAAAATTTTGGGTAT
TTCTGAAGCTTATAATAATGATTCTAATCCACAAAAAATTAATTTAGGAGTT
GGTGCTTATAGAGATAATTCTGGTAAACCAATTATTTTCCCATCAGTTAAAA
AAGCTGAAGAAATTTTATTAAATAAAGAAACTGAAAAAGAATATACTGCTA
TTATTGGTTCAAAAAATTTCCAATCAATTGTGAAAAATTTTATTTTTAATAAT
TCTAATAAAGATATTAATGGTAAACAATTAATTGATGATAATAGAATTGTTA
CTGCTCAAACCATTTCTGGTACTGGATCACTTAGAGTTATTGCTGATTTTTTA
AATCGATTT

> *AAT1b*-1
TTAACTAAATTGGCTAATGAAAATAAAATCCCTTCATTTGCTTTGTGTCAAT
CATTTGCTAAAAATATGGGACTTTATGGAGAAAGAACTGGATCAATTTCTAT
TATTAATTCATCGAGTGAAGATTCTAAAGCAGTTGAATCTCAATTGAAAAA
ATTAATTAGACCAATTTATTCTTCTCCACCAATTCATGGATCTAAAATTGTTG
AAATTATTTTTGATGAAAATTCTGGTTTATTACCTCAATGGTTAGATGAATT
AGATAAAGTTGTTGGAAGATTAAATACTGTTCGTTCTAAATTATATGAAAAA
TTAGATAAATCTAATTATAATTGGGATCA

> *AAT1b*-2
TTAACTAAATTGGCTAATGAAAATAAAATCCCTTCATTTGCTTTGTGTCAAT
CATTTGCTAAAAATATGGGACTTTATGGAGAAAGAACTGGATCAATTTCTAT
AATTAATTCATCGAGTAAAGATTCTAAAGCAGTTGAATCTCAATTGAAAAA
ATTAATTAGACCAATTTATTCTTCTCCACCAATTCATGGATCTAAAATTGTTG
AAATTATTTTTGATGAAAATTCTGGTTTATTACCTCAATGGTTAGATGAATT
AGATAAAGTTGTTGGAAGATTAAATACTGTTCGTTCTAAATTATATGAAAAA
TTAGATAAATCTAATTATAATTGGGATCA

> *AAT1b*-3
TTAACTAAATTGGCTAATGAAAATAAAATCCCTTCATTTGCTTTGTGTCAAT
CATTTGCTAAAAATATGGGACTTTATGGAGAAAGAACTGGATCAATTTCTAT
TATTAATTCTTCGAGTGAAGATTCTAAAGCAGTTGAATCTCAATTGAAAAAA
TTAATTAGACCAATTTATTCTTCTCCACCAATTCATGGATCTAAAATTGTTGA
AATTATTTTTGATGAAAATTCTGGTTTATTACCTCAATGGTTAGATGAATTA
GATAAAGTTGTTGGAAGATTAAATACTGTTCGTTCTAAATTATATGAAAAAT
TAGATAAATCTAATTATAATTGGGATCA

> *AAT1b*-4
TTAACTAAATTGGCTAATGAAAATAAAATCCCTTCATTTGCTTTGTGTCAAT
CATTTGCTAAAAATATGGGACTTTATGGAGAAAGAACTGGATCAATTTCTAT
TATTAATTCATCGAGTGAAGATTCGAAAGCAGTTGAATCTCAATTGAAAAA
ATTAATTAGACCAATTTATTCTTCTCCACCAATTCATGGATCTAAAATTGTTG
AAATTATTTTTGATGAAaATTCTGGTTTATTACCTCAATGGTTAGATGAATTA
GATAAAGTTGTTGGAAGATTAAATACTGTTCGTTCTAAATTATATGAAAAAT
TAGATAAATCTAATTATAATTGGGATCA

> *AAT1b*-5
TTAACTAAATTGGCTAATGAAAATAAAATCCCTTCATTTGCTTTGTGTCAAT
CATTTGCTAAAAATATGGGACTTTATGGAGAAAGAACTGGATCAATTTCTAT
AATTAATTCATCGAGTRAAGATTCTAAAGCAGTTGAATCTCAATTGAAAAA
ATTAATTAGACCAATTTATTCTTCTCCACCAATTCATGGATCTAAAATTGTTG
AAATTATTTTTGATGAAAATTCTGGTTTATTACCTCAATGGTTAGATGAATT
AGATAAAGTTGTTGGAAGATTAAATACTGTTCGTTCTAAATTATATGAAAAA
TTAGATAAATCTAATTATAATTGGGATCA

> *AAT1b*-6
TTAACTAAATTGGCTAATGAAAATAAAATCCCTTCATTTGCTTTGTGTCAAT
CATTTGCTAAAAATATGGGACTTTATGGAGAAAGAACTGGATCAATTTCTAT
TATTAATTCATCGAGTGAAGATTCTAAAGCAGTTGAATCTCAATTGAAAAA
ATTAATTAGACCAATTTATTCTTCTCCACCAATTCATGGATCKAAAATTGTT
GAAATTATTTTTGATGAAAATTCTGGTTTATTACCTCAATGGTTAGATGAAT
TAGATAAAGTTGTTGGAAGATTAAATACTGTTCGTTCTAAATTATATGAAAA
ATTAGATAAATCTAATTATAATTGGGATCA

> *ACC1*-1
TTTTGAGATCATCTGTTGTTCAAACCTCGTATGGTGAAATTTTTGCTAAACAT
AGAGAACCAAATTTGGAAATTATTCGTGAAGTTGTTGATTCCAAACACATTG
TTTTTGATGTGTTGTCACAATTTTTGATCAATCCAGATCCATGGGTTGCCATT
GCTGCAGCTGAAGTGTATGTCAGACGTTCATACCGTGCTTATGATTTGGGTA
AAATTGAATATCATGTTAACGACAGACTTCCTATTGTTGAATGGAAATTCAA
GTTAGCTAGTATGGGAGCCGCTGGAGTAAATGATGCTCAACAGGCTGCTGT
TGCTGGTGGTGATGACTCAACATCAATGAAACATGCTGCTTCAGTTTCTGAT
TTGACCTTTGTTGTTGATTCCAAAACCGAGCATACTACAAGA

> *ACC1*-2
TTTTGAGATCATCTGTTGTTCAAACCTCGTATGGTGAAATTTTTGCTAAACAT
AGAGAACCAAATTTGGAAATTATTCGTGAAGTTGTTGATTCCAAACACATTG
TTTTTGATGTGTTGTCACAATTCTTGATCAATCCAGATCCATGGGTTGCCATT
GCTGCAGCTGAAGTGTATGTCAGACGTTCATACCGTGCTTATGATTTGGGTA
AAATTGAATATCATGTTAACGACAGACTTCCTATTGTTGAATGGAAATTCAA
GTTAGCTAGTATGGGAGCCGCTGGAGTAAATGATGCTCAACAGGCTGCTGC
TGCTGGTGGTGATGACTCAACATCAATGAAACATGCTGCTTCAGTTTCTGAT
TTGACCTTTGTTGTTGATTCCAAAACCGAGCATACTACAAGA

> *ACC1*-3
TTTTGAGATCATCTGTTGTTCAAACCTCGTATGGTGAAATTTTTGCTAAACAT
AGAGAACCAAATTTGGAAATTATTCGTGAAGTTGTTGATTCCAAACACATTG
TTTTTGATGTGTTGTCACAATTTTTGATCAATCCAGATCCATGGGTTGCCATT
GCTGCAGCTGAAGTGTATGTCAGACGTTCATACCGTGCTTATGATTTGGGTA
AAATTGAATATCATGTTAACGACAGACTTCCTATTGTTGAATGGAAATTCAA
GTTAGCTAGTATGGGAGCCGCTGGAGTAAATGATGCTCAACAGGCTGCTGC
TGCTGGTGGTGATGACTCAACATCAATGAAACATGCTGCTTCAGTTTCTGAT
TTGACCTTTGTTGTTGATTCCAAAACCGAGCATACTACAAGA

> *ACC1*-4
TTTTGAGATCATCTGTTGTTCAAACCTCGTATGGTGAAATTTTTGCTAAACAT
AGAGAACCAAATTTGGAAATTATTCGTGAAGTTGTTGATTCCAAACACATTG
TTTTTGATGTGTTGTCACAATTTTTGATCAATCCAGATCCATGGGTTGCCATT
GCTGCAGCTGAAGTGTATGTCAGACGTTCATACCGTGCTTATGATTTGGGTA
AAATTGAATATCATGTTAACGACAGACTTCCTATTGTTGAATGGAAATTCAA
GTTAGCTAGTATGGGAGCCGCTGGAGTAAATGATGCTCAACAGGCTGCTGC
TGCTGGTGGTGATGACTCAACATCAATGAAACATGCTGCTTCAGTTTCTGAT
TTGACCTTTGTTGTTGATTCCAATACCGAGCATACTACAAGA

> *ADP1*-1
TACGTTGCAATGTAAATCAGGTGAGTGTATCCATAAAAGTGAAATACCAGG
TTACACTATCCCAGAGAATCCAAAGTTCACTCTTGGTAATTTGTTTGTGATA
ATTGGGGTCTTGTTGGTTTGTGTTTTGGCAGTTGTCTCTCTCTTGAGAAATAT
TAGTGAGTCGGCCTTGTTCAAGAAAAATGGGTATGAGCCATTGGATTCAGA
TCCTAGCGTCATGAATCAAAACTTTGAGCCTACAACATTGTCGTTTGAAGAT
ATAAAGTATGAGGTTGCTGGTGGTCGACAAATTTTGAATGGAGTCTTTGGGT
TTGTAAAACCAAGAGAATGTTTGGCAATAATGGGGGGTTCGGGTGCTGGTA
AAACTACGTTATTGGATATTTTGGCCGGTAAGAACAAGGATGGTATGATTA
ATGGATCGATTTATGTTAATGGAAATCCAA

> *ADP1*-2
TACGTTGCAATGTAAATCAGGTGAGTGTATTCATAAAAGTGAAATACCAGG
TTACACTATCCCAGAGAATCCAAAGTTCACTCTTGGTAATTTGTTTGTGATA
ATTGGGGTCTTGTTGGTTTGTGTTTTGGCAGTTGTCTCTCTCTTGAGAAATAT
TAGTGAGTCGGCCTTGTTCAAGAAAAATGGGTATGAGCCATTGGATTCAGA
TCCTAGTGTCATGAATCAAAACTTTGAGCCTACAACATTGTCGTTTGAAGAT
ATAAAGTATGAGGTTGCTGGTGGTCGACAAATTTTGAATGGAGTCTTTGGGT
TTGTAAAACCAAGAGAATGCTTGGCAATAATGGGGGGTTCGGGTGCTGGTA

AAACTACGTTATTGGATATTTTGGCCGGTAAGAACAAGGATGGTATGATTA
ATGGATCGATTTATGTTAATGGAAATCCAA

> *ADP1*-3
TACGTTGCAATGTAAATCAGGTGAGTGTATCCATAAAAGTGAAATACCAGG
TTACACTATCCCAGAGAATCCAAAGTTCACTCTTGGTAATTTGTTTGTGATA
ATTGGGGTCTTGTTGGTTTGTGTTTTGGCAGTTGTCTCTCTCTTGAGAAATAT
TAGTGAGTCGGCCTTGTTCAAGAAAAATGGGTATGAGCCATTGGATTCAGA
TCCTAGYGTCATGAATCAAAACTTTGAGCCTACAACATTGTCGTTTGAAGAT
ATAAAGTATGAGGTTGCTGGTGGTCGACAAATTTTGAATGGAGTCTTTGGGT
TTGTAAAACCAAGAGAATGYTTGGCAATAATGGGGGGTTCGGGTGCTGGTA
AAACTACGTTATTGGATATTTTGGCCGGTAAGAACAAGGATGGTATGATTA
ATGGATCGATTTATGTTAATGGAAATCCAA

> *ADP1*-4
TACGTTGCAATGTAAATCAGGTGAGTGTATCCATAAAAGTGAAATACCAGG
TTACACTATCCCAGAGAATCCAAAGTTCACTCTTGGTAATTTGTTTGTGATA
ATTGGGGTCTTGTTGGTTTGTGTTTTGGCAGTTGTCTCTCTCTTGAGAAATAT
TAGTGAGTCGGCCTTGTTCAAGAAAAATGGGTATGAGCCATTGGATTCAGA
TCCTAGTGTCATGAATCAAAACTTTGAGCCTACAACATTGTCGTTTGAAGAT
ATAAAGTATGAGGTTGCTGGTGGTCGACAAATTTTGAATGGAGTCTTTGGGT
TTGTAAAACCAAGAGAATGCTTGGCAATAATGGGGGGTTCGGGTGCTGGTA
AAACTACGTTATTGGATATTTTGGCCGGTAAGAACAAGGATGGTATGATTA
ATGGATCGATTTATGTTAATGGAAATCCAA

> *ADP1*-5
TACGTTGCAATGTAAATCAGGTGAGTGTATCCATAAAAGTGAAATACCAGG
TTACACTATCCCAGAGAATCCAAAGTTCACTCTTGGTAATTTGTTTGTGATA
ATTGGGGTCTTGTTGGTTTGTGTTTTGGCAGTTGTCTCTCTCTTGAGAAATAT
TAGTGAGTCGGCCTTGTTCAAGAAAAATGGGTATGAGCCATTGGATTCAGA
TCCTAGCGTCATGAATCAAAACTTTGGGCCTACAACATTGTCGTTTGAAGAT
ATAAAGTATGAGGTTGCTGGTGGTCGACAAATTTTGAATGGAGTCTTTGGGT
TTGTAAAACCAAGAGAATGTTTGGCAATAATGGGGGGTTCGGGTGCTGGTA
AAACTACGTTATTGGATATTTTGGCCGGTAAGAACAAGGATGGTATGATTA
ATGGATCGATTTATGTTAATGGAAATCCAA

> *ADP1*-6
TACGTTGCAATGTAAATCAGGTGAGTGTATCCATAAAAGTGAAATACCAGG
TTACACTATCCCAGAGAATCCAAAGTTCACTCTTGGTAATTTGTTTGTGATA
ATTGGGGTCTTGTTGGTTTGTGTTTTGGCAGTTGTCTCTCTCTTGAGAAATAT
TAGTGAGTCGGCCTTGTTCAAGAAAAATGGGTATGAGCCATTGGATTCAGA
TCCTAGCGTCATGAATCAAAACTTTGAGCCTACAACATTGTCGTTTGAAGAT
ATAAAGTATGAGGTTGCTGGTGGTCGACAAATTTTGAATGGAGTCTTTGGGT
TTGTAAAACCAAGAGAATGTTTGGCAATAATGGGGGGTTCGGGTGCTGGTA
AAACTACGTTATTGGATATTTTGGCCGGTAAGAACAAGGATGGTATGATTA
ATGGATCTATTTATGTTAATGGAAATCCAA

> *ADP1*-7
TACGTTGCAATGTAAATCAGGTGAGTGTATCCATAAAAGTGAAATACCAGG
TTACACTATCCCAGAGAATCCAAAGTTCACTCTTGGTAATTTGTTTGTGATA
ATTGGGGTCTTGTTGGTTTGTGTTTTGGCAGTTGTCTCTCTCTTGAGAAATAT
TAGTGAGTCGGCCTTGTTCAAGAAAAATGGGTATGAGCCATTGGATTCAGA
TCCTAGCGTCATGAATCAAAACTTTGAGCCTACAACATTGTCGTTTGAAGAT
ATAAAGTATGAGGTTGCTGGTGGTCGACAAATTTTGAATGGAGTCTTTGGGT
TTGTAAAACCAAGAGAATGTTTGGCAATAATGGGGGGTTCGGGTGCTGGTA
AAACTACGTTATTGGATATTTTGGCCGGTAAGAACAAGGATGGTATGATTA
ATGGATCGATTTATGTTAATGGAAATCCAA

> *GLN4*-1
TCTGCTTTATACAATATTTTAGGCACAAACTTCCCCAAAACCGATGATAAGA
AATTATCTTTGTTACATCAATTGGCTATACATGAAAGCAAAAATGGAGAGG
AACCAAACCATGACTTTGTCATTAATGGAATTCAAAATGGAGATTTGAAAA
CTGCTTTGCAAGTGACTGAAGGTATCAAGTATTTACAGAATAACACAACTGT
TGCTAAAGAAAAATTCGATGAAGCTAGTGGTGTGGGTATTGAAATTACCCC
TCAACAAGCCAAGGCAGAAATTTCAAAATACCTTGATTCTATTAAAACTGA
TTTGGAAAGTAAAAGATATTCAATATTACCTAAGGTTTTGGGTGAAGTGAA
AACTCAGCCATCCTTGAAATGGGCCCCACCTCAATTATTCAAACC

> *GLN4*-2
TCTGCTTTATACAATATTTTAGGCACAAACTTCCCCAAAACCGATGATAAGA
AATTATCTTTGTTACATCAATTGGCTATACATGAAAGCAAAAATGGAGAGG
TACCAAACCATGACTTTGTCATTAATGGAATTCAAAACGGAGACTTGAAAA
CTGCTTTGCAAGTGACTGAAGGTATCAAGTATTTACAGAATAACACAACTGT
TGCTAAAGAAAAATTCGATGAAGCTAGTGGTGTGGGTGTTGAAATTACCCC
TCAACAAGCCAAGGCAGAAATTTCAAAATACCTTGATTCTATTAAAACTGA
TTTGGAAAGTAAAAGATATTCAATATTACCTAAGGTTTTGGGTGAAGTGAA
AACTCAGCCATCCTTGAAATGGGCCCCACCTCAATTATTCAAACC

> *GLN4*-3
TYTGCTTTATACAATATTTTAGGCACAAACTTCCCCAAAACCGATGATAAGA
AATTATCTTTGTTACATCAATTGGCTATACATGAAAGCAAAAATGGAGAGG
AACCAAACCATGACTTTGTCATTAATGGAATTCAAAATGGAGATTTGAAAA
CTGCTTTGCAAGTGACTGAAGGTATCAAGTATTTACAGAATAACACAACTGT
TGCTAAAGAAAAATTCGATGAAGCTAGTGGTGTGGGTATTGAAATTACCCC
TCAACAAGCCAAGGCAGAAATTTCAAAATACCTTGATTCTATTAAAACTGA
TTTGGAAAGTAAAAGATATTCAATATTACCTAAGGTTTTGGGTGAAGTGAA
AACTCAGCCATCCTTGAAATGGGCCCCACCTCAATTATTCAAACC

> *PMIb*-1
TTTAAGCCTTTGGATCAATTGGCTAAAACTTTGACTACAGTTCCTGAATTAA
ACGAAATTATTGGTCAAGAATTGGTTGATGAATTTGTTAGTGGTATTAGACT
ACCAGCAGAAATTGGAAGTCAAGATGATGTTAACAATAGAAAATTGTTACA
AAAAGTGTTTGGTAAATTAATGAACACCAATGAAGATATTATAAAGCAACA
AACTACTAAATTACTTGAAAGAACAGAAAGGGAACCTCAAGTGTTTAAAAA
CATTGATTCTAGATTACCAGAGTTAATACAAAGATTGAACAAGCAATTTCCT

AATGATATTGGATTATTTTGTGGATGTCTCTTGTTGAACCACGTTGGCTtGAA
CAAAGGGGAAGCA

> *PMIb*-2
TTTAAGCCTTTGGATCAATTGGCTAAAACTTTGACTACAGTTCCTGAATTAA
ACGAAATTATTGGTCAAGAATTGGTTGATGAATTTGTTAGTGGTATTAAACT
ACCAGCAGAAATTGGAAGTCAAGATGATGTTAACAATAGAAAATTGTTACA
AAAAGTGTTTGGTAAATTAATGAACACCAATGAAGATATTATAAAGCAACA
AACTACTAAATTACTTGAAAGAACAGAAAGGGAACCTCAAGTGTTTAAAAA
CATTGATTCTAGATTACCAGAGTTAATACAAAGATTGAACAAGCAATTTCCT
AATGATATTGGATTATTTTGTGGATGTCTCTTGTTGAACCACGTTGGCTTGA
ACAAAGGGGAAGCA

> *PMIb* -3
TTTAAGCCTTTGGATCAATTGGCTAAAACTTTGACTACAGTTCCTGAATTAA
ACGAAATTATTGGTCAAGAATTGGTTGATGAATTTGTTAGTGGTATTAAACT
ACCAGCAGAAATTGGAAGTCAAGATGATGTTAACAATAGAAAATTGTTACA
AAAAGTGTTTGGTAAATTAATGAACACCAATGAAGATATTATAAAGCAACA
AACTACTAAATTACTTGAAAGAACAGAAAGGGAACCTCAAGTGTTTAAAAA
CATTGATTCTAGATTACCAGAGTTAATACAAAGATTGAACAAGCAATTTCCT
AATGATATTGGATTATTTTGTGGATGTCTCTTGTTGAACCACGTTGGCTTGA
ACAAAGGGGAAGCG

> *PMIb* -4
TTTAAGCCTTTGGATCAATTGGCTAAAACTTTGACTACAGTTCCTGAATTAA
ACGAAATTATTGGTCAAGAATTGGTTGATGAATTTGTTAGTGGTATTARACT
ACCAGCAGAAATTGGAAGTCAAGATGATGTTAACAATAGAAAATTGTTACA
AAAAGTGTTTGGTAAATTAATGAACACCAATGAAGATATTATAAAGCAACA
AACTACTAAATTACTTGAAAGAACAGAAAGGGAACCTCAAGTGTTTAAAAA
CATTGATTCTAGATTACCAGAGTTAATACAAAGATTGAACAAGCAATTTCCT
AATGATATTGGATTATTTTGTGGATGTCTCTTGTTGAACCACGTTGGCTTGA
ACAAAGGGGAAGCR

> *PMIb* -5
TTTAAGCCTTTGGATCAATTGGCTAAAACTTTGACTACAGTTCCTGAATTAA
ACGAAATTATTGGTCAAGAATTGGTTGATGAATTTGTTAGTGGTATTAGACT
ACCAGCAGAAATTGGAAGTCAAGATGATGTTAACAATAGAAAATTGTTACA
AAAAGTGTTTGGTAAATTAATGAACACCAATGAAGATATTATAAAGCAACA
AACTACTAAATTACTTGAAAGAACAGAAAGGGAACCTCAAGTGTTTAAAAA
CATTGATTCTAGATTACCAGAGTTAATACAAAGATTGAACAAGCAATTTCCT
AATGATATTGGATTATTTTGTGGATGTCTCTTGTTGAACCACGTTGGCTTGA
ACAAAGGGGAAGCR

> *PMIb* -6
TTTAAGCCTTTGGATCAATTGGCTAAAACTTTGACTACAGTTCCTGAATTAA
ACGAAATTATTGGTCAAGAATTGGTTGATGAATTTGTTAGTGGTATTAGACT
ACCAGCAGAAATTGGAAGTCAAGATGATGTTAACAATAGAAAATTGTTACA
AAAAGTGTTTGGTAAATTAATGAACACCAATGAAGATATTATAAAGCAACA
AACTACTAAATTACTTGAAAGAACAGAAAGGGAACCTCAAGTGTTTAAAAA

CATTGATTCTAGATTACCAGAGTTAATACAAAGATTGAACAAGCAATTTCCT
AATGATATTGGATTATTTTGTGGATGTCTCTTGTTGAACCACGTTGGCTTGA
ACAAAGGGGAAGCG

> *PMIb* -7
TTTAAGCCTTTGGATCAATTGGCTAAAACTTTGACTACAGTTCCTGAATTAA
ACGAAATTATTGGTCAAGAATTGGTTGATGAATTTGTTAGTGGTATTARACT
ACCAGCAGAAATTGGAAGTCAAGATGATGTTAACAATAGAAAATTGTTACA
AAAAGTGTTTGGTAAATTAATGAACACCAATGAAGATATTATAAAGCAACA
AACTACTAAATTACTTGAAAGAACAGAAAGGGAACCTCAAGTGTTTAAAAA
CATTGATTCTAGATTACCAGAGTTAATACAAAGATTGAACAAGCAATTTCCT
AATGATATTGGATTATTTTGTGGATGTCTCTTGTTGAACCACGTTGGCTTGA
ACAAAGGGGAAGCA

> *RPN2*-1
TTGGTCCAAGTTTTCCGCTACTGCAGCATTGGGTGTTATTCATAAAGGTAAT
TTGTCACAAGGACGTACTATATTGAAACCATATTTGCCAGGATCATCTGGTT
CAGCCAATAACAAAGGTGGTTCCTTATTTGCCCTTGGATTAATCTTTGCTGG
CCATGGTAGAGAAGTCATTAAAACTTTGAAATCATTTATTGACCAAAATGGT
AATGCAGCAGGCTCAAATGATATTGATATTCAATTGCATGGGGCAGCTTTG
GGTGCTGGTGTTGCTGGTATGGGTTCAAAAAGTGAAAGTCTTTACGA

> *RPN2*-2
TTGGTCCAAGTTTTCCGCTACTGCAGCATTGGGTGTTATTCATAAAGGTAAT
TTGTCACAAGGACGTACTATATTGAAACCATATTTGCCAGGATCATCTGGTT
CAGCCAATAACAAAGGTGGTTCATTATTTGCCCTTGGATTAATCTTTGCTGG
CCATGGTAGAGAAGTCATTAAAACTTTGAAATCATTTATTGACCAAAATGGT
AATGCAGCAGGCTCAAATGATATTGATATTCAATTGCATGGGGCAGCTTTG
GGTGCTGGTGTTGCTGGTATGGGTTCAAAAAGTGAAAGCCTTTACGA

> *RPN2*-3
TTGGTCCAAGTTTTCCGCTACTGCAGCATTGGGTGTTATTCATAAAGGTAAT
TTGTCACAAGGACGTACTATATTGAAACCATATTTGCCAGGATCATCTGGTT
CAGCCAATAACAAAGGTGGTTCCTTATTTGCTCTTGGATTAATCTTTGCTGG
CCATGGTAGAGAAGTCATTAAAACTTTGAAATCATTTATTGACCAAAATGGT
AATGCAGCAGGCTCAAATGATATTGATATTCAATTGCATGGGGCAGCTTTG
GGTGCTGGTGTTGCTGGTATGGGTTCAAAAAGTGAAAGTCTTTACGA

> *ALA1*-1
TAAATCCAAGGAATTAGGGGTTGCGTTAAAGAAATTATCAATTTCCGTATTG
GATAAACAAAGATTAACTGAAAAATTCAACAAATTGGATAAATCAATCAAA
GATAATTTGAAGGCTAGACAAAAAGAGGAAACCAAAAAGACATTAGATGT
AGTTAACAATTGGTTGAATGATAAAGAAAATGTTTCATCATTTTTGGTTTCT
CATGTTCCAATTACTGCTAATGCAAAGGCAATTACTGAAGCTATTAACTTGA
TCAAAAAGCAAGACAAATCTAAATCAATTTATTTGTTGACCGGCGAAACCG
ATAAAGTTGCTCATGGATGTTATGTTAGTGATGAAGCTATTGCCAAGGGAAT
TAATGCCAATGAGTTAGCTAAGTCTGTATCT

> *ALA1*-2
TAAATCCAAGGAATTAGGGGTTGCGTTAAAGAAATTATCAATTTCCGTATTG
GATAAACAAAGATTAACTGAAAAATTCAATAAATTGGATAAATCAATCAAA
GATAATTTGAAGGCTAGACAAAAAGAGGAAACCAAAAAGACATTGGATGT
AGTTAACAATTGGTTGAATGATAAAGAAAATGCTTCATCATTTTTGGTTGCT
CATGTTCCAATTACTGCTAATGCAAAGGCAATTACTGAAGCTATTAACTTGA
TCAAAAAGCAAGACAAATCTAAATCAATTTATTTGTTGACCGGCGAAACCG
ATAAAGTTGCTCATGGATGTTATGTTAGTGATGAAGCTATTGCCAAGGGAAT
TAATGCCAATGAGTTAGCTAAGGCTGTATCT

> *ALA1*-3
TAAATCCAAGGAATTAGGGGTTGCGTTAAAGAAATTATCAATTTCCGTATTG
GATAAACAAAGATTAACTGAAAAATTCAAyAAATTGGATAAATCAATCAAA
GATAATTTGAAGGCTAGACAAAAAGAGGAAACCAAAAAGACATTGGATGT
AGTTAACAATTGGTTGAATGATAAAGAAAATGCTTCATCATTTTTGGTTGCT
CATGTTCCAATTACTGCTAATGCAAAGGCAATTACTGAAGCTATTAACTTGA
TCAAAAAGCAAGACAAATCTAAATCAATTTATTTGTTGACCGGCGAAACCG
ATAAAGTTGCTCATGGATGTTATGTTAGTGATGAAGCTATTGCCAAGGGAAT
TAATGCCAATGAGTTAGCTAAGGCTGTATCT

> *ALA1*-4
TAAATCCAAGGAATTAGGGGTTGCGTTAAAGAAATTATCAATTTCCGTATTG
GATAAACAAAGATTAACTGAAAAATTCAACAAATTGGATAAATCAATCAAA
GATAATTTGAAGGCTAGACAAAAAGAGGAAACCAAAAAGACATTAGATGT
AGTTAACAATTGGTTGAATGATAAAGAAAATGTTTCATCATTTTTGGTTTCT
CATGTTCCAATTACTGCTAATGCAAAGGCAATTACTGAAGCTATTAACTTGA
TCAAAAAGCAAGACAAATCTAAATCAATTTATTTGTTGACCGGCGAAACCG
ATAAAGTTGCTCATGGATGTTATGTTAGTGATGAAGCTATTGCCAAGGGAAT
TAATGCCAATGAGTTAGCTAAGGCTGTATCT

> *ALA1*-5
TAAATCCAAGGAATTAGGGGTTGCGTTAAAGAAATTATCAATTTCCGTATTG
GATAAACAAAGATTAACTGAAAAATTCAACAAATTGGATAAATCAATCAAA
GATAATTTGAAGGCTAGACAAAAAGAGGAAACCAAAAAGACATTAGATGT
AGTTAACAATTGGTTGAATGATAAAGAAAATGTTTCATCATTTTTGGTTTCT
CATGTTCCAATTACTGCTAATGCAAAGGCAATTACTGAAGCTATTAACTTGA
TCAAAAAGCAAGACAAATCTAAATCAATTTATTTGTTGACCGGCGAAACCG
ATAAAGTTGCTCATGGATGTTATGTTAGTGATGAAGCTATTGCCAAGGGAAT
TAATGCCAATGAGTTAGCTAAGkCTGTATCT

> *ALA1*-6
TAAATCCAAGGAATTAGGGGTTGCGTTAAAGAAATTATCAATTTCCGTATTG
GATAAACAAAGATTAACTGAAAAATTCAACAAATTGGATAAATCAATCAAA
GATAATTTGAAGGCTAGACAAAAAGAGGAAACCAAAAAGACATTGGATGT
AGTTAACAATTGGTTGAATGATAAAGAAAATGCTTCATCATTTTTGGTTGCT
CATGTTCCAATTACTGCTAATGCAAAGGCAATTACTGAAGCTATTAACTTGA
TCAAAAAGCAAGACAAATCTAAATCAATTTATTTGTTGACCGGCGAAACCG
ATAAAGTTGCTCATGGATGTTATGTTAGTGATGAAGCTATTGCCAAGGGAAT
TAATGCCAATGAGTTAGCTAAGGCTGTATCT

> *VPS13*-1
CCTTGATATGCAAATCAAACCTTTTGAAGCACGGGCAATCAACTGGTCCAC
AGATCTTAATGCTGAAGTTCATATTGAGCATTATATCAATATTTTCAATTAT
GCACGATCGTCTTGGGAACCTTTGGTTGAAAGCTGGCCAATAGCAGTTTACA
TGTCAAAAGCTCGACACCCGAAACCTCAATTATTAGTGGAAATAATTTCCA
GACAAGTAGCTCAAGTGACGCTTACATCTAAGGCAGTAGCATTGTTATCAC
AAGTATCTGATTTGATTACTTCTGGAGAGAAATTGAAACCAAGAGGTGAAG
ATTATCCATATGTAATAGTAAATGAAACTGGATTAGATTTGGAAGTTTGGAA
TGATACAAAGGAATCCGAAACTAGAACCGAAATAAAATCT
TGG

> *VPS13*-2
CCTTGATATGCAAATCAAACCTTTTGAAGCACGGGCAATCAACTGGTCCAC
AGATCTTAATGCTGAAGTTCATATTGAGCATTATATCAATATTTTCAATTAT
GCACGATCGTCTTGGGAACCTTTGGTTGAAAGCTGGCCAATAGCAGTTTACA
TGTCAAAAGCTCGACACCCGAAACCTCAATTATTAGTGGAAATAATTTCCA
GACAAGTAGCTCAAGTGACGCTTACATCTAAGGCAGTAGCATTGTTATCAC
AAGTATCTGATTTGATTACTTCTGGAGAGAAATTGAAACCAAGAGGTGAAG
ATTATCCATATGTAATAGTAAATGAAACTGGATTAGATTTGGAAGTTTGGAA
TGATACAAAGGAATCCGAAACTAGAACCGAAATAAAGTCT
TCG

> ex*VPS13*-1
AGAGCAAACGACTTGTACATGAAGTACCAAGGTATGGTTGAAGAGACATCA
AATTCAGAAGAATTCCAGTTCTCGAAGAATTTCAAAAAGAGGCTTTCGCAG
TATGCTCCTAGTGTGCTTTCAACATATTCTGATGAATCCGATTATGAGACTC
CCCGAATTCCACCAGGTGTTGCTATTGTAAAAGGTGAAGAGTTGAATGCCA
ATTTTGGTGGTTTGAGATTTGTTTTGATTGGTGATGTATCTGAGTTGCCTGTC
CTTGATATGCAAATCAAACCTTTTGAAGCACGGGCAATCAACTGGTCCACA
GATCTTAATGCTGAAGTTCATATTGAGCATTATATCAATATTTTCAATTATG
CACGATCGTCTTGGGAACCTTTGGTTGAAAGCTGGCCAATAGCAGTTTACAT
GTCAAAAGCTCGACACCCGAAACCTCAATTATTAGTGGAAATAATTTCCAG
ACAAGTAGCTCAAGTGACGCTTACATCTAAGGCAGTAGCATTGTTATCACA
AGTATCTGATTTGATTACTTCTGGAGAGAAATTGAAACCAAGAGGTGAAGA
TTATCCATATGTAATAGTAAATGAAACTGGATTAGATTTGGAAGTTTGGAAT
GATACAAAGGAATCCGAAACTAGAACCGAAATAAAATCTTGGGACTCTAAA
CCTTGG

> ex*VPS13*-2
AGAGCAAACGACTTGTACATGAAGTACCAAGGTATGGTTGAAGACACATCA
AATTCAGAAGAATTCCAGTTCTCGAAGAATTTCAAAAAGAGGCTTTCGCAG
TATGCTCCTAGTGTGCTTTCAACATATTCTGATGAATCCGATTATGAGACTC
CCCGAATTCCACCAGGTGTTGCTATTGTAAAAGGTGAAGAGTTGAATGCCA
ATTTTGGTGGTTTGAGATTTGTTTTGATTGGTGATGTATCTGAGTTGCCTGTC
CTTGATATGCAAATCAAACCTTTTGAAGCACGGGCAATCAACTGGTCCACA
GATCTTAATGCTGAAGTTCATATTGAGCATTATATCAATATTTTCAATTATG
CACGATCGTCTTGGGAACCTTTGGTTGAAAGCTGGCCAATAGCAGTTTACAT
GTCAAAAGCTCGACACCCGAAACCTCAATTATTAGTGGAAATAATTTCCAG
ACAAGTAGCTCAAGTGACGCTTACATCTAAGGCAGTAGCATTGTTATCACA

AGTATCTGATTTGATTACTTCTGGAGAGAAATTGAAACCAAGAGGTGAAGA
TTATCCATATGTAATAGTAAATGAAACTGGATTAGATTTGGAAGTTTGGAAT
GATACAAAGGAATCCGAAACTAGAACCGAAATAAAATCTTGGGACTCTAAA
CCTTGG

> ex*VPS13*-3
AGAGCAAACGACTTGTACATGAAGTACCAAGGTATGGTTGAAGAGACATCA
AATTCAGAAGAATTCCAGTTCTCGAAGAATTTCAAAAAGAGGCTTTCGCAG
TATGCTCCTAGTGTGCTTTCAACATATTCTGATGAATCCGATTATGAGACTC
CCCGAATTCCACCAGGTGTTGCTATTGTAAAAGGTGAAGAGTTGAATGCCA
ATTTTGGTGGTTTGAGATTTGTTTTGATTGGTGATGTATCTGAGTTGCCTGTC
CTTGATATGCAAATCAAACCTTTTGAAGCACGGGCAATCAACTGGTCCACA
GATCTTAATGCTGAAGTTCATATTGAGCATTATATCAATATTTTCAATTATG
CACGATCGTCTTGGGAACCTTTGGTTGAAAGCTGGCCAATAGCAGTTTACAT
GTCAAAAGCTCGACACCCGAAACCTCAATTATTAGTGGAAATAATTTCCAG
ACAAGTAGCTCAAGTGACGCTTACATCTAAGGCAGTAGCATTGTTATCACA
AGTATCTGATTTGATTACTTCTGGAGAGAAATTGAAACCAAGAGGTGAAGA
TTATCCATATGTAATAGTAAATGAAACTGGATTAGATTTGGAAGTTTGGAAT
GATACAAAGGAATCCGAAACTAGAACCGAAATAAAGTCTTCGGACTCTAAA
CCTTGG

> ex*VPS13*-4
AGAGCAAACGACTTGTACATGAAGTACCAAGGTATGGTTGAAGACACATCA
AATTCAGAAGAATTCCAGTTCTCGAAGAATTTCAAAAAGAGGCTTTCGCAG
TATGCTCCTAGTGTGCTTTCAACATATTCTGATGAATCCGATTATGAGACTC
CCCGAATTCCACCAGGTGTTGCTATTGTAAAAGGTGAAGAGTTGAATGCCA
ATTTTGGTGGTTTGAGATTTGTTTTGATTGGTGATGTATCTGAGTTGCCTGTC
CTTGATATGCAAATCAAACCTTTTGAAGCACGGGCAATCAACTGGTCCACA
GATCTTAATGCTGAAGTTCATATTGAGCATTATATCAATATTTTCAATTATG
CACGATCGTCTTGGGAACCTTTGGTTGAAAGCTGGCCAATAGCAGTTTACAT
GTCAAAAGCTCGACACCCGAAACCTCAATTATTAGTGGAAATAATTTCCAG
ACAAGTAGCTCAAGTGACGCTTACATCTAAGGCAGTAGCATTGTTATCACA
AGTATCTGATTTGATTACTTCTGGAGAGAAATTGAAACCAAGAGGTGAAGA
TTATCCATATGTAATAGTAAATGAAACTGGATTAGATTTGGAAGTTTGGAAT
GATACAAAGGAATCCGAAACTAGAACCGAAATAAAGTCTTCGGACTCTAAA
CCTTGG

> ex*VPS13*-5
AGAGCAAACGACTTGTACATGAAGTACCAAGGTATGGTTGAAGAGACATCA
AATTCAGAAGAATTCCAGTTCTCGAAGAATTTCAAAAAGAGGCTTTCGCAG
TATGCTCCTAGTGTGCTTTCAACATATTCTGATGAATCCGATTATGAGACTC
CCCGAATTCCACCAGGTGTTGCTATTGTAAAAGGTGAAGAGTTGAATGCCA
ATTTTTGTGGTTTGAGATTTGTTTTGATTGGTGATGTATCTGAGTTGCCTGTC
CTTGATATGCAAATCAAACCTTTTGAAGCACGGGCAATCAACTGGTCCACA
GATCTTAATGCTGAAGTTCATATTGAGCATTATATCAATATTTTCAATTATG
CACGATCGTCTTGGGAACCTTTGGTTGAAAGCTGGCCAATAGCAGTTTACAT
GTCAAAAGCTCGACACCCGAAACCTCAATTATTAGTGGAAATAATTTCCAG
ACAAGTAGCTCAAGTGACGCTTACATCTAAGGCAGTAGCATTGTTATCACA
AGTATCTGATTTGATTACTTCTGGAGAGAAATTGAAACCAAGAGGTGAAGA

TTATCCATATGTAATAGTAAATGAAACTGGATTAGATTTGGAAGTTTGGAAT
GATACAAAGGAATCCGAAACTAGAACCGAAATAAAATCTTGGGACTCTAAA
CCTTGG

> *ZWF1b*-1
CAAACCAGGTTATTTAGATGATAAAACGGTGAATCCTGATTCTAAAGCTGTT
ACTTATGCTGCTTTTAGAGTAAACATTCATAATGAAAGATGGGATGGTGTTC
CAATTGTTTTACGTGCTGGTAAAGCTTTAGATGAAAGTAAAGTTGAAATTAG
AATTCAATTTAAACCTGTTGCTAAAGGAATGTTTAAAGAAATTCAAAGAAA
TGAATTAGTTATTAGAATACAACCAAATGAAGCAATTTATTTAAAAATTAAT
TCAAAGATCCCCGGGATTTCTACTGAAACTTCTTTAACTGATTTAGATTTAA
CTTATGCTACTCGTTATTCAAAAGATTTTTGGATTCCAGAAGCTTATGAAGC
TTTAATTAGAGATTGTTATTTAGGTAATCATTCTAATTTTGTTAGAGATGATG
AATTGGATGTTTCTTGGAAATTATTTACTCCATTATTAAAAGCAGTTGAAGA
TCCTGCTAATAAAATAGAATTAC

> *ZWF1b*-2
TAAACCAGGTTATTTAGATGATAAAACGGTGAATCCTGATTCTAAAGCTGTT
ACTTATGCTGCTTTTAGAGTAAACATTCATAATGAAAGATGGGATGGTGTTC
CAATTGTTTTACGTGCTGGTAAAGCTTTAGATGAAAGTAAAGTTGAAATTAG
AATTCAATTTAAACCTGTTGCTAAAGGAATGTTTAAAGAAATTCAAAGAAA
TGAATTAGTTATTAGAATACAACCAAATGAAGCAATTTATTTAAAAATTAAT
TCAAAAATCCCCGGGATTTCTACTGAAACTTCTTTAACTGATTTAGATTTAA
CTTATGCTACTCGTTATTCAAAAGATTTTTGGATTCCAGAAGCTTATGAAGC
TTTAATTAGAGATTGTTATTTAGGTAATCATTCTAATTTTGTTAGAGATGATG
AATTGGATGTTTCTTGGAAATTATTTACTCCATTATTAAAAGCAGTTGAAGA
TCCTGCTAATAAAATAGAATTAC

> *ZWF1b*-3
CAAACCAGGTTATTTAGATGATAAAACrGTGAATCCTGATTCTAAAGCTGTT
ACTTATGCTGCTTTTAGAGTAAACATTCATAATGAAAGATGGGATGGTGTTC
CAATTGTTTTACGTGCTGGTAAAGCTTTAGATGAAAGTAAAGTTGAAATTAG
AATTCAATTTAAACCTGTTGCTAAAGGAATGTTTAAAGAAATTCAAAGAAA
TGAATTAGTTATTAGAATACAACCAAATGAAGCAATTTATTTAAAAATTAAT
TCAAAAATCCCCGGGATTTCTACTGAAACTTCTTTAACTGATTTAGATTTAA
CTTATGCTACTCGTTATTCAAAAGATTTTTGGATTCCAGAAGCTTATGAAGC
TTTAATTAGAGATTGTTATTTAGGTAATCATTCTAATTTTGTTAGAGATGATG
AATTGGATGTTTCTTGGAAATTATTTACTCCATTATTAAAAGCAGTTGAAGA
TCCTGCTAATAAAATAGAATTAC

> *ZWF1b*-4
CAAACCAGGTTATTTAGATGATAAAACGGTGAATCCTGATTCTAAAGCTGTT
ACTTATGCTGCTTTTAGAGTAAACATTCATAATGAAAGATGGGATGGTGTTC
CAATTGTTTTACGTGCTGGTAAAGCTTTAGATGAAAGTAAAGTTGAAATTAG
AATTCAATTTAAACCTGTTGCTAAAGGAATGTTTAAAGAAATTCAAAGAAA
TGAATTAGTTATTAGAATACAACCAAATGAAGCAATTTATTTAAAAATTAAT
TCAAAGATCCCCGGGATTTCTACTGAAACTTCTTTAACTTATTTAGATTTAA
CTTATGCTACTCGTTATTCAAAAGATTTTTGGATTCCAGAAGCTTATGAAGC
TTTAATTAGAGATTGTTATTTAGGTAATCATTCTAATTTTGTTAGAGATGATG

AATTGGATGTTTCTTGGAAATTATTTACTCCATTATTAAAAGCAGTTGAAGA
TCCTGCTAATAAAATAGAATTAC

> *ZWF1b*-5
CAAACCAGGTTATTTAGATGATAAAACGGTGAATCCTGATTCTAAAGCTGTT
ACTTATGCTGCTTTTAGAGTAAACATTCATAATGAAAGATGGGATGGTGTTC
CAATTGTTTTACGTGCTGGTAAAGCTTTAGATGAAAGTAAAGTTGAAATTAG
AATTCAATTTAAACCTGTTGCTAAAGGAATGTTTAAAGAAATTCAAAGAAA
TGAATTAGTTATTAGAATACAACCAAATGAAGCAATTTATTTAAAAATTAAT
TCAAAGATCCCCGGGATTTCTACTGAAACTTCTTTAACTGATTTAGATTTAA
CTTATGCTACTCGTTATTCAAAGGATTTTTGGATTCCAGAAGCTTATGAAGC
TTTAATTAGAGATTGTTATTTAGGTAATCATTCTAAT
TTTGTTAGAGATGATGAATTGGATGTTTCTTGGAAATTATTTACTCCATTATT
AAAAGCAGTTGAAGATCCTGCTAATAAAATAGAATTAC

> *ZWF1b*-6
CAAACCAGGTTATTTAGATGATAAAACGGTGAATCCTGATTCTAAAGCTGTT
ACTTATGCTGCTTTTAGAGTAAACATTCATAATGAAAGATGGGATGGTGTTC
CAATTGTTTTACGTGCTGGTAAAGCTTTAGATGAAAGTAAAGTTGAAATTAG
AATTCAATTTAAACCTGTTGCTAAAGGAATGTTTAAAGAAATTCAAAGAAA
TGAATTAGTTATTAGAATACAACCAAATGAAGCAATTTATTTAAAAATTAAT
TCAAAGATCCCCGGGATTTCTACTGAAACTTCTTTAACTGATTTAGATTTAA
CTTATGCTACTCGTTATTCAAAAGATTTTTGGATTCCAGAAGCTTATGAAGC
TTTAATTAGAGATTGTTATTTAGGTAATCATTCTAATTTTGTTAGAGATGATG
AATTGGATGTTTCTTGGAAATTATTTACTCCATTATTAAAAGCAGTTGAAGA
TCCTGCTAATAGAATAGAATTAC

> *ZWF1b*-7
CAAACCAGGTTATTTAGATGATAAAACGGTGAATCCTGATTCTAAAGCTGTT
ACTTATGCTGCTTTTAGAGTAAACATTCATAATGAAAGATGGGATGGTGTTC
CAATTGTTTTACGTGCTGGTAAAGCTTTAGATGAAAGTAAAGTTGAAATTAG
AATTCAATTTAAACCTGTTGCTAAAGGAATGTTTAAAGAAATTCAAAGAAA
TGAATTAGTTATTAGAATACAACCAAATGAAGCAATTTATTTAAAAATTAAT
TCAAAGATCCCCGGGATTTCTACTGAAACTTCTTTAACTGATTTAGATTTAA
CTTATGCTACTCGTTATTCAAAAGATTTTTGGATTCCAGAAGCTTATGAAGC
TTTAATTAGAGATTGTTATTTAGGTAATCATTCTAATTTTGTTAGAGATGATG
AATTGGATGTTTCTTGGAAATTATTTACTCCATTATTAAAAGCAGTTGAAGA
TCCTGCTAATARAATAGAATTAC

> ex*ZWF1b*-1
AAAGTTTTAAAAGCTTTTGATAATATTGATGTTAATGATGTTATATTGGGTC
AATATACTAAATCAGAAGATGGTACCAAACCAGGTTATTTAGATGATAAAA
CGGTGAATCCTGATTCTAAAGCTGTTACTTATGCTGCTTTTAGAGTAAACAT
TCATAATGAAAGATGGGATGGTGTTCCAATTGTTTTACGTGCTGGTAAAGCT
TTAGATGAAAGTAAAGTTGAAATTAGAATTCAATTTAAACCTGTTGCTAAA
GGAATGTTTAAAGAAATTCAAAGAAATGAATTAGTTATTAGAATACAACCA
AATGAAGCAATTTATTTAAAAATTAATTCAAAGATCCCCGGGATTTCTACTG
AAACTTCTTTAACTGATTTAGATTTAACTTATGCTACTCGTTATTCAAAAGAT
TTTTGGATTCCAGAAGCTTATGAAGCTTTAATTAGAGATTGTTATTTAGGTA

ATCATTCTAATTTTGTTAGAGATGATGAATTGGATGTTTCTTGGAAATTATTT
ACTCCATTATTAAAAGCAGTTGAAGATCCTGCTAATAAAATAGAATTACAA
TATTATCCTTATGGTTCAAAAGGTCCAAAAGATTTAAGAAAATATTTGAAA

> ex*ZWF1b*-2
AAAGTTTTAAAAGCTTTTGATAATATTGATGTTAATGATGTTATATTGGGTC
AATATACTAAATCAGAAGATGGTACTAAACCAGGTTATTTAGATGATAAAA
CGGTGAATCCTGATTCTAAAGCTGTTACTTATGCTGCTTTTAGAGTAAACAT
TCATAATGAAAGATGGGATGGTGTTCCAATTGTTTTACGTGCTGGTAAAGCT
TTAGATGAAAGTAAAGTTGAAATTAGAATTCAATTTAAACCTGTTGCTAAA
GGAATGTTTAAAGAAATTCAAAGAAATGAATTAGTTATTAGAATACAACCA
AATGAAGCAATTTATTTAAAAATTAATTCAAAAATCCCCGGGATTTCTACTG
AAACTTCTTTAACTGATTTAGATTTAACTTATGCTACTCGTTATTCAAAAGAT
TTTTGGATTCCAGAAGCTTATGAAGCTTTAATTAGAGATTGTTATTTAGGTA
ATCATTCTAATTTTGTTAGAGATGATGAATTGGATGTTTCTTGGAAATTATTT
ACTCCATTATTAAAAGCAGTTGAAGATCCTGCTAATAAAATAGAATTACAA
TATTATCCTTATGGTTCTAAAGGTCCAAAAGATTTAAGAAAATATTTGAAA

> ex*ZWF1b*-3
AAAGTTTTAAAAGCTTTTGATAATATTGATGTTAATGATGTTATATTGGGTC
AATATACTAAATCAGAAGATGGTACCAAACCAGGTTATTTAGATGATAAAA
CRGTGAATCCTGATTCTAAAGCTGTTACTTATGCTGCTTTTAGAGTAAACAT
TCATAATGAAAGATGGGATGGTGTTCCAATTGTTTTACGTGCTGGTAAAGCT
TTAGATGAAAGTAAAGTTGAAATTAGAATTCAATTTAAACCTGTTGCTAAA
GGAATGTTTAAAGAAATTCAAAGAAATGAATTAGTTATTAGAATACAACCA
AATGAAGCAATTTATTTAAAAATTAATTCAAAAATCCCCGGGATTTCTACTG
AAACTTCTTTAACTGATTTAGATTTAACTTATGCTACTC
GTTATTCAAAAGATTTTTGGATTCCAGAAGCTTATGAAGCTTTAATTAGAGA
TTGTTATTTAGGTAATCATTCTAATTTTGTTAGAGATGATGAATTGGATGTTT
CTTGGAAATTATTTACTCCATTATTAAAAGCAGTTGAAGATCCTGCTAATAA
AATAGAATTACAATAyTATCCTTATGGTTCAAAAGGTCCAAAAGATTTAAGA
AAATATTTGAAA

> ex*ZWF1b*-4
AAAGTTTTAAAAGCTTTTGATAATATTGATGTTAATGATGTTATATTGGGTC
AATATACTAAATCAGAAGATGGTACCAAACCAGGTTATTTAGATGATAAAA
CGGTGAATCCTGATTCTAAAGCTGTTACTTATGCTGCTTTTAGAGTAAACAT
TCATAATGAAAGATGGGATGGTGTTCCAATTGTTTTACGTGCTGGTAAAGCT
TTAGATGAAAGTAAAGTTGAAATTAGAATTCAATTTAAACCTGTTGCTAAA
GGAATGTTTAAAGAAATTCAAAGAAATGAATTAGTTATTAGAATACAACCA
AATGAAGCAATTTATTTAAAAATTAATTCAAAGATCCCCGGGATTTCTACTG
AAACTTCTTTAACTTATTTAGATTTAACTTATGCTACTCGTTATTCAAAAGAT
TTTTGGATTCCAGAAGCTTATGAAGCTTTAATTAGAGATTGTTATTTAGGTA
ATCATTCTAATTTTGTTAGAGATGATGAATTGGATGTTTCTTGGAAATTATTT
ACTCCATTATTAAAAGCAGTTGAAGATCCTGCTAATAAAATAGAATTACAA
TATTATCCTTATGGTTCAAAAGGTCCAAAAGATTTAAGAAAATATTTGAAA

> ex*ZWF1b*-5
AAAGTTTTAAAAGCTTTTGATAATATTGATGTTAATGATGTTATATTGGGTC
AATATACTAAATCAGAAGATGGTACCAAACCAGGTTATTTAGATGATAAAA
CGGTGAATCCTGATTCTAAAGCTGTTACTTATGCTGCTTTTAGAGTAAACAT
TCATAATGAAAGATGGGATGGTGTTCCAATTGTTTTACGTGCTGGTAAAGCT
TTAGATGAAAGTAAAGTTGAAATTAGAATTCAATTTAAACCTGTTGCTAAA
GGAATGTTTAAAGAAATTCAAAGAAATGAATTAGTTATTAGAATACAACCA
AATGAAGCAATTTATTTAAAAATTAATTCAAAAATCCCCGGGATTTCTACTG
AAACTTCTTTAACTGATTTAGATTTAACTTATGCTACTCGTTATTCAAAGGAT
TTTTGGATTCCAGAAGCTTATGAAGCTTTAATTAGAGATTGTTATTTAGGTA
ATCATTCTAATTTTGTTAGAGATGATGAATTGGATGTTTCTTGGAAATTATTT
ACTCCATTATTAAAAGCAGTTGAAGATCCTGCTAATAAAATAGAATTACAA
TATTATCCTTATGGTTCAAAAGGTCCAAAAGATTTAAGAAAATATTTGAAA

> ex*ZWF1b*-6
AAAGTTTTAAAAGCTTTTGATAATATTGATGTTAATGATGTTATATTGGGTC
AATATACTAAATCAGAAGATGGTACCAAACCAGGTTATTTAGATGATAAAA
CRGTGAATCCTGATTCTAAAGCTGTTACTTATGCTGCTTTTAGAGTAAACAT
TCATAATGAAAGATGGGATGGTGTTCCAATTGTTTTACGTGCTGGTAAAGCT
TTAGATGAAAGTAAAGTTGAAATTAGAATTCAATTTAAACCTGTTGCTAAA
GGAATGTTTAAAGAAATTCAAAGAAATGAATTAGTTATTAGAATACAACCA
AATGAAGCAATTTATTTAAAAATTAATTCAAAAATCCCCGGGATTTCTACTG
AAACTTCTTTAACTGATTTAGATTTAACTTATGCTACTCGTTATTCAAAAGAT
TTTTGGATTCCAGAAGCTTATGAAGCTTTAATTAGAGATTGTTATTTAGGTA
ATCATTCTAATTTTGTTAGAGATGATGAATTGGATGTTTCTTGGAAATTATTT
ACTCCATTATTAAAAGCAGTTGAAGATCCTGCTAATAAAATAGAATTACAA
TAyTATCCTTATGGTTCAAAAGGTCCAAAAGATTTAAGAAAATATTTGAAA

> ex*ZWF1b*-7
AAAGTTTTAAAAGCTTTTGATAATATTGATGTTAATGATGTTATATTGGGTC
AATATACTAAATCAGAAGATGGTACCAAACCAGGTTATTTAGATGATAAAA
CGGTGAATCCTGATTCTAAAGCTGTTACTTATGCTGCTTTTAGAGTAAACAT
TCATAATGAAAGATGGGATGGTGTTCCAATTGTTTTACGTGCTGGTAAAGCT
TTAGATGAAAGTAAAGTTGAAATTAGAATTCAATTTAAACCTGTTGCTAAA
GGAATGTTTAAAGAAATTCAAAGAAATGAATTAGTTATTAGAATACAACCA
AATGAAGCAATTTATTTAAAAATTAATTCAAAGATCCCCGGGATTTCTACTG
AAACTTCTTTAACTGATTTAGATTTAACTTATGCTACTCGTTATTCAAAAGAT
TTTTGGATTCCAGAAGCTTATGAAGCTTTAATTAGAGATTGTTATTTAGGTA
ATCATTCTAATTTTGTTAGAGATGATGAATTGGATGTTTCTTGGAAATTATTT
ACTCCATTATTAAAAGCAGTTGAAGATCCTGCTAATAGAATAGAATTACAA
TATTATCCTTATGGTTCAAAAGGTCCAAAAGATTTAAGAAAATATTTGAAA

> ex*ZWF1b*-8
AAAGTTTTAAAAGCTTTTGATAATATTGATGTTAATGATGTTATATTGGGTC
AATATACTAAATCAGAAGATGGTACCAAACCAGGTTATTTAGATGATAAAA
CGGTGAATCCTGATTCTAAAGCTGTTACTTATGCTGCTTTTAGAGTAAACAT
TCATAATGAAAGATGGGATGGTGTTCCAATTGTTTTACGTGCTGGTAAAGCT
TTAGATGAAAGTAAAGTTGAAATTAGAATTCAATTTAAACCTGTTGCTAAA
GGAATGTTTAAAGAAATTCAAAGAAATGAATTAGTTATTAGAATACAACCA

AATGAAGCAATTTATTTAAAAATTAATTCAAAGATCCCCGGGATTTCTACTG
AAACTTCTTTAACTGATTTAGATTTAACTTATGCTACTCGTTATTCAAAAGAT
TTTTGGATTCCAGAAGCTTATGAAGCTTTAATTAGAGATTGTTATTTAGGTA
ATCATTCTAATTTTGTTAGAGATGATGAATTGGATGTTTCTTGGAAATTATTT
ACTCCATTATTAAAAGCAGTTGAAGATCCTGCTAATARAATAGAATTACAAT
ATTATCCTTATGGTTCAAAAGGTCCAAAAGATTTAAGAAAATATTTGAAA