Molecular Typing and Phylogenetic Analysis of *Candida albicans* Isolates from Different Patient Populations

A thesis presented in partial fulfilment of the requirements for the degree of PhD

in

Molecular Biology and Genetics

at Massey University, Palmerston North,

New Zealand.

Nicole Beate Freifrau von Maltzahn

2000

ABSTRACT

An important question in understanding the epidemiology of *Candida albicans* is to determine whether certain strains, e.g. HSP strains (highly successful strains), more prevalent in Candidosis patients than in healthy individuals, replace commensal strains under certain conditions and whether the replacing strains may cause Candidosis. This question was investigated in experiments which monitored the genetic diversity of commensal *C. albicans* isolates obtained from individuals before and after they were exposed to conditions that may predispose them to the development of Candidosis. The distinctiveness of *C.albicans* strains isolated from individuals was analysed using the phylogenetic method of split decomposition. The method was found to provide a good representation of the phylogenetic information in strain replacement Ca3 data.

Our study highlighted difficulties in monitoring of strain replacement with Ca3 methodology. An indication for strain replacement was observed in one patient at low risk to acquire Candidosis. However, the observations that cancer patients, who were at a high risk of developing Candidosis, were colonised with diverse strains and that healthy individuals could be colonised with different commensal *C.albicans* strains within one body location, cautioned against overinterpretation of this finding. These results demonstrated the need for extensive sampling of larger numbers of isolates from different body locations when evaluating replacement hypotheses.

In investigating potential sources of *C.albicans* infections we successfully isolated this fungus from the hospital environment of high risk patients, demonstrating the potential of the hospital environments as a source for infection causing strains. In characterising strains, a nonradioactive fingerprinting protocol was developed for a more convenient use of Ca3 fingerprinting.

Until now the existence of the HSP group has been based entirely on Ca3 fingerprinting data. To test for the existence of this group we have analysed amplified fragment length polymorphisms (AFLP) of 36 *C.albicans* isolates from different geographical regions. Phylogenetic reconstruction from both data forms (Ca3 and

AFLP data) were highly congruent and suggest a worldwide distribution of HSP strains. Study of the tree building properties of AFLP and Ca3 data using Quartet Puzzling and a tree comparison metric showed that AFLP data were more treelike than the Ca3 data.

However, whilst both AFLP and Ca3 methods provided high resolution data to identify strains and substrains of *C.albicans*, the need for population based studies to test for strain replacement makes the use of either method limited. For this reason, both nuclear ITS and AFLP derived PCR markers were investigated for their potential use in such studies. In particular, one AFLP derived PCR marker that was partially characterised appears very promising for future strain replacement studies. It is likely to provide a simple diagnostic test for rapid identification of HSP strains.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. Pete Lockhart for providing excellent scientific guidance and encouragement. Thanks for all the scientific discussions, the financial support, and the wonderful working environment. It was a real pleasure working with you.

I am grateful to my supervisor Dr. Paul O'Toole for your guidance and encouragement, which helped a lot to finish this thesis.

Prof. Dr. Patrick Sullivan I would like to thank for providing his scientific expertise and encouragement. I am also thankful for the opportunity of attending your lab meetings and the scientific discussions. Moreover I like to thank you for the financial support which allowed me to attend a conference overseas.

To my supervisor Dr. Bart Baker from Palmerston North Hospital, many thanks for your support at the hospital and for helping in the selection and sampling of the patients. I also like to thank the hospital staff, especially Susan for their help in sampling the patients. Thank you for your time and effort.

I like to thank my supervisor and initiator of this project Dr. Jan Schmid for the provision of the scientific background regarding *C.albicans* epidemiology. I am also thankful for the financial support to attend a conference outside of New Zealand. I thankfully acknowledge financial support from the award of a Massey University Doctoral Scholarship.

I am very grateful to Dr. Linda Giblin for her scientific contribution to this thesis and for her friendship.

Many thanks to the DEB lab people for their help. A special thanks to Trish for the introduction into the AFLP world, for always giving a helping hand, and for your cheerfulness. Thanks to Karen for her help to find my way in a new

environment and for listening. Thanks to Richard, Leon, Carmel, Abbey, Matt, Nena, Owen, and Dan, it was real fun working with you.

Special thanks to the people of the "other" labs. Thanks to Martin for bearing me during all the times without losing any patience and for his friendship. Also many thanks to Sheralee for a brief but very enjoyable working relationship which transformed into in hopefully long friendship. I also like to thank Tash, Michelle, Rochelle, Rebecca, and Yong for their help and for the provision of some "urgently" needed substances.

Finally, I'd like to thank my husband Christoph for always being there for me and to whom I'd like to dedicate this thesis.

TABLE OF CONTENTS

ABBREVIATIONS

A	adriamycin, immunosuppressive anticancer drug
ABI	Applied Biosystems
ABCM	combination therapy of four immunosuppressive anticancer drugs: adriamycin , BCNU (cisplatin), cyclophosphamide, melphalan
AFLP	amplified fragment length polymorphism
ALL	acute lymphoid leukaemia
AML	acute myeloid leukaemia
APS	ammonium persulphate
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BLAST	basic local alignment search tool
bp	base pair
BSA	bovine serum albumine
C	cyclophosphamide, immunosuppressive anticancer drug
°C	degree Celsius
Ca3	moderately repetitive sequence of the *Candida albicans* genome
CBP1	corticosteroid binding protein gene
CHOP	combination therapy of three immunosuppressive anticancer drugs cyclophosphamide, doxorubicin, vincristine, and the adrenocorticoid predisone
CLL	chronic lymphoid leukaemia
CML	chronic myeloid leukaemia
CMI	cell-mediated immunity
COP	combination therapy of two immunosuppressive anitcancer drugs cyclophosphamide, vincristine, and the adrenocorticoid predisone
CSPD	disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'chloro) tricyclo[3.3.1.1.]decan}-4-yl)phenyl phosphate
C1-C3	cleavage sites
DIG	digoxigenin
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dATP	2' deoxyadenosine triphosphate
dCTP	2' deoxycytidine triphosphate
dGTP	2' deoxyguanosine triphosphate
dNTP	deoxynucleoside triphosphate
dTTP	2' deoxythymidine triphosphate
dUTP	2' deoxyuridine triphosphate
EDTA	ethylenediaminetetraacetic acid
ESR	Environmental Science Research, New Zealand
FA	folinic acid, immunosuppressive anticancer drug
F_{ab}	variable sequence fragment of immunoglobulin
g	gram
GC	content of deoxyguanylate and deoxycytidylate in DNA

h	hour
HAS	hospitalisation for at least 7 days, surgery (hip- or knee replacement), and antibiotic treatment
HSP	highly successful pathogens
IPTG	isopropyl-1-thio-ß-D-galactoside
ITS	internal transcribed spacer
kb	kilo base
l	liter
LB	Luria-Bertani broth
LTR	long terminal repeats
M	mole
m	milli
min	minute
MLEE	multilocus enzyme electrophoresis
µ	micro
n.a.	not available
NBT	4-nitro blue tetrazolium chloride
NJ	neighbour-joining
OD	optical density
PCR	polymerase chain reaction
RAPD	random amplified polymorphic DNA
RCF	relative centrifugal force
rDNA	DNA that encodes for ribosomal ribonucleic acid
RFLP	restriction fragment length polymorphism
rRNA	ribosomal ribonucleic acid
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RT	room temperature
SAB	Sabouraud
S_{AB}	Similarity value
SDS	sodium dodecyl sulphate
subsp.	sub species
TBS	tris-buffered saline
TEMED	**NNN'N'** tetramethylethylenediamine
T_m	melting temperature
Tris	tris(hydroxymethyl)methylamine
U	unit
UPGMA	unweighted pairgroup method with algorithmic means
URS	upstream regulatory sequence
USA	United States of America
V	volt
Vol	volume
vs.	versus

w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside
YPD	Yeast-peptone-dextrose medium
5FU	5-fluorouacil, immunosuppressive anticancer drug

In addition, the conventional one-letter codes for deoxyribonucleosides was applied:

deoxyribonucleosides: A, C, G, T for deoxyadenylate, deoxycytidylate, deoxyguanylate and deoxythymidylate, respectively.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - GENERAL INTRODUCTION

1.1 Morphological Characteristics of *Candida albicans*

Candida albicans is a dimorphic yeast which can assume a variety of shapes. These range from the spherical, budding yeast form to an elongate hyphal form, which can be either pseudohypha or true hypha. Hyphae arise as branching of existing hyphae or from blastospores. In the latter case a blastospore gives rise to new cellular material in the shape of a cylinder, the germ tube (Odds, 1988). Formation of hyphae can be induced, for example, by dilution in serum-containing medium (Odds, 1988), and this is the basis for the traditional germ tube test used for identification of *C.albicans* (see section 2.3.2).

Yeast cell and hyphae phenotypes do not represent the entire phenotypic repertoire of *C.albicans*. For more than 50 years, numerous investigators have reported a phenotypic variability and instability in colony phenotypes in *C.albicans*, known as high frequency switching (Soll, 1991). Morphological variants which alter in colony form, colour, size or texture arise spontaneously at frequencies between 1.4% and 10% (Odds, 1988; Rutschenko-Bulgac *et al.*, 1990). Because of this instability it is not possible to discriminate between different *C.albicans* strains based on their phenotypes. Therefore several methods have been developed to differentiate *C.albicans* strains on a genetic level.

1.2 Genetic Characteristics of *Candida albicans*

C.albicans is a permanently diploid organism with no known sexual cycle (Odds, 1994). Eight pairs of homologous chromosomes have been detected and these may show length polymorphisms in different *C.albicans* strains (Magee and Magee, 1987; Iwaguchi *et al.*, 1990). Recombination events, such as mitotic

crossing-over involving DNA translocations and reciprocal translocations between repetitive DNA sequences, may contribute to this phenomenon (Trash-Bingham and Gorman, *1992*). The genome consists of 13% repetitive DNA (Wills *et al.*, 1984), of which several species-specific moderately repetitive DNA sequences have been used as probes in Southern blot hybridisation analysis to discriminate *C.albicans* strains (Cutler *et al.*, 1988; Scherer and Stevens, 1988; Schmid *et al.*, 1990; Lasker *et al.*, 1992). One of the most accurate and reproducible methods is Ca3 fingerprinting, which has been employed to answer basic questions about epidemiology of *C.albicans* (Schmid *et al.*, 1990; Soll *et al.*, 1991; Schröppel *et al.*, 1994).

1.3 Epidemiology of *Candida albicans*

C.albicans is an opportunistic pathogen commonly found as part of the commensal flora of humans but capable of causing disease, termed Candidosis. Under certain conditions Candidosis can become a life-threatening disease. One study has shown that healthy individuals are often colonised with genetically similar strains, but they can also be colonised with unrelated strains in different body locations (Soll *et al.,* 1991). The fact that genetically nonidentical strains simultaneously colonise different anatomical locations of the same individual has suggested the existence of vaginotropic, orotropic, and perhaps anorectotropic strains (Soll *et al.,* 1991). It has been suggested that recurrent infections may be caused by strains showing minor genetic differences and which have undergone microevolution (Schröppel *et al.*, 1994). Strains from one geographical locale have been shown to be more similar than strains from different geographical areas (Schmid, 1993; Pfaller *et al.*, 1998, Schmid *et al.*, 1999).

However, some of the most important questions concerning *C.albicans* epidemiology have not been answered yet. For example, it is not clear whether infection-causing strains derive from the commensal flora or from exogenous

sources. Nevertheless, it is generally believed that commensal strains are the source for subsequent infection (Pechere, 1993; Dembry *et al.*, 1994; Pfaller, 1994; Voss *et al.*, 1995), and that the host defence plays the most important role during the transition from commensal to a pathogenic state (see section 3.1.2). The importance of virulence factors in *C.albicans* (see section 1.5) is still controversial. It is possible that infection-causing strains are transferred from an exogenous source (Sanchez *et al.*, 1993; Schmid, 1993; van Belkum *et al.*, 1994; Schmid *et al.*, 1995) and are capable of replacing commensal strains.

1.4 Replacement of *Candida albicans* Strains?

Recently, it has been suggested that Candidosis may be associated with the occurrence of particular strains of *C.albicans*, which may replace the normal commensal flora (Schmid, 1993; Schmid *et al.*, 1995; Schmid *et al.* 1999). This hypothesis was originally based on phylogenetic reconstruction of *C.albicans* strains using a UPGMA (unweighted pairgroup) clustering of S_{AB} values derived from Ca3 fingerprints. A more recent analysis has used neighbour joining clustering of a related distance measure derived for similar data (Schmid *et al.,* 1999). Both analyses have identified a group of genetically similar *C.albicans* strains. Schmid et al. (1995) report that this group has been found to be more prevalent in Candidosis patients (up to 70%) than in healthy individuals (approximately 30%), and that this cluster of strains, termed "group 13 isolates" (Schmid *et al.*, 1995) and "general-purpose genotype" (cluster A) strains (Schmid *et al.,* 1999) is the predominant aetiological agent of Candidosis. Physiological properties of these strains thought to be associated with the success of this group include resistance to a larger number of chemicals (tested in a resistogram assay) and a better adherence to saliva coated surfaces (Schmid *et al.*, 1995). These may be advantageous when the host becomes exposed to antiseptics or disinfectants (Schmid *et al.*, 1995). Because of the high prevalence of this group, the cluster is referred to as "HSP" (Highly Successful Pathogens) in the present

study. The higher prevalence of HSP strains in Candidosis patients than in healthy individuals may be explained by their potential ability to replace other commensal strains and cause infection (Schmid, 1993). It is possible that HSP strains may be more virulent than other strains due to specific pathogenicity factors.

1.5 *Candida* Pathogenicity and Virulence Factors

Opportunistic pathogens like *C.albicans* need an immunocompromised host to do more than colonise. This indicates that virulence attributes alone are not sufficient for causing infection. Nevertheless, it has been suggested that the high prevalence of HSP strains in healthy individuals and in Candidosis patients may be associated with characteristics that would be advantageous for these strains as commensal and pathogenic organisms (Schmid *et al.,* 1999). Pathogenic characteristics of *C.albicans* (adapted from Odds (1994)) are:

- rapid phenotype switching, which is based on the capability to rearrange and regulate genetic expression and to select for phenotypes suiting different host environments (Soll, 1991)
- hyphae formation (true and pseudo hyphae) (Sobel *et al.*, 1984)
- thigmotropism, the response to surface topography (Sherwood *et al.*, 1992) which may help the organism to penetrate some tissue by following surface discontinuities and microscopic breaks
- surface hydrophobicity, which influences the adherence to epithelial surfaces and the speed of germ tube formation (Hazen and Hazen, 1992; Hazen and Glee, 1994)
- surface virulence molecules, like adhesins, which are surface components responsible for the attachment of *C.albicans* cells to epithelial cells
- molecular mimicry, which is the theoretical ability of an microorganism to be coated on the surface with molecules that imitate host components and to become less recognisable by the host (Robert *et al.*, 1991)

- the production of enzymes that affect growth rate, such as lytic enzymes, e.g. secreted aspartyl proteinases (Ross *et al.*, 1990); because a slow growing organism has a better chance of being eliminated by the host defence mechanism than a fast growing organism
- less restrictive nutrient requirements; because organisms with minimal nutrient requirements are less specific for particular environments they therefore have a greater chance of finding a micro-niche than does an organism with highly specific requirements, e.g. the majority of *C.albicans* strains can grow in a simple medium containing biotin, salts, a carbon source, and a nitrogen source other than nitrate (Odds, 1988).

1.6 Thesis Overview

In developing a better understanding of *C.albicans* epidemiology, it is important to determine whether HSP strains replace other strains under certain conditions and whether the replacing strains may cause Candidosis. Addressing these questions requires suitable genetic markers to study the population biology of *C.albicans* strains. It also requires implementation of a suitable phylogenetic framework that will sufficiently well represent population diversity.

To date, measurement of the genetic diversity between HSP and other *C.albicans* strains has been based entirely on Ca3 fingerprinting data, and in Chapter 3 use of this methodology for strain replacement studies is reported on. In Chapter 4, some other new and potentially useful genetic marker systems have also been investigated. These include sequence analysis of the nuclear rDNA ITS regions (see section 4.2.1) and AFLP fingerprinting (see section 4.2.2). The phylogenetic approach of "split decomposition" (see section 2.6.1.3) is used to provide a novel means of assessing whether strain replacement has occurred.

The characterisation of AFLP and ITS data described in Chapter 4 provides the means for an independent test of the HSP cluster identified from Ca3 profiles. Also tested for are whether commensal and pathogenic isolates of this cluster occur in geographically diverse localities. In addition for their use in testing phylogenetic hypotheses, these data were used to try and identify specific PCR markers associated with the HSP cluster that might be of use in future strain replacement studies.

CHAPTER 2 - MATERIALS AND METHODS

Centrifugations were performed in a table top microcentrifuge (Eppendorf), unless otherwise stated. Since these centrifuges have approximately the same rotating radius, the same speed will yield the same RCF (relative centrifugal force measured in x g; 14 000 min $^{-1}$ ≙ 20 800 x g) (Ausubel *et al.*, 1989). Therefore only the speed in rpm was used as reference in the present thesis.

2.1 Fungal and Bacterial Strains

C.albicans strains supplied by Dr. Jan Schmid are listed in Table 2-1.

Table 2-1: *C.albicans* strains isolated in previous studies by Dr. Jan Schmid *et al.* (Schmid *et al.*, 1990; Schmid, 1993; Schmid *et al.*, 1995; Schmid *et al.*, 1999).

C.albicans strain	geographical origin	clinical status
RIHO6	USA	pathogenic
HUN125	Europe	pathogenic
CLB42	South America	pathogenic
hp10bt	USA	commensal
RIHO1	USA	pathogenic
var1.10	USA	pathogenic
YSU649	Asia	pathogenic
AU25	New Zealand	pathogenic
hp 38an	USA	commensal
CLB43	South America	pathogenic
jam 2c	USA	pathogenic
OD8826	Europe	commensal
Otg16	New Zealand	pathogenic
Fj11	Pacific Islands	pathogenic
HmHc4	USA	commensal
W13	New Zealand	pathogenic
HUN121	Europe	pathogenic
AU34	New Zealand	pathogenic
YSM1	Asia	pathogenic
Fj10	Pacific Islands	pathogenic

C.albicans strain	geographical origin	clinical status
HUN92	Europe	pathogenic
W136	New Zealand	pathogenic
HUN65	Europe	pathogenic
hp6ch	USA	commensal
hp50an	USA	commensal
hol-c	USA	commensal
var1.7	USA	pathogenic
hp5bt	USA	commensal
YSU709	Asia	pathogenic
CH20	New Zealand	pathogenic
CLB44	South America	pathogenic
YSM42	Asia	pathogenic
var1.11	USA	pathogenic
rom-c	USA	pathogenic
AU47	New Zealand	pathogenic
AU52 (standard)	New Zealand	-
3153A (standard)	unknown	-

Commensal and environmental *C.albicans* strains isolated during the present study in Palmerston North (New Zealand) are listed in Table 2-2.

Table 2-2: *C.albicans* strains isolated during this study

C.albicans isolate (multiple isolates separated by ',')	host no./ source	cell count of sample per ml	date of isolation
C1-1S 1, 2, 4	C1/saliva	510	14/03/97
C1-1R.1, 2, 3	C1/rectum	n.a.	14/03/97
C1-4R.1, 2, 3,	C1/rectum	n.a.	17/04/97
C1-4S1, 2, 3	C1/saliva	633	17/04/97
C3F1, 2, 3	C3/saliva	61	04/09/97
C3L1, 2, 3	C3/saliva	70	24/09/97
C4F1-F3	C4/saliva	3343	08/09/97
C4L1-L3	C4/saliva	6×10^4	23/10/97
C7F1, 2, 3	C7/saliva	200	24/09/97
C7L1, 2, 3	C7/saliva	168	20/10/97
C8F1, 2, 3	C8/saliva	240	05/11/97
C8L1, 2, 3	C8/saliva	1095	26/11/97
C9F1, 2, 3	C9/saliva	32	12/11/97
C10F1, 2, 3	C10/saliva	2000	24/11/97
C13F1, 2, 3	C13/saliva	1.03×10^4	10/12/97
C13L1, 2, 3	C13/saliva	10	06/01/98
C14F1, 2, 3	C14/saliva	885	17/12/97
C14L1, 2, 3	C14/saliva	155	12/02/98
C15F1, 2, 3	C15/saliva	278	03/02/98
C15L1	C15/saliva	105	25/02/98
C16F1	C16/saliva	5	03/02/98
C16L1, 2, 3	C16/saliva	47	25/02/98
C21F1, 2, 3	C21/saliva	1.3×10^4	16/02/98
C21L1, 2, 3	C21/saliva	2×10^4	18/03/98
C22F1, 2, 3	C22/saliva	228	25/2/98
C22L1, 2, 3	C22/saliva	185	18/03/98
C23F1,2,3	C23/saliva	60	02/03/98
C24F1,2,3,	C24/saliva	465	02/03/98
A26.1, 2, 4, 6, 7, 8, 9, 10, 13	A26/saliva	n.a.	26/02/97
FM1-1S1,2,3,	FM1/saliva	n.a.	14//03/97
FM1-4S1,2,3	FM1/saliva	n.a.	17/04/97
CH2F1, 3	CH2/saliva	3000	03/09/97
CH2L1, 3	CH2/saliva	2×10^4	10/09/97
CH4F1, 3	CH4/saliva	155	08/09/97
CH4L1, 3	CH4/saliva	141	16/09/97
CH5F1, 3	CH5/saliva	1.3×10^4	10/09/97
CH6F1, 2, 3	CH6/saliva	80	10/09/97
CH7F1, 3	CH7/saliva	5.5×10^4	11/09/97
CH7L4, 5	CH7/saliva	150	18/09/97
CH9L1, 2, 3	CH9/saliva	1350	24/09/97

C.albicans isolate (multiple isolates separated by ',')	host no./ source	cell count of sample per ml	date of isolation
CH11L1, 3, 4	CH11/saliva	332	25/09/97
CH13F1, 2, 3	CH13/saliva	5×10^4	08/10/97
CH13L1, 2, 3	CH13/saliva	3000	15/10/97
CH14F1, 2, 3,	CH14/saliva	137	09/10/97
CH14L1, 2, 3	CH14/saliva	3000	16/10/97
CH16F1, 2, 3	CH16/saliva	425	09/10/97
CH16L1, 2, 3	CH16/saliva	440	16/10/97
CH18F1, 2, 3	CH18/saliva	710	15/10/97
CH18L1	CH18/saliva	495	22/10/97
CH19F1	CH19/saliva	5	16/10/97
CH19L4	CH19/saliva	531	23/10/97
CH20F5, 7	CH20/saliva	342	16/10/97
CH20L3	CH20/saliva	4258	16/10/97
CH21F1, 2, 3	CH21/saliva	7057	22/10/97
CH22F1	CH22/saliva	1893	22/10/97
CH22L1	CH22/saliva	7730	29/10/97
CH23F1	CH23/saliva	6754	22/10/97
CH23L1	CH23/saliva	1×10^4	29/10/97
CH24F1, 3	CH24/saliva	205	22/10/97
CH24L1	CH24/saliva	260	29/10/97

n.a. = not available

Bacterial strains and λ-clones used in this study are listed in Table 2-3.

Table 2-3: Bacterial strains and λ-clones used in this study

strain	charcteristics	reference/source	application
E.coli DH5 α	competent cells	Gibco BRL	cloning
E.coli C600	lambda sensitive	Gibco BRL	preparation of Ca3 probe
clone λCa3	isolated from a λgt 10 clone library of strain 3153A ⇒ Ca3 DNA	(Soll *et al.*, 1988)	preparation of Ca3 probe

2.2 Media and Supplements

All media prepared for fungal and bacterial culture during this study are listed in Table 2-4. Water used for preparation of media was purified by the MilliQ

Reagent Water System (Millipore). Media were subsequently sterilised at 121°C for 20 min. For preparation of solid media, if not stated otherwise,15 g agar or agarose per litre were added.

Table 2-4: Media for fungal and bacterial culture

medium	composition per litre	reference
Luria-Bertani broth (LB)	10 g tryptone (Difco), 5 g yeast extract (Merck), 5 g NaCl (BDH), 1 ml 1M NaOH (Sigma)	(Sambrook *et al.*, 1989)
lambda broth	10 g tryptone (Difco), 2.5 g NaCl (BDH), if required, 2.46 g $MgSO_4$ (Sigma), 2 mg maltose (BDH),	(Ausubel *et al.*, 1989)
lambda top agar	10g tryptone (Difco), 2.5 g NaCl (BDH), 7g agar (BDH)	(Ausubel *et al.*, 1989)
suspension medium	5.8 g NaCl (BDH), 2 g $MgSO_4$ (Sigma), 50 ml 1M Tris-HCl, pH 7.5, 0.01% gelatine (Merck)	(Ausubel *et al.*, 1989)
Sabouraud (SAB)	10 g neopeptone, 40 g glucose, pH 5.0 (Oxoid);	Oxoid
e.g., SAB ++	SAB + 1mg/l cycloheximide, 100mg/l chloramphenicol	
yeast-peptone-dextrose medium (YPD)	20 g dextrose (BDH), 20 g bacto-peptone (Difco), 10 g yeast extract (Difco)	(Sambrook *et al.*, 1989)
freezing medium	65 % (w/v) glycerol (Sigma), 0.1 M $MgSO_4$, 0.025 M Tris-HCl, pH 8.0	(Ausubel *et al.*, 1989)

Supplements and antibiotics added to the media (if required) are listed in Table 2-5.

Table 2-5: Supplements and antibiotics

supplement	stock concentration	dissolving agent	mode of action / characteristics
cycloheximide	10 mg/ml	ethanol	inhibition of peptyltranferase of cytoplamic ribosomes; suppression of filamentous fungi
chloramphenicol	100 mg/ml	methanol	bacteriostatic; inhibits protein synthesis by interacting with 50Sribosomal subunit and inhibiting tne peptidyltransferase reaction
ampicillin	100 mg/ml	ethanol	bacteriocidal; inhibits cell wall synthesis by inhibiting formation of the peptidoglycan cross-link; kills growing bacterial cells
5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside (Xgal)	50 mg/ml	N, N dimethylformamide	noninducing chromogenic substrate of ß-galactosidase; cleavage of Xgal results in blue colour
isopropyl-1-thio-ß-D-galactoside (IPTG)	100 mM	sterile H_2O	effective inducer

2.3 Fungal and Bacterial Cultures

2.3.1 *Candida albicans*

Candida albicans isolates were cultured either on Sabouraud (SAB) or YPD (see section 2.2) agar or alternatively in liquid broth on a shaker at 200 rpm at 37°C. For isolation of *C.albicans* strains samples were cultured on SAB++ plates. *C.albicans* cultures were maintained on slants containing SAB medium at 4°C. *C.albicans* culture stocks were maintained at – 70°C in 50% freezing medium (see section 2.2).

2.3.2 Germ Tube Assay

For the rapid identification of *C.albicans* the traditional germ tube test was applied, in which *C.albicans* alone produces hyphal outgrowth from blastospores when incubated at 37°C in serum (Taschdjian *et al.*, 1960). A single colony which was grown on SAB plate (see section 2.2) was inoculated in 300 µl bovine serum using a thin inoculating loop, and incubated for 2.5-3 h at 37°C. To monitor the production of germ tubes approximately 15 µl of the cell mixture was applied to a slide and viewed on a phase contrast microscope at 400x.

2.3.3 *Escherichia coli*

For cloning experiments *Escherichia coli* was grown on Luria-Bertani (LB) agar, or in liquid broth on a shaker at 200 rpm at 37°C. For preparation of phage plaques and lysates *E.coli* was grown in lambda broth or alternatively in lambda top agar (see section 2.2) at 37°C. *E.coli* cultures were stored at –70°C in 50% (w/v) freezing medium (see section 2.2).

2.4 Patient Selection and Sampling Procedures

Patient criteria were based on their risk of acquiring Candidosis.

2.4.1 Patient Selection

2.4.1.1 Criteria 1 ("Healthy") Individuals

- no signs of compromised immunocompetence towards *C.albicans*, e.g. neutrophil count over 2.0×10^9/l.
- nonhospitalised (or at day of admission within the first 3 hours), last hospitalisation longer than 1 year ago,

- asymptomatic for Candidosis
- must not suffer from any underlying disease known to be associated with immunosuppression
- has not recently undergone antimicrobial therapy

2.4.1.2 Criteria 2 (Highest Risk) Patients

- diagnosis of multiple myeloma or chronic leukaemia
- neutrophil count at first sampling at least 2.0×10^9/l
- asymptomatic for Candidosis
- no former treatment with chemotherapy
- sampling possible before and after first course of chemotherapy

2.4.1.3 Criteria 3-5 (Lower Risk) Patients

Patient groups were categorised according to their risk of acquiring Candidosis. These groups were surveyed for changes in the *C.albicans* flora as indicator groups for possible strain replacement.

2.4.1.3.1. Criteria 3 (Hospitalised) Patients

- asymptomatic with respect to Candidosis
- immunocompetent with respect to *C.albicans*; neutrophil count over 2.0×10^9/l.
- sampling possible before and after hospitalisation for at least 7 days, surgery (hip- or knee replacement), and antibiotic treatment (HAS).

Note: Isolates obtained before HAS were also part of criteria 1 group.

2.4.1.3.2. Criteria 4 (Solid Cancer) Patients

In an early stage, nonhaematological solid cancer patients have presumably a normal host defence against *C. albicans* (before chemotherapy treatment) (Ortiz and Stoliar, 1988; Dollinger and Rosenbaum, 1998). Due to chemotherapy treatment, the major host defence becomes impaired and strain replacement might then occur.

- neutrophil count at first sampling preferably at least 2.0 x 10^9/l
- diagnosis of solid malignancy, by preference in an early stage
- asymptomatic for candidosis
- no former treatment with chemotherapy
- sampling possible before and after first course of immunosuppressive chemotherapy

2.4.1.3.3. Criteria 5 (Haematological Cancer) Patients

The last group was considered highly predisposed to candidosis because the major host defence against *C.albicans* was already impaired due to disease and became more impaired due to immunosuppressive chemotherapy. The requirements for highly predisposed patients were:

- neutrophil count at first sampling preferably at least 2.0 x 10^9/l
-diagnosis of haematological cancer (except multiple myeloma or chronic leukaemia)
- asymptomatic for candidosis
- no former treatment with chemotherapy
- sampling possible before and after first course of chemotherapy

For the prevention of microbial infections (bacterial, viral, and fungal) cancer patients (moderately and highly predisposed patients) were treated during

chemotherapy with mouthwashes containing chlorohexidine, a bacteriocidal, fungistatic, and virostatic agent.

All individuals sampled throughout this study are listed in Table 2-6.

Table 2-6: Individuals sampled throughout this study

individual	sex / age	underlying disease / treatment	neutrophil count	relevance for this study (please refer to section 2.4.)
A26	male / 21	healthy	n.a.	criteria 1 individual
CH2	female / 72	HAS[1]	3.22 x10^9 / l	criteria 1 and 3 individual
CH4	male / 61	HAS[1]	4.69 x10^9 / l	criteria 1 and 3 individual
CH5	female / 67	HAS[1]	1.3 x10^9 / l	no detectable colonisation with *C.albicans*
CH6	female / 67	HAS[1]	5.5 x10^9 / l	criteria 1 individual
CH7	male / 66	HAS[1]	4.8 x10^9 / l	criteria 1 and 3 individual
CH8	male / 70	HAS[1]	3.96 x10^9 / l	no detectable colonisation with *C.albicans*
CH9	female / 60	HAS[1]	4.31 x10^9 / l	no detectable colonisation with *C.albicans*
CH10	female / 84	HAS[1]	4.72 x10^9 / l	no detectable colonisation with *C.albicans*
CH11	male / 77	HAS[1]	n.a.	no detectable colonisation with *C.albicans*
CH12	female / 73	HAS[1]	n.a.	no detectable colonisation with *C.albicans*
CH13	female / 42	HAS[1]	7.81 x10^9 / l	criteria 1 and 3 individual
CH14	female / 79	HAS[1]	5.73 x10^9 / l	criteria 1 individual
CH15	female / 74	HAS[1]	4.52 x10^9 / l	no detectable colonisation with *C.albicans*
CH16	female/ n.a.	HAS[1]	5.10 x10^9 / l	excluded; was hospitalised 5 month ago
CH17	female / 71	HAS[1]	3.66 x10^9 / l	no detectable colonisation with *C.albicans*
CH18	male / 72	HAS[1]	n.a.	criteria 1 individual
CH19	female / 69	HAS[1]	4.78 x10^9 / l	criteria 1 and 3 individual
CH20	male / 82	HAS[1]	2.78 x10^9 / l	criteria 1 individual
CH21	male / 67	HAS[1]	n.a.	excluded, no neutrophil count
CH22	male / 75	HAS[1]	3.46 x10^9 / l	criteria 1 and 3 individual
CH23	female / 84	HAS[1]	2.2 x10^9 / l	criteria 1 and 3 individual
CH24	female / 67	HAS[1]	3.52 x10^9 / l	criteria 1 individual
CH25	female / 71	HAS[1]	n.a.	no detectable colonisation with *C.albicans*
FM1	female / 57	healthy	n.a.	partner of C1
C1	male / 58	multiple myeloma / ABCM[2]	2.07 x10^9 / l	criteria 2/ cancer patient
C3	male / 41	lymphoma / COP[3]	3.05 x10^9 / l	criteria 5/ cancer patient
C4	female / 44	breast cancer / A[4], C[5]	5.4 x10^9 / l	criteria 4/ cancer patient
C6	male / 71	colon cancer / 5FU[6], FA[7]	2.88 x10^9 / l	criteria 4/ cancer patient
C7	male / 59	colon cancer / 5FU[6], FA[7]	6.87 x10^9 / l	criteria 4/ cancer patient
C8	female / 55	breast cancer / A[4], C[5]	2.7 x10^9 / l	criteria 4/ cancer patient
C9	female / 57	non Hodgkin lymphoma / CHOP[8]	3.3 x10^9 / l	criteria 5/ cancer patient
C10	female / 59	breast cancer / A[4], C[5]	4.86 x10^9 / l	criteria 4/ cancer patient
C11	male / 70	rectum cancer / 5FU[6], FA[7]	5.03 x10^9 / l	no detectable colonisation with *C.albicans*
C12	female / 66	colon cancer/ 5FU[6] / radiotherapy	4.23 x10^9 / l	no detectable colonisation with *C.albicans*
C13	female / 62	rectum cancer/ 5FU[6]	5.7 x10^9 / l	criteria 4/ cancer patient
C14	female / 53	colon cancer / 5FU[6],	6.1 x10^9 / l	criteria 4/ cancer patient

individual	sex / age	underlying disease / treatment	neutrophil count	relevance for this study (please refer to section 2.4.)
C15	female / 40	breast cancer / A[4], C[5]	n.a.	no detectable colonisation with *C.albicans*
C16	female / 41	breast cancer / A[4], C[5]	4.9 x10^9 / l	criteria 4/ cancer patient
C17	female / 48	breast cancer / A[4], C[5]	5.0 x10^9 / l	no detectable colonisation with *C.albicans*
C18	female / 78	ovary cancer / carboplatin	5.08 x10^9 / l	no detectable colonisation with *C.albicans*
C19	male / 66	colon cancer / 5FU[6], FA[7], radiotherapy	6.3 x10^9 / l	no detectable colonisation with *C.albicans*
C20	male / 72	small cell lung cancer / carboplatin, etoposide	10.27 x10^9 /l	no detectable colonisation with *C.albicans*
C21	male / 71	rectum cancer / 5FU[6]	2.1 x10^9 / l	criteria 4/ cancer patient
C22	female / 48	breast cancer / A[4], C[5]	5.1 x10^9 / l	criteria 4/ cancer patient
C23	female /47	non Hodgkin lymphoma / CHOP[8]	4.2 x10^9 / l	criteria 5/ cancer patient
C24	female / 60	rectum cancer / 5FU[6]	3.4 x10^9 / l	criteria 4/ cancer patient
C25	female / 54	ovary cancer / carboplatin	5.1 x10^9 / l	no detectable colonisation with *C.albicans*
C26	male / 50	rectum cancer /5FU[6], radiotherapy	4.1 x10^9 / l	no detectable colonisation with *C.albicans*
C27	female / 47	breast cancer / mitomycin, methotrexate, mitozantrane	2.8 x10^9 / l	no detectable colonisation with *C.albicans*
C28	male / 57	rectum cancer / 5FU[6]	3.03 x10^9 / l	no detectable colonisation with *C.albicans*

n.a. = not available

[1] HAS = 7 days of hospitalisation, antibiotic treatment, and surgery (hip or knee replacement)

[2]ABCM = combination therapy of four immunosuppressive anticancer drugs: adriamycin, BCNU (cisplatin), cyclophosphamide, melphalan (Oldham, 1991)

[3]COP = combination therapy of two immunosuppressive anitcancer drugs cyclophosphamide, vincristine, and the adrenocorticoid predisone (Oldham, 1991)

[4] A = adriamycin, immunosuppressive anticancer drug (Oldham, 1991)

[5] C = cyclophosphamide, immunosuppressive anticancer drug (Oldham, 1991)

[6] 5FU= 5-fluorouracil, immunosuppressive anticancer drug (Oldham, 1991)

[7] FA = folinic acid, immunosuppressive anticancer drug (Oldham, 1991)

[8] CHOP = combination therapy of three immunosuppressive anticancer drugs cyclophosphamide, doxorubicin, vincristine, and the adrenocorticoid predisone (Oldham, 1991)

2.4.2 Sampling Procedures

To take into account endo- and exogenous sources of replacement strains, different body locations (saliva, hands, and rectum) of patient 1 (C1) as well as cohabitating individuals (saliva, hands), and their inanimate environment (hospital formites) were sampled. However, to include a greater number of patients for valid statistical analysis, sampling of the following patients (A26, CH2-CH23, C3-C28)

was reduced to the oral cavity. From individuals colonised with *C.albicans*, one to two *C.albicans* clones of each specimen were fingerprinted and screened for an indication of replacement.

Preliminary experiments were performed to evaluate the recovery rates of different sampling techniques. A cell culture of strain 3153A was grown overnight at 37°C and diluted with saline (0.9% NaCl-solution) to final concentrations of 10^3-10^7 cfu/ml. The cell concentrations per ml were determined by plating 100 µl of each dilution onto a SAB- plate and counting colonies grown overnight at 37°C (viable cell count). To compare the recovery rates of direct and indirect swabs, a cotton swab was dipped in 1 ml of each dilution and the suspension was spread either directly (9 strikes) or indirectly onto a SAB- plate. Indirectly spread means that the swab was vortexed for 1 min in 500 µl sterile saline and 100 µl of this suspension was spread onto a SAB- plate. Plates were incubated overnight at 37°C. Colonies were counted and the recovery rate for each dilution was calculated. In these preliminary experiments the indirect technique gave better recovery rates on average than the direct technique (data not shown).

2.4.2.1 Saliva

Patients were asked to spit at least 1 ml saliva into a 50 ml falcon tube. The saliva samples were stored at 4°C and transported within a maximum of 4 h to the laboratory. By viable cell count it was determined that no significant growth occurred during that time at 4°C. For determination of colonisation intensity the saliva was diluted 1:2, 1:5 and 1:10 in sterile saline, spread onto SAB++ plates (see section 2.2), and incubated at 37°C. To test for susceptibility to cycloheximide and chloramphenicol, 200 µl of neat saliva was plated onto SAB– plates (see section 2.2). Colonisation intensity was determined after at least 24 h of incubation.

2.4.2.2 Hands

Hands were sampled by Dr. Baker (Consultant Haematologist, Palmerston North Hospital) with a modified tape stripping method previously described by Barnetson *et al.* (Barnetson and Milne, 1973). An imprint of hands was taken by a stamp covered with double sided tape and pressed subsequently onto a SAB–plate (see section 2.2). Colonies were counted after at least 24 h of incubation at 37° C. Negative plates were incubated for at least 7 days.

2.4.2.3 Rectum

Rectal samples were obtained by Dr. Baker by gently rubbing on the rectal surface with a sterile cotton swab (Culturette Swab Transportsystem, Difco laboratories, USA). Immediately after sampling, each swab was placed in a sterile containment tube and moistened with sterile salt solution by crushing the ampoule in the tube. The tubes were stored at 4°C and transported within 4 h of sampling to the laboratory. Each swab was inserted into 0.5 ml of sterile saline in a universal bottle, vigorously mixed for 1 min using a laboratory table top vortex mixer, and 200 µl of the suspension was spread onto SAB- and SAB++ medium (see section 2.2). The plates were incubated at 37°C for at least 4 days.

2.4.2.4 Environment

Environmental samples were taken using a sterile cotton swab premoistened with sterile saline (0.9% (w/v) NaCl), and directly spread on SAB++ plates (see section 2.2). The air of each patient room was sampled using air settle plates containing SAB++ medium, for 24 h. The following facilities on ward 25 (haematology ward) were sampled twice:

- clean utility (sinks, table, trundler, door, handle of cupboard),
- kitchen (sinks, taps, handles of fridge, lid of rubbish bin, table)

- dirty utility (sinks, table, washing machine, soap donor, bed pan washer, clean bed pan)
- male bathroom (sinks, bath tub, chairs, showers, toilet seats)
- unisex toilet (sinks, toilet seats, showers, chairs)
- female toilet (sinks, toilet seats, chairs, showers)
- visitor toilet (sink, toilet seat)
- patient rooms 1-12 (sink, bed, table, chair)
- air settle plates in patient rooms 1-12

2.5 DNA Analysis Methods

2.5.1 DNA Isolation

2.5.1.1 Genomic DNA

Genomic *C.albicans* DNA was isolated from each strain by modified method of Scherer and Stevens (Scherer and Stevens, 1987). A single colony was inoculated in 25 ml YPD (see section 2.2) and grown overnight in 125 ml Erlenmeyer flasks at 200 rpm at 37°C. Cells of 5 ml saturated culture were pelleted by spinning for 10 min at 5000 xg (Beckmann centrifuge), resuspended in 1 ml of 1 M sorbitol, transferred to a 1.5 ml eppendorf tube, and spun at 4000 rpm for 5 min. The pellet was resuspended in 1 ml of spheroblast buffer {1 mM sorbitol, 50 mM potassium phosphate (pH 7.5), 0.1% (w/v) ß-mercaptoethanol, 200 U lyticase (Sigma)}, vigorously mixed by vortexing, and incubated at 30°C until cells had turned into spheroblasts (about 1 h). Spheroblasts were harvested and lysed in 0.5 ml lysis buffer {50 mM EDTA (pH 8.5), 2 % (w/v) SDS} by incubation at 70-80°C for 30 min. After lysis, 50 µl of 5 M potassium acetate (pH 5.8) were added and the solution put on ice for at least 30 min. Cell debris was pelleted and the supernatant poured into 1 ml absolute ethanol to precipitate DNA, spun down briefly, and washed with 70% (w/v) ethanol. DNA was resuspended in TE buffer {10 mM Tris-HCl pH 8.0 / 1 mM EDTA} and treated with 100 µg RNase A (Sigma) per ml. DNA was then precipitated with 2 volumes of 2-isopropyl alcohol (Sigma), washed with 70% (w/v) ethanol, and resuspended in 50 µl TE buffer.

2.5.1.2 Plasmid DNA

Plasmid preparations were performed by the method described by Holmes and Quigley (Holmes and Quigley, 1981). A single colony of transformed *E.coli*

DH5α cells was grown overnight in LB broth supplemented with ampicillin (final concentration 100 µg/ml). An aliquot of 1.5 ml was pelleted by centrifugation at 14000 rpm in a microcentrifuge (Eppendorf) at room temperature (RT) for 1 min, and subsequently resuspended in 400 µl STET buffer {10 mM Tris-HCl (pH 8.0) 0.1 M NaCl, 1 mM EDTA, 5% (w/v) Triton x-100}. The suspension was treated with 0.2 mg/ml RNase A (Sigma) and 2 mg/ml lysosyme (Sigma), left for 5-10 min at RT and then boiled for 40 sec, chilled on ice for 60 sec and immediately spun at 14000 rpm for 10 min at RT. The cell debris was removed using a sterile toothpick and plasmid DNA precipitated for 5 min at RT by addition of 40 µl 3 M sodium acetate and 420 µl 2-isopropyl alcohol. DNA was pelleted by centrifugation for 10 min at RT, washed with ethanol, and resuspended in 25-50 µl TE buffer {10 mM Tris-HCl (pH 8.0) / 1 mM EDTA}. If plasmid was required for sequence analysis, a phenol/chloroform extraction was performed, followed by ethanol precipitation (see section 2.5.3.2), and the air dried pellet was resuspended in 20-50 µl H_2O.

2.5.2 Restriction Enzyme Digestion

For Ca3 fingerprinting experiments, 10.8 µg *C.albicans* DNA were digested using 15-20U *Eco*RI (New England Biolabs) in a volume of 40 µl. Digestion was performed for 4-5 h at 37°C in a water bath using the supplied buffer (New England Biolabs). To test for complete digestion 20 µl of 3 x loading buffer (see section 2.5.5.1) was added and a 15 µl aliquot run on a test gel. As an additional test for complete digestion and to ensure equal concentrations of DNA from different strains, Ca3 blots were stripped (see section 2.5.6.8) and reprobed using part of the single copy gene actin (see section 2.5.7.1).

For other restriction enzyme analysis DNA was digested according to the recommendations of the respective manufacturer for 1.5-2 h using the supplied buffer. To test for complete digestion an aliquot was run on a mini agarose gel (see section 2.5.5.1).

2.5.3 Purification of DNA

2.5.3.1 Phenol/Chloroform Purification of DNA

To remove proteins from DNA solutions an equal volume of phenol (tris-equlibrated phenol, pH 8.0) was added, the mixture vortexed vigorously, and then centrifuged at 14000 rpm in a microcentrifuge (Eppendorf) for 10 min at RT. The aqueous phase was transferred to a fresh tube and extracted with an equal volume 50% (v/v) phenol/chloroform. The mixture was vortexed vigorously, and centrifuged at 14000 rpm for 10 min at RT. This extraction was repeated until no protein was visible at the interface of the organic and aqueous phases. Traces of phenol were removed by mixing with an equal volume of chloroform/isoamyl (24:1) and centrifuged as described above (Ausubel *et al.*, 1989). The upper phase was collected and concentrated by ethanol precipitation (see section 2.5.3.2.).

2.5.3.2 Concentration of DNA by Ethanol Precipitation

To concentrate DNA solutions, 1/10 volume of 3 M sodium acetate pH 4.8, and 2.5-3 volumes of absolute ethanol were added, and chilled for either 30 min at –20 °C or 15 min at –80°C. The DNA was pelleted by centrifugation for 15 min, washed with an equal volume of 70% (v/v) ethanol, subsequently dried (vacuum or air) and resuspended in TE buffer {10 mM Tris-HCl (pH 8.0), 1 mM EDTA} or H_2O (Ausubel *et al.*, 1989).

2.5.3.3 Column Purification of PCR Products

PCR products were purified using the QIAquick PCR Purification kit (Qiagen) or the Concert PCR Purification kit (Life Technologies) according to the instructions of the manufacturers. Briefly, the DNA sample was diluted in the

respective binding buffer, mixed, and bound to the silica-gel membrane in the presence of chaotropic salts by centrifugation. One washing step was performed using the respective washing solutions supplied by the manufacturer, and the bound DNA was eluted using low salt solutions, e.g., TE buffer {10 mM Tris-HCl pH 8.0 / 1 mM EDTA} or H_2O.

2.5.3.4 Column Purification of Gel Extracted Products

DNA fragments were fractionated by agarose gel electrophoresis (see section 2.5.5.1), cut out and purified from agarose gels using the QIAquick Gel Extraction Kit (Qiagen) following the supplier's protocol. Firstly, the desired gel fragment was excised using a clean scalpel blade, and dissolved in binding buffer (100 µl per 100 mg gel) by incubation at 50°C for 10 min. DNA purification then proceeded as described above (see section 2.5.3.3).

DNA fragments separated by polyacrylamide gel electrophoresis were excised using a scalpel from the air dried gel. Gel fragments were soaked in H_2O, vortexed vigorously, and stored overnight at 4°C. The desired DNA fragment was then reamplified by PCR (see section 2.5.7) using the appropriate primers.

2.5.4 DNA Quantification

The concentration of genomic *C.albicans* DNA for Ca3 fingerprinting was determined by fluorometric assay in a Hoefer scientific TKO 100 Fluorometer (Hoefer) according to the manufacturer's instructions. Briefly, DNA was diluted in working buffer B {10 mM Tris-HCl; 1mM Na_2EDTA; 100 mM NaCl, pH 7.4; 1 µg/ml Hoechst 33258 dye} as recommended by the manufacturer for DNA samples with concentrations between 100 and 2000 ng/ml. Standard DNA dilutions (200 µg/ml, 1000 µg/ml, and 1500 µg/ml) were prepared with calf thymus DNA from a 1mg/ml

stock solution (Hoefer) in TE buffer {10 mM Tris-HCl pH 8.0 / 1 mM EDTA}. After the fluorometer had prewarmed for 15 min, the scale was set to zero with 2 ml working buffer B, and then set to 1000 after adding 2 µl of the 1000 µg/ml standard DNA dilution. This adjustment was repeated until the display read 1000 using the 1000 µg/ml standard DNA dilution. An aliquot of 2 µl per sample was added to working buffer B, the reading recorded, and subsequently divided by the dilution factor (1000), giving the concentration of DNA in µg/µl of the sample. If it was not possible to scale the fluorometer reliably to 1000, the readings for the standard DNA dilutions were recorded and a calibration curve defined. Based on this calibration curve the DNA concentrations of the samples were then determined.

2.5.5 Gel Electrophoresis

2.5.5.1 Agarose Gel Electrophoresis

For DNA visualisation and size fractionation, horizontal agarose gel electrophoresis was performed in a Horizon® 58 mini gel apparatus (Life Technology) following standard protocols described by Ausubel *et al.,* (Ausubel *et al.*, 1989). Typically, agarose was added to 1 x TAE (40 mM Tris, 20 mM glacial acetic acid, 1 mM EDTA), to a concentration of 0.8% (w/v) depending on the size of the fragments to be resolved. The mixture was microwaved until the agarose was completely dissolved, and cooled down to approximately 55°C before being poured into the apparatus. DNA samples were mixed with loading dye {0.5% (w/v) bromophenol blue, 0.5% (w/v) xylene cyanol, 50% (w/v) glycerol, 50 mM EDTA} and loaded in wells. Electrophoresis was performed at 70-90 V.

Large agarose gels (20.5 cm x 14 cm) of 0.8%-1.2% (w/v) in 1 x TBE buffer {0.89 mM Tris-HCl, 2.5 mM Na_2EDTA, 89 mM boric acid, pH 8.3} were prepared as described above using a Hoefer Submarine Large Gel Unit (Hoefer). Loading buffer Type 2 {25% (w/v) Ficoll 4000, 0.25% (w/v) bromophenol blue, 0.25% (w/v)

xylene cyanol} was added to DNA samples and loaded into the wells of a large gel. The gel was run at 25-70 V for 4-20 hours, e.g., 0.8% gels prepared for Ca3 fingerprinting were run at 30 V until the bromophenol blue marker had run 16 cm (16-20 h).

Following electrophoresis, gels were stained in ethidium bromide solution (5 µg/ml ethidium bromide in 1 x electrophoresis buffer) for 10-30 min, and destained in deionised H_2O for 5 min. DNA bands were visualised directly by illumination with UV light on a TMW-20 Transilluminator (Alpha Innotech), and photographed using the IS-1000 Digital Imaging System (Alpha Innotech).

2.5.5.2 Polyacrylamide Gel Electrophoresis

For Amplified Fragment Length Polymorphisms (AFLP) fingerprinting DNA was separated on 5% (w/v) denaturing polyacrylamide gels according to the method by Maxam and Gilbert (Maxam and Gilbert, 1980). The gels consisted of 5 % (v/v) Long Ranger gel solution (FMC Bio Products), 8 M urea in 1 x TBE (0.89 mM Tris-HCl, 2.5 mM Na_2EDTA, 89 mM boric acid, pH 8.3). The gel solution was filtered through two sheets of Whatman paper. Before pouring 350 µl of 10% (w/v) ammonium persulphate (APS) stock solution and 35 µl NNN'N'tetramethylethylenediamine (TEMED) was added to the solution and mixed. All solutions were prepared using deionised H_2O. Gels were poured between clean glass plates separated by 0.4 mm spacers. To prevent the gel from sticking to both glass plates the longer plate was coated with an application of Rainex® and the shorter plate was treated with 2 ml of 95% ethanol, 0.5% glacial acetic acid, and 0.05% (v/v) Bind-Silane (Pharmacia Biotech). The gel solution was poured between the glass plates, sharktooth combs inserted upside down, and stored after polymerisation, overnight at 4°C covered in saran wrap. The gel apparatus was assembled using the Horizon® 11•14 Gel Apparatus (Gibco BRL) according to the recommendations of the manufacturer, the unit filled with 1 x TBE

as running buffer, and the combs were removed. To prewarm the gel it was run without samples at 55 W for 30 min. DNA samples and standards were mixed with 1/2 volume formamide loading buffer {98% (v/v) formamide, 10 mM EDTA, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol}, denatured at 94°C for 4 min, and cooled on ice before loading. Electrophoresis was performed at 55W till the bromophenol dye front had run about 5 cm from the base of the gel.

The gel was stained according to a modified DNA Silver Staining System Manual (Promega) as follows: First the gel was fixed in 10% (v/v) acetic acid for at least 2 h, rinsed 3 x for 5 min in deionised water, before being placed in a silver staining solution {6 mM $AgNO_3$, 0.15 % (v/v) formaldehyde (37%, Sigma)} with agitation for 30–60 min. The banding patterns were developed in pre-chilled developer {280 mM Na_2CO_3, 0.16% (v/v) formaldehyde (37%, Sigma)}and 50 µM sodium thiosulphate (BDH). When the desired staining intensity and low background coloration was achieved the reaction was stopped by adding chilled 10% (w/v) acetic acid for 2-3 min. The gel was then rinsed in deionised H_2O and air dried at room temperature. The banding pattern was scored for the presence (1) and absence (0) of bands using a light box and entered into a computer datafile. Further analysis was performed as described below (see section 2.6.1).

2.5.6 Southern Blotting and Hybridisation

2.5.6.1 Southern Blotting

Capillary Southern transfer of DNA was based on the method of Southern (Southern, 1975).

DNA fragments were separated electrophoretically on conventional agarose gels (see section 2.5.5.1). Following electrophoresis, the gel was equilibrated for 15 min at room temperature in 0.2 M HCl for depurination. This step was followed

by denaturation, twice for 15 min in 0.5 M NaOH/1.5 M NaCl, with subsequent neutralisation for 30-40 min in 0.5 M Tris-HCl, (pH7.5) 1.5 M NaCl.

The capillary DNA transfer was performed overnight either to a nitrocellulose membrane (BA 84/0.45 µm, Schleicher and Schuell) or a nylon membrane (positively charged nylon membrane, Boehringer Mannheim/Roche) as described by Sambrook *et al.* (Sambrook *et al.*, 1989) with 10 x SSPE {1.8 M NaCl, 0.1 M Na_2HPO_4/NaH_2PO_4, 10 mM Na_2EDTA; pH 7.0} when radiolabelled probe was used, or 20 x SSC buffer {3 M NaCl, and 0.3 M sodium citrate}, when nonradiolabelled probe was used. The transferred DNA was covalently bound to the membrane by UV irradiation for 3.5 min on a TMW-20 transilluminator (Alpha Innotech) and afterwards rinsed in 2 x SSPE or 2 x SSC, buffer respectively, for 10 min at RT, and finally airdried.

2.5.6.2 Preparation of Ca3 DNA For Probe Making

In previous studies, a 15 kb *Eco*RI fragment containing the 11 kb repetitive Ca3 sequence had been isolated from a λgt10 library (Schmid *et al.*, 1990). DNA was isolated from a phage lysate as follows: For the preparation of single plaques, 0.2 ml of a saturated *E.coli* cell cuture was added to 2 ml of melted top agar (see section 2.2), and poured evenly on to a prewarmed agarose LB plate (see section 2.2). After the top agar had hardened, plates were cooled for at least 15 min at 4°C, and subsequently 200 µl of λ-phage stock culture was streaked out using a thin loop wire. Plates were incubated at 37°C until single plaques appeared (10-16 h).

Single plaques were used to prepare a stock of phages by plate lysis as follows. A single colony of lambda-sensitive *E.coli* strain C600 was grown to a density of 2-3 x 10^8/ml (OD_{600} of 0.4) in LB medium supplemented with 0.2% (w/v) maltose (see section 2.2). From this culture 1 ml aliquots were transferred to test

tubes. Single fresh phage plaques were picked using a toothpick, added to the cell cultures, and vortexed lightly for 10 sec. To each test tube 2 ml of melted top agar (see section 2.2) cooled to about 45°C, were added, and poured on prewarmed agarose lambda plates (see section 2.2). Plates were incubated 37 °C until 90-100% of lawn was lysed. For suspension, 2 ml of SM (see section 2.2) was added, the top agar scraped into a centrifuge tube, mixed with approximately 3 drops of chloroform per sample, and incubated for 10 min at room temperature. The resulting lysate was collected and the cell debris removed by centrifugation at 10 000 xg (Sovall centrifuge RC-5C, rotor SS-34) for 10 min at 4°C. The supernatant was saved, treated with DNase and RNase to a final concentration of 1 µg per ml, and incubated at 37°C for 1 h. This treatment degraded bacterial DNA and RNA during lysis, while phage DNA was still packed in phage heads. Phages were pelleted by spinning at 48 000 xg (Sorvall centrifuge RC-5C, rotor SS-34) for 2 h 15 min at 4°C. Phage DNA was subsequently extracted from the capsids by phenol/chloroform extraction (see section 2.5.3.1), concentrated by ethanol precipitation (see section 2.5.3.2), and finally resuspended in 100 µl TE buffer {10 mM Tris-HCl (pH 8.0), 1 mM EDTA}.

2.5.6.3 Radioactive [^{32}P] Probe Labelling

Radiolabelled probe was prepared using the nick translation kit (Gibco BRL) with 0.5 µg DNA, and 5 µl of [^{32}P] dCTP (3000Ci/mmol, Amersham) according to the instructions of the manufacturer (Gibco BRL). In the nick translation reaction DNase I is used to create single-stranded nicks in double-stranded DNA. The 5'-3' exonuclease activity of *E.coli* DNA polymerase I entered the nicks and removed stretches of single-stranded DNA. The degraded DNA was regenerated by the polymerase activity of the polymerase I, incorporating dATP, dTTP, dGTP, and [^{32}P] labelled dCTP. Unincorporated nucleotides were separated from labelled DNA using a ProbeQuant™ G-50 micro column according to the manufacturer (Pharmacia Biotech). DNA was diluted to a final volume of 500

µl in TE buffer {10 mM Tris-HCl (pH 8.0), 1 mM EDTA}. The counts per minute (cpm) / ml labelled probe was estimated using a scintillation counter (full energy spectrum), and approximately 1 x 10^7 cpm were used for Ca3 Southern blot hybridisation.

2.5.6.4 Southern Blot Hybridisation Using [^{32}P]-Labelled Probes

All hybridisations of probes to DNA fragments immobilised on membranes were performed in roller tubes using a Bachofer hybridisation oven. Firstly, membranes were prehybridised for 2 h at 65°C in 30 ml of 5 x SSPE {5% (w/v) dextrane sulfate, and 0.3 % (w/v) SDS} hybridisation buffer. Hybridisation took place overnight at 65°C using a total of approximately 1 x 10^7 cpm [^{32}P] labelled probe per 30 ml hybridisation buffer. Hybridisation was followed by four stringent washes for 30 min in 2 x SSPE at 45 °C, and the blot exposed to X-ray film (Kodak).

2.5.6.5 Nonradioactive (DIG) Labelling of Probes

Digoxigenin (DIG)-11-dUTP DNA probes were prepared by random primed labelling using the DIG-High Prime kit according to the manufacturer's instructions (Boehringer Mannheim/Roche). First, Ca3 DNA was digested with *Eco*RI (see section 2.5.2), and purified using phenol/chloroform extraction (see section 2.5.3.1), and ethanol precipitation (see section 2.5.3.2). Labelling of Ca3 DNA was performed at 37°C overnight using 1 µg DNA. The yield of DIG-labelled DNA was estimated by comparing serial dilutions of DIG-labelled sample to dilutions of labelled control DNA (5 µg/ml DIG-labelled DNA) supplied by the manufacturer (Boehringer Mannheim/Roche). DNA was spotted on positively charged nylon membrane, cross-linked for 3 min by UV irradiation and detected (see section 2.5.6.7) according to the manufacturer's instructions (Boehringer Mannheim/Roche).

Between 10 and 20 ng of DIG-labelled DNA were used for Southern blot hybridisation. DIG-labelled probe was stored at –20°C for up to 1 year.

2.5.6.6 Southern Blot Hybridisation Using DIG-Labelled Probes

To prevent background problems as a result of the probe being too concentrated, the optimal Ca3 probe concentration was determined by "mock" hybridisation as recommended by the manufacturer (Boehringer Mannheim/Roche). Briefly, small blank (no DNA) pieces of membrane were incubated at different probe concentrations and detection was performed as described below (see section 2.5.6.7). The highest probe concentration that gave acceptable background was 25ng/ml, which was subsequently used for Ca3 Southern blot hybridisations. Hybridisations using P1CTC and actin (see section 2.5.7.1) as probes were performed at a concentration of 10 ng/ml.

In general, membranes were prehybridised in DIG Easy Hyb hybridisation buffer (Boehringer Mannheim/Roche) without probe for at least 1 h at 42°C. After prehybridisation, denatured probe of appropriate concentration was added and hybridisation performed overnight at 45°C. Washes of different stringencies were performed using glass dishes and an incubator (Controlled Environment Incubator Shaker, Model G-25, New Brunswick Scientific Co. Inc., New Brunswick, N.J., USA.): (i) for Ca3 fingerprinting, 4 washes for 20 min in 2 x SSC, 0.1% (w/v) SDS or (ii) 2 washes of 5 min in 2 x SSC, 0.2% (w/v) SDS, followed by 2 washes of 15 min in 0.5 x SSC, 0.2% (w/v) SDS at 45°C or (iii) 68°C. Sites of the hybridised probes were detected as light signals by chemiluminescent detection (see section 2.5.6.7).

2.5.6.7 Detection of DIG-Labelled DNA Probes

Probe binding sites were detected using the chemiluminscent detection protocol as recommended by the manufacturer (Boehringer Mannheim/Roche). First, membranes were equilibrated in washing buffer {0.3% (w/v) Tween 20®, 0.1 M maleic acid, 0.15 M NaCl pH 7.5} for 1 min, followed by a blocking step for 30-60 min in blocking buffer {0.1 M maleic acid (Sigma), 0.15 M NaCl pH 7.5, 1% (v/v) blocking reagent (Boehringer Mannheim/Roche)}, to prevent nonspecific attraction of antibody to the membrane. Subsequently, anti-DIG-F_{AB} fragments conjugated with alkaline phosphatase (Boehringer Mannheim/Roche) were added to a concentration of 1:10,000 to the blocking solution, and incubated for 30 min. Antibody incubation was followed by 2 washes for 15 min in washing buffer. Hybridised probe and bound antibody conjugate were detected by reaction with chemiluminescent substrate CSPD (Boehringer Mannheim/Roche). CSPD-substrate was diluted 1:100 in detection buffer {100 mM Tris-HCl, pH 9.5, 100 mM NaCl} and approximately 0.5 ml per 100 cm^2 membrane was distributed over the membrane. After incubation at 37°C for 20 min, membranes were exposed to X-ray film (Kodak).

2.5.6.8 Stripping of Membranes

To remove the nonradioactive (DIG) labelled probe, the membrane was rinsed in deionised H_2O, then incubated twice for 20 min in 0.2 M NaOH, 0.1% (w/v) SDS at 37°C, and subsequently rinsed in 2 x SSC {0.3 M NaCl, and 0.03 M sodium citrate} before being dried at room temperature and reused for hybridisation.

2.5.7 Polymerase Chain Reaction (PCR)

2.5.7.1 Oligonucleotide Primers

Oligonucleotide primers were designed following the recommendations described by Fenton Wiliams and Perkin-Elmer (1989) and obtained from Life Technologies, unless stated otherwise. Each primer was rehydrated to a final concentration of 1nmol/µl and stored at –80°C. All primers used throughout this study, their orientations, and application, are listed in Table 2-7.

Table 2-7: Oligonucleotides used as primers and adapters in this study

oligonucleotides	sequence 5'→3'	position	application
pITS1 primer	CTTGTGAAACTCCGTCGTGC	*C.albicans* 18S rRNA sequence, →	sequencing of ITS1
pITS2 primer	GGAGAAACGACGCTCAAACAG	*C.albicans* 5.8S rRNA sequence, ←	sequencing of ITS1
pITS3 primer	CAACGGATCTCTTGGTTCTCGC	*C.albicans* 5.8S rRNA sequence, →	RFLPs and sequencing of ITS2
pITS4 primer	GGACGCCAAAGACGCCAG	*C.albicans* 26S rRNA sequence, ←	RFLPs and sequencing of ITS2
M–C primer	GATGAGTCCTGAGTAAC	preselective AFLP primer based on *Mse* adapter sequence +N, → + ←	AFLP analysis
M–CTC primer	GATGAGTCCTGAGTAACTC	selective AFLP primer based on *Mse* adapter sequence +3N, → + ←	AFLP analysis
M–CAG primer	GATGAGTCCTGAGTAACAG	selective AFLP primer based on *Mse* adapter sequence +3N, → + ←	AFLP analysis
M–CAC primer	GATGAGTCCTGAGTAACAC	selective AFLP primer based on *Mse* adapter sequence +3N, → + ←	AFLP analysis
M–CTG primer	GATGAGTCCTGAGTAACTG	selective AFLP primer based on *Mse* adapter sequence +3N, → + ←	AFLP analysis

oligonucleotides	sequence 5'→3'	position	application
M13F primer	CCCAGTCACGACGTTGTAAAAC G	pGEM®-T Easy Vector forward primer, →	sequencing of cloned fragments
M13R primer	AGCGGATAACAATTTCACACAG G	pGEM®-T Easy Vector reverse primer, ←	sequencing of cloned fragments
P1CTCF primer	TCGCTGAGGCCATTTGC	AFLP fragment P1CTC, →	amplification of P1CTC for sequencing, probe for Southern blot hybridisations
P1CTCR primer	GGGATCATAAGCAAGTGCTG	AFLP fragment P1CTC, ←	sequencing of P1CTC
Candida actin sense primer[1]	TGAAGCCCAATCCAAAAGAGGT AT	portion of actin gene, →	standard in multiplex PCR, probe for Southern blot hybridisations
Candida actin anti sense primer[1]	CGAAATCCAAAGCAACGTAACA CAA	portion of actin gene, ←	probe for Southern blot hybridisations
ana (actin antisense) primer	CTTTCTGGTGGAGCAATAATCTT	portion of actin gene, ←	standard in multiplex PCR
ans (actin sense) primer	ATGAAGCCCAATCCAAAAGAGG TAT	portion of actin gene, →	standard in multiplex PCR
CBP1R primer	CCACGATGCCCCTAAAC	*C.albicans* CBP1 (corticosteroid binding protein) gene sequence, ←	PCR amplification of CBP1 upstream region
P1 primer	GGTACTGCATGTAGTTGC	AFLP fragment P1CTC, ←	inverse PCR, AFLP
P3 primer	CTTTCTTACATCCAACAACAACC	AFLP fragment P1CTC, →	inverse PCR, AFLP
Eco primer	GACTGCGTACCAATTC	based on *Eco* adapter sequence, → + ←	AFLP to obtain flanking regions of P1CTC
188F primer	TGTTGTCTTCTCCCATTCTACC	based on sequence of *C.albicans* clone 265153HO1, →	PCR to obtain flanking regions of P1CTC
3'endR primer	CATTAGTGCTATTTGCTCATCAT TG	based on sequence of 3' flanking region of P1CTC, ←	PCR to amplify P1CTC including flanking regions

oligonucleotides	sequence 5'→3'	position	application
Mse adapter	5`-GACGATGAGTCCTGAG-3` 3`- TACTCAGGACTCAT-5`	ligated to restriction fragments	AFLP
Eco adapter	5`-CTCGTAGACTGCGTACC-3` 3`- CTGACGCATGGTTAA-5`	ligated to restriction fragments	inverse PCR

[1] primers were kindly provided by Dr. Peter Farley, Institute of Molecular BioSciences, Massey University

→ , ← relative to Figures presented in this thesis and/or relative to gene orientation

2.5.7.2 PCR Protocols

All PCR reactions were performed in a final volume of 20 µl containing 1 U Taq polymerase (Qiagen), 1x PCR buffer {Tris-HCl, KCl, $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, (pH 8.7)} supplied by the manufacturer (Qiagen), 10 pmol (= 1 Volume, refer to Table 2-8 below) of each primer, 200 µM of each dNTP (Boehringer Mannheim/Roche), and 10-100 ng DNA. PCR protocols are listed in Table 2-8. Some PCR reactions were supplemented by Q-buffer (Quiagen), enhancer solution (Gibco BRL), or dimethyl sulfoxide (DMSO) (Gibco BRL) as described below.

Table 2-8: PCR reaction conditions

primer combination/ ratio	PCR protocol	amplified product(s)
pITS1/pITS2 / 1:1	94°C, 2 min; 34 cycles of 94°C, 1 min; 50°C, 1 min; 72°C, 1 min; final extension 72°C for 5min	ITS1
pITS3/pITS4 / 1:1	94°C, 2 min; 34 cycles of 94°C, 1 min; 50°C, 1 min; 72°C, 1 min; final extension 72°C for 5 min	ITS2
M–C	19 cycles of 1°C/sec to 94°C; 94°C, 30 sec; 1°C/sec to 56°C; 56°C, 1 min; 1°C/sec to 72°C; 72°C, 1 min	preselective amplifications for AFLP

primer combination/ ratio	PCR protocol	amplified product(s)
alternatively *Mse*I-CTC, *Mse*I-CAG, *Mse*I-CAC, *Mse*I-CTG	94°C 2 min,; 5 cycles of 94°C, 30 sec; 65°C, 30 sec; 72°C, 1 min; 5 cycles of 94°C, 30 sec; 60°C, 30 sec; final extension 72°C for 5 min	selective amplifications for AFLP
P1CTCF/P1CTCR / 1:1	94°C ,2 min; 30 cycles of 94°C, 1 min; 55°C, 1 min; 72°C, 1 min; final extension :72°C for 5 min	P1CTC
P1CTCF/CBP1R / 1:1	94°C, 2 min; 34 cycles of 94°C, 1 min; 50°C, 1 min; 72°C, 1 min; final extension 72°C for 5 min	upstream region of CBP1
Candida actin sense/antisense / 1:1	94°C ,1 min,; 30 cycles of 94°C, 10 sec; 55°C, 30 sec; 72°C, 10 sec; final extension 72°C for 1 min	actin as probe for Southern blot hybridisations
ans/ ana and P1CTCF/P1CTCR / 1.2:1.2:1:1	94°C 2 min; 30 cycles of 94°C, 30 sec; 55°C, 60 sec; 72°C, 60 sec; final extension at 72°C for 5 min	multiplex PCR
P1/P3 / 2:1	94°C 2 min,30 cycles of 94°C 1 min; 50°C, 1 min; 72°C 1 min; final extension at 72°C for 5 min	inverse PCR, first amplification
P1/P3 / 2:1	94°C 2 min; 30 cycles of 94°C 1 min; 50°C, 1 min; 72°C 2 min; final extension at 72°C for 5 min	inverse PCR, reamplification of isolated bands
P1CTCF/*Eco* / 5:1	94°C 2 min; 5 cycles of 94°C, 30 sec; 65°C, 30 sec; 72°C, 1 min; 5 cycles of 94°C, 30 sec; 60°C, 30 sec; final extension at 72°C for 5 min	isolation of flanking regions of P1CTC by AFLP
188F/3'endR / 1:1	94°C 2 min; 34 cycles of 94°C 1 min; 54°C, 1 min; 72°C90 sec; final extension at 72°C for 5 min	amplification of P1CTC including flanking regions
as required	25 cycles of 96°C, 10 sec,; 50°C, 5 sec; 60°C, 4 min	autoseq (sequencing reaction)
M13F/M13R / 1:1	1 cycle of 94°C 5 min; 58°C, 1 min; 75°C, 1 min; 23 cycles of 94°C, 1 min; 58°C, 1 min; 75°C, 2 min; 1 cycle 94°C, 1 min; 58°C, 1 min; 75°C, 10 min	screening of *E.coli* colonies transformed with recombinant plasmid DNA

2.5.7.3 Amplified Fragment Length Polymorphism (AFLP)

Genomic *C.albicans* DNA was isolated (see section 2.5.1.1) and 600 µg restricted with 10 U of endonuclease *Mse*I (New England Biolabs) for 4 h at 37°C

(see section 2.5.2). Restriction fragments were then separated electrophoretically on a 0.8 % agarose gel (see section 2.5.5.1) to ensure complete digestion. Bands were visualised by UV transillumination after ethidium bromide staining. An aliquot (approximately 150 ng) was ligated (see section 2.5.8.1) to *Mse*I adapter (Life Technologies, see Table 2-7) and a subset of tagged restriction fragments was preamplified (see section 2.5.7). All PCR amplifications were performed in a volume of 20 µl using *Taq* DNA polymerase and buffer supplied by the manufacturer (Qiagen). The PCR products of preamplification were diluted 1:10 and 1 µl was used for selective amplification (see section 2.5.7). Following amplification, 3.5 µl of products were electrophoretically resolved on 5% (w/v) denaturing polyacrylamide (see section 2.5.5.2). Polymorphisms were visualised by silver staining according to the DNA Silver Staining System Manual (Promega) (see section 2.5.5.2).

The relationship between isolates was visualised by reconstructing phylogenetic trees (see section 2.6.1).

2.5.8 Cloning Procedures

2.5.8.1 Ligations

Ligations of PCR products with an A overhang at the 3'end (added by Taq polymerase) were performed using the pGEM®-T Easy Vector System according to the instructions of the manufacturer (Promega). Prior to ligation, PCR products were purified using the QIAquick PCR purification kit (Qiagen) (see section 2.5.3.3) or by gel extraction (see section 2.5.3.4). Ligations were performed in 10 µl volumes using 1 x ligation buffer, 3 Weiss units T4DNA Ligase, and 50 ng pGEM®-T Easy Vector supplied by the manufacturer (Promega). The molar ratio of insert:vector was 3:1 and the amount of PCR product needed for ligation reaction was calculated by the following equation:

$$\frac{\text{ng of vector} \times \text{kb of insert}}{\text{3.0 kb vector}} \times \text{insert:vector molar ratio} = \text{ng of insert}$$

as recommended by the manufacturer (Promega).
Ligations were performed overnight at 4°C.

2.5.8.2 Transformations

High efficiency competent *E.coli* DH5α cells were purchased from Gibco BRL and stored at –80°C. Transformations were performed as follows: an aliquot of 30 µl competent cells was thawed on ice for 10-30 min, mixed with the whole ligation reaction, heat-shocked at 42°C for 45 sec, then put on ice for 2 min. The suspension was diluted 10-fold in LB medium (see section 2.2), and incubated at 37°C on a shaker at 200 rpm for 1 h. The whole cell suspension was plated on selective media plates (LB, Amp, Xgal, IPTG, see section 2.2), and incubated at 37°C overnight. White colonies most likely contained the transformed fragment and were therefore selected for further analysis (see section 2.5.1.2). Positive and negative transformation controls were performed regularly as recommended by the manufacturer (Promega).

2.5.8.3 Sequencing of DNA Fragments

Templates were purified using the QIAquick PCR Purification Kit (see section 2.5.3.3). Sequencing reactions were performed based on the technique described by Sanger *et al.* (Sanger *et al.*, 1974) using the autoseq protocol (see PCR protocols, section 2.5.7.2) with the recommended quantities of 1.6 µl primer (0.8 pmol/µl), 2-3 µl template (PCR product, 20-45 ng/µl or plasmid DNA, 200ng/µl), and 3.5 µl of Dye Terminator Sequencing Mix (Perkin-Elmer Corporation) in a final volume of 10 µl. Labelled DNA was precipitated for 5 min on ice in the presence of absolute ethanol (see section 2.5.3.2), centrifuged for 15

min at 14 000 rpm, washed with 70% (v/v) ethanol, centrifuged for 15 min at 14 000 rpm, and the pellet dried at RT or 37°C. The reactions were resuspended and run by the MUSeq Massey University DNA Analysis Service (Massey University, Palmerston North, New Zealand) on an ABI PRISM™ DNA sequencer, model 377 (Perkin-Elmer Corporation). Generated data were visualised as chromatograms using ABI PRISM software (Perkin-Elmer Corporation).

2.5.8.4 Analysis of DNA Sequences

Editing of DNA sequences was performed using the GeneJockey software program (Biosoft). For comparative sequence analysis, multiple sequence alignments were produced using the software program ClustalX (1.62b) with the default parameters (Higgins and Sharp, 1989). To characterise sequences and to identify homologous sequences, sequence data were compared to the databases GenBank, Stanford (unpublished sequences of *Candida albicans* sequencing project), and European Molecular Biology Laboratory using the BLAST algorithm (Altschul *et al.*, 1990) using the default parameters.

2.6 Phylogenetic Analysis

The aim of most phylogenetic analyses is to reconstruct evolutionary relationships. Molecular phylogenetics is typically based on information that is derived from DNA (sequence differences or indications of such, e.g. variable patterns in fingerprint profiles). These are coded as site patterns and/or converted into distances, then typically represented as "evolutionary" or "phylogenetic trees" (Page and Holmes, 1998).

In the present study phylogenetic trees were reconstructed to test whether an HSP cluster (see section 1.4 for more details) was a feature of both Ca3 and AFLP fingerprinting data. Tree building methods used in the present study were

those implemented in the computer software packages Dendron (Schmid *et al.*, 1990), Paup 4.0 (Swofford, 1998), and SplitsTree (Bandelt and Dress, 1992).

2.6.1 Tree Building Methods

As suggested from Table 2-9, tree building methods can be divided into dissimilarity (the general term for "distance") and discrete methods. Both approaches can involve either local clustering or global search methods. In the latter case these can be based on finding an optimal criteria for all the taxa considered together or for subsets of taxa. The first classification (distance vs discrete) is based on the kind of data which is used for the reconstruction of trees, while the second classification is based on the mathematical criteria used to select between possible trees. An overview is given in Table 2-9.

Table 2-9: Classification of tree building methods (adapted from Page and Holmes, 1998)

tree selection criteria	type of data: clustering	distance	discrete
	clustering	distance UPGMA NJ	discrete
	global search	minimum evolution	maximum parsimony
			maximum likelihood
		split decomposition	split decomposition
		quartet puzzling	quartet puzzling

(i) Dissimilarity and similarity methods

Distance methods are based on the idea that the evolutionary relationships between taxa can be reconstructed from estimates of evolutionary distance for all pairs of taxa calculated either from their homologous DNA sequences or from other data, such as AFLP profiles, Ca3 profiles etc. (Page and Holmes, 1998). With such methods, the actual data set, i.e. Ca3 and AFLP scoring data, is converted into a pairwise distance matrix, which is held to represent the evolutionary distance of one taxon from another. Several approaches (using either "cluster" or "global search" criteria) can then be employed for tree reconstruction (see below). Examples of distance methods which use clustering criteria to build trees are UPGMA and neighbour-joining (NJ). An example of a distance method which uses global search criteria is minimum evolution.

In contrast to the use of dissimilarities or distance estimates for building trees, early tree building methods used "similarity" values. In these, "S_{AB}" values are held to represent the degree of similarity between taxa. Here, for the purpose of making the present study directly comparable with earlier work, they are calculated for Ca3 fingerprints. Values range from 1.0 (two patterns are identical) to 0.0 (the two patterns have no bands in common). S_{AB} values between two strains, A and B, are calculated using the following formula

$$S_{AB} = \frac{\sum_{i=1}^{k}(a_i + b_i - |a_i - b_i|)}{\sum_{i=1}^{k}(a_i + b_i)}$$

where a_i and b_i are the intensities of band i in patterns A and B, respectively, and k is the number of bands. S_{AB} values for all isolates analysed are comprised in a pairwise distance matrix.

(ii) Discrete methods

In contrast to distance methods, discrete methods operate directly on sequences, or on functions of it (e.g. hybridisation patterns), rather than on pairwise distances. Examples of this approach include maximum parsimony and maximum likelihood.

2.6.1.1 Maximum Parsimony and Maximum Likelihood Tree Building

For maximum parsimony (minimal length), only "parsimony informative sites", that is, sites which discriminate at least two taxa from at least two other remaining taxa, are used to choose the tree (or trees) that require the fewest evolutionary changes. Branch lengths are not considered in the calculation of maximum parsimony trees. Nor are patterns in the data that correspond to the external branches of the tree. In contrast, maximum likelihood chooses the tree with branch lengths that are most likely to have evolved the observed data (Page and Holmes, 1998).

Maximum parsimony and maximum likelihood trees are often very similar, because both are global search methods (Page and Holmes, 1998). However, because of this and their need to evaluate every possible tree that might fit the data, they are often computationally too extensive to use when analyses involve many taxa. For the analysis of more taxa, subsets of the whole dataset can be selected and analysed (e.g., as in quartet methods such as split decomposition or quartet puzzling). Alternatively heuristic global searches which cannot guarantee to find the best tree, but which do seek a global solution to phylogeny, e.g. quartet puzzling (Strimmer and von Haeseler, 1996) can be used for larger numbers of taxa or subset comparisons.

2.6.1.2 UPGMA and Neighbour-joining (NJ) Clustering

(i) UPGMA

UPGMA (Sneath and Sokal, 1973) is an algorithm which is typically used to cluster similarity or dissimilarity values in a data matrix. The method first joins the two least dissimilar taxa, then it joins an additional taxon which is closest to the midpoint of these. This process is continued until all taxa are joined. The data is displayed as a tree (dendrogram), in which all tips are equally distant from the root of the tree (i.e. the analysis assumes that the data evolved in a clock-like fashion). Branch lengths suggest times of divergence but this interpretation requires the assumption that all evolutionary changes in the data have occurred at the same rate (an assumption which is rarely true even for sequence data). A disadvantage of the method is that the resulting tree can be highly dependent on the order in which the data has been considered in the analysis (the reason being that when distances between taxa are the same length, the method arbitrarily joins into the growing tree, the first taxon of equal distance that it encounters). Nevertheless, ease of implementation of the method has made it very popular.

(ii) Neighbour-joining

Neighbour-joining (NJ) (Saitou and Nei, 1987) is also a clustering method. Unlike UPGMA it takes into account the net divergence of each taxa from every other taxon in the data matrix. It therefore calculates the shape of a tree without needing to assume a molecular clock. Like UPGMA it is also susceptible to the problem of the order of taxon addition. In the resulting tree, branch lengths correspond to the amount of evolutionary change between taxa (Page and Holmes, 1998).

2.6.1.3 Quartet Methods

Some tree building methods are more difficult to classify than others. Such methods include "quartet puzzling" (Strimmer and von Haeseler, 1996; Strimmer, Goldman *et al.*, 1997) and "split decomposition" (Bandelt and Dress, 1992; Huson, 1998). Both are used in this thesis. As will be discussed, split decomposition provides a useful tool for examining data structure, while quartet puzzling is a useful method for building trees for population samples (large numbers of taxa). Both methods employ a common principle, which can be used for both distance and discrete character methods: taxa are analysed in subsets of four (quartets) and then these subsets are used to build a tree or network describing the relationships between all taxa in the data set. As described below, split decomposition and quartet puzzling use the quartet information differently to build graphs.

(i) Quartet puzzling

This method first examines all combinations of four taxa and determines which (if any) of the three bifurcating trees is best supported for each quartet. In doing this it could use e.g. parsimony, maximum likelihood or distance criteria. A global heuristic search strategy is used to join all such subtrees into a final tree (so called "puzzling steps"). The method is very fast and simulation studies have suggested that the heuristic search works well to find optimal trees.

(ii) Split decomposition

In the first step, all possible quartets are searched for weakly compatible splits (the two best supported splits in each quartet) using distance or discrete methods. In the present study, a distance criterion was used to calculate the

support for different splits. These values are called "isolation indicies" (Lockhart *et al.*, 1999).

The second step is to obtain a "split system" which reflects the relationship between all taxa. Since a split system contains all taxa, it is possible that some trees which were not excluded in the previous quartet analysis will identify the same split with different isolation indices. In this case the smallest index of each compatible split in the split system is chosen as the one to represent the value of this split (Lockhart *et al.*, 1999).

In the final step, the split system is displayed in the form of a multidimensional graph called a splitsgraph. The splitsgraph will have a treelike structure when no incompatibilities are present in the split system. In this case there will be a good correspondence with isolation index values in the split system and in the graph. However, if there are incompatibilities in the split system, these are indicated by boxes in the splitsgraph, and poor correspondence between split values in the split system and in the splitsgraph (Lockhart *et al.,* 1999).

2.6.1.4 Measurement of Fit Between Data and Trees

(i) Discrete data

Numerous methods have been used to compare the fit of data onto different evolutionary trees. Different approaches have been discussed in Swofford *et al.*, (Swofford *et al.*, 1996) and by Page and Holmes (Page and Holmes, 1998).

One of the most commonly used approaches involves "bootstrapping" of data matrices (Swofford *et al.*, 1996). In this method, data sets are resampled with replacement to generate new data sets and trees are reconstructed for each of these. A consensus tree is then calculated and the number of times taxa group

together in such trees is reported. When bootstrap values are high, this indicates a relatively large number of uncontradicted patterns in the original data matrix (as interpreted by the tree building method) supporting a relationship. When values are low, there may be either a low number of patterns in the data set (which may be uncontradicted) in comparison to other pattern types or there may be a relatively similar number of patterns but which suggest contradictory relationships. Bootstrapping has been used in the present study.

With quartet puzzling "reliability" values (Strimmer and von Hasesler, 1996) are given for the internal branches in the reconstructed tree. These values indicate the number of times, in the independent heuristic puzzling steps that certain relationships were obtained. Often these values will be expected to be similar to bootstrap values. However, if there are relatively few informative patterns in a data set, but they are largely compatible, the puzzle values tend to be higher than the boostrap values. They indicate the extent of internal compatibility in a data set.

(ii) Distance data

Observed distances, calculated from pairwise comparisons, rarely perfectly fit the evolutionary distances visualised in reconstructed trees. This occurs because the true number of evolutionary changes that lead to the observed data is difficult to estimate. In practise, the true amount of genetic change between taxa may be greater or less than that contained in the distance matrix and/or represented on the tree. In selecting the evolutionary tree that best describes the evolution of distance data, a fit statistic can be used. Measures of best fit can be used to identify the tree that minimises the difference between the distances in the distance matrix and the distances represented in the tree. A fit statistic is associated with the splitsgraphs reconstructed in the present work (the statistic is

calculated as the sum of all the distances in the graph divided by the sum of pairwise distances between taxa).

2.6.1.5 Tree Comparison Metrics

A common procedure used for comparing the degree of similarity of two evolutionary trees is to count the number of internal branches that are unique to both trees and to then compare this number to that expected from a Poisson distribution; this is the tree comparison metric (Steel and Penny, 1993; Page and Holmes, 1998). This approach is used here for comparing the similarity of quartet puzzle trees derived from Ca3 and AFLP profiles.

2.6.2 z-test

To test whether a difference between two proportions, e.g. the frequencies of S_{AB} values within different study groups, was statistically significant, the z-test was applied using the following equation:

Null hypothesis: $p_1 = p_2$

Alternative hypothesis: $p_1 \neq p_2$

$$z = \frac{|p_1 - p_2|}{\sqrt{p(1-p)[(1/n_1) + (1/n_2)]}}$$

with n_1 = no. tested, n_2 = no. tested, p_1 = no. of events, p_2 = no. of events

If t was less than t_{α} in the table of critical value (Sachs, 1991), at 99 % confidence the null hypothesis was accepted, if t was greater than t_{α} the null hypothesis was declined.

CHAPTER 3 - CA3 TYPING AND THE STUDY OF *C.ALBICANS* EPIDEMIOLOGY

3.1 Introduction

Despite the prevalence of *C.albicans* as a commensal organism and the frequency of Candidosis, some of the most fundamental questions related to commensal carriage, infection, or the transition from the commensal to the pathogenic state have not been answered (Soll, 1992; Soll, 1993). Epidemiological studies have focussed on colonisation and predisposition of a host to the disease. At present it is not clear whether infection-causing strains derive from the endogenous flora or from an exogenous source. However, recent findings suggest the prevalence of a closely related group of *C. albicans* strains in individuals suffering from Candidosis (Schmid *et al*, 1993; 1999). An hypothesis is that these strains replace the commensal flora and cause Candidosis. In this chapter, the use of Ca3 fingerprint methodology to test this hypothesis in individuals which are expected to have different levels of predisposition to Candidosis is reported. This methodology was also used to identify potential sources of nosocomial infections and results from this study are reported and discussed here. Background information on the nature of colonisation, predisposition and nosocomial infections are first described.

3.1.1 Colonisation by *Candida* Subspecies

"Colonisation is the presence of a microorganism in or on a host with growth and multiplication but without any overt clinical expression or detected immune response in the host at the time the microorganism is isolated" (Brachman, 1986). Normal oral colonisation in humans begins during the birth process and through subsequent contacts with inanimate or animate environments

until a sensitively balanced mircroflora is established (Jarvis, 1996). *Candida* often colonises the human epidermis and the gastro-intestinal tract. A large proportion of the population carries *Candida* subspecies in the mouth, with *C.albicans* as the most prevalent species. A study by Russell and Lay (1973) showed that the frequency of oral carriage at birth is low (5.7% neonates), more than doubles by the time of hospital discharge (14.2%, usually 7 days after birth) and sharply increases after 1 month at home (82%). A different study suggested that the frequency decreases to 50% at an age of 1 year (Kleinegger *et al.*, 1996).

The isolation of *C.albicans* or other *Candida* subsp. from the commensal flora of the oral cavity without lesions does not indicate a clinical Candidosis (Odds, 1988; Cannon *et al.*, 1995). The term 'commensal isolates' implies nonpathogenic isolates as part of the commensal microflora. However, since roughly one third to one-half of adult individuals do not carry measurable levels of *Candida* species in the oral cavity, it may be possible that carriage itself predisposes an individual to subsequent infection (Odds, 1988; Soll, 1993).

3.1.2 Predisposition to Candidosis

3.1.2.1 Host Defence

Any alteration in antimicrobial defences can reduce the ability of a host to withstand a microbial attack. For most infectious diseases, the seriousness of the illness and its eventual outcome are determined by the balance between the pathogenicity of the organism and the strength of the host defences (Bodey, 1993).

Table 3-1: Overview of host defence mechanisms against *Candida* subsp. and predisposing factors

nonimmune host defence		**immune host defence**	
microbial flora competition for nutrients production of antifungal substances	**skin and mucosal barrier** high turn over of cells eliminate attached organism skin lipids can inhibit growth	**cell- mediated immunity (CMI)** antigenpresenting cell (macrophage) and T-lymphocyte complex activates inhibition or killing of *Candida* subsp.	**phagocytosis** neutrophils (white blood cells) are able to digest *Candida* subsp.
predisposing factors			
antibiotics	burns surgical wounds catheters	defects in CMI, e.g. - AIDS patients	neutropenia (low neutrophil count), e.g., - haematological malignancies - other malignancies treatment with chemotherapy

The most prominent examples of predisposing factors are (Odds, 1988):

- diseases, e.g. immunopathological diseases and malignant diseases,
- digression from normal physiological status, e.g., infancy, old age, pregnancy, diets,
- mechanical factors, burns/trauma and local disruption, the preference of the *C.albicans* for moist habitats, means that any situation involving disruption of skin or membrane or which increases the local humidity may predispose for overgrowth and infection
- iatrogenic factors (associated with medical or surgical treatment), e.g. antibiotics, corticosteroids, hormonal contraceptives, drugs that induce neutropenia, surgery, intravascular catheter.

Often minor defects in host defences will allow *Candida* subsp. to invade and cause illness (Bodey, 1993). A wide variety of host defence mechanisms are thought to play an important role in preventing infections with *Candida subsp.* Table 3-1 gives an overview of the most dominant mechanisms. The commensal microflora and skin and mucosal barriers are considered to be parts of the nonimmunal host defence. The skin is frequently colonised by *Candida* subsp., but invasion of this tissue occurs only under special circumstances. An increased prevalence of local and systemic Candidosis occurs when these barriers are physically disrupted by trauma such as burns, surgical wounds, and the placement of intravenous catheters and other types of catheter (Odds, 1988; Bross *et al.*, 1989; Levitz, 1992).

The normal endogenous bacterial flora protects humans by competing for nutrients and, perhaps, by local production of antifungal substances. Therefore the administration of broad spectrum antibiotics increases the risk of Candidosis by disrupting the normal flora (Odds, 1988; Bross *et al.*, 1989; Levitz, 1992).

Phagocytic defence and cell-mediated immunity (CMI) are part of the immune defence against *Candida* subsp.. Phagocytes include the granulocytic phagocytes (neutrophils, eosinophils and basophils) and the mononuclear phagocytes (monocytes and macrophages). Neutrophils and mononuclear phagocytes are able to kill *Candida* hyphae and spores (Levitz, 1992; Bodey, 1993). Patients who are neutropenic, usually because they have received cytotoxic chemotherapy for haematological malignancies, are particularly predisposed to invasive Candidosis (DeGregorio *et al.*, 1982; Tomada and Nakano, 1984; Odds *et al.*, 1989; Reagan *et al.*, 1990; Bodey, 1993). The CMI response to a fungal cell is known to require an antigen-presenting cell, in general a mononuclear phagocyte, to process and present fungal antigens to T-lymphocytes (generally CD4+ helper cells). Specific lymphocytes that express receptors for the antigens are then generated, and these lymphocytes proliferate and release cytokines, which activate effector cells to inhibit or kill the fungus. Cells that are postulated to function as the effector arm of the CMI response include macrophages, natural killer cells, and cytotoxic T-lymphocytes (Levitz, 1992).

3.1.2.2 Haematological Cancer and the Impact on Host Defence

Haematological cancer patients are expected to be the most promising group of patients for strain replacement studies, since they are at a very high risk for acquiring Candidosis. Haematological malignancies can be distinguished as leukaemias and tumours in the lymphoid tissues.

Leukaemia occurs in a number of forms. The two main criteria used in the classification are the clinical course of the disease, and the differentiation of the predominant leukaemia cell population. The myeloid leukaemias involve unregulated proliferation of primitive haematopoietic cells in which maturation, and therefore, functional capability is impaired. The defect in maturation is quantitative,

being most severe in poorly differentiated (acute) forms and less severe in well differentiated (chronic) forms. In acute myeloid leukaemia (AML) neutropenia may lead to infections, whereas in chronic myeloid leukaemia (CML) only evidence for mild functional impairment of phagocytosis exists (Lichtmann, 1980). Lymphoid leukaemias are malignancies affecting the lymphoid system. In both forms, acute lymphoid leukaemia (ALL) and chronic lymphoid leukaemia (CLL), phagocytosis can be extremely impaired.

Tumours in the lymphoid tissue comprise solid tissue tumours, malignant lymphomas, and tumours of plasma cells which are generally categorised as multiple myeloma. Lymphoma is a family of diseases in which the primary expression of the malignant process takes place in the tissues, usually in the lymph nodes, spleen, and other lymphoid organs. The immune defects make these patients susceptible to fungal infection. Multiple Myeloma is a neoplastic monoclonal expansion of plasma cells. Growth of malignant plasma cells throughout the bone marrow of multiple bones causes pancytopenia. It is expected that in an early stage the bone marrow failure is minor and consequential neutropenia might not have been developed.

3.1.2.3 Nonhaematological Cancer and Impact on Host Defence

In this study the term "nonhaematological (solid) cancer" comprises all diseases that are not associated with a malignant transformation of normal haemopoietic tissue (haematological cancer). At an early stage, nonhaematological solid cancer (before chemotherapy treatment) patients are thought to have a normal host defence against *C.albicans* which subsequently becomes compromised (Ortiz and Stoliar, 1988; Clarke, 1995; Dollinger and Rosenbaum, 1998). Thus, such patients would also be expected to be useful in strain replacement study.

3.1.3 Nosocomial *Candida* Infections

Between 1980 and 1990 *Candida* infections accounted for 78.3% of nosocomial (hospital obtained) fungal infections (Fridkin and Jarvis, 1996). Candidosis remains the most frequently encountered fungal infection in cancer (Anaissie, 1992; Meunier *et al.*, 1992), where the frequency of fungal infections is higher among patients with haematological malignancies than among patients with solid tumours (Bodey, 1984). The proportion of hospital diagnoses in which leukaemia or lymphoma was associated with Candidosis rose from 7.3% in 1976 to 19.7% in 1980-1982 in US hospitals (Rheingold *et al.*, 1986).

C.albicans is the most frequent cause of fungemia (candidemia) and accounts for 8%-10% of all hospital-acquired bloodstream infections (Pfaller, 1994). The ESR Health (Environmental Science Research, New Zealand) reported a five-fold increase of fungemia for New Zealand during 1988 – 1994 (Biannual summary of opportunistic mycoses infections in New Zealand, ESR). The estimated overall crude mortality rate is 50-85% (Bodey, 1984; Bross *et al.*, 1989; Wey *et al.*, 1989) with about one third of infected patients dieing as a result of candidemia, one third as a result of their underlying disease, and another third surviving hospitalisation and their infection (Bross *et al.*, 1989; Wey *et al.*, 1989).

Nosocomial fungemia is a major complication, even in patients without haematological malignancies (Klein and Watanakunakorn, 1979; Horn *et al.*, 1985; Meunier *et al.*, 1992). The incidence of all types of *Candida* infections has risen during the eras of antibiotic and immunosuppressive chemotherapy (Odds, 1988). More advanced technologies and therapies, such as bone marrow or solid-organ transplants and chemotherapy agents, have become common at many medical centres, resulting in many immunocompromised individuals (Fridkin and Jarvis, 1996). The increase in numbers of severely ill, immunocompromised, hospitalised patients who are highly susceptible to fungal infections may account for the rising

number of nosocomial infections. (Fridkin and Jarvis, 1996). However, the true rise of incidence is probably smaller than is indicated by some authors. The factors such as increased consciousness of Candidosis and improved methods for detection of *Candida* certainly contribute to the elevated frequency of diagnosed Candidosis (Odds, 1988).

Investigations of nosocomial acquisition must take into account the possibility of endogenous vs. exogenous reservoirs for infection (Pfaller, 1996)., and it is often suggested that many *Candida* infections arise from an endogenous source (Seeliger and Patzelt, 1991; Pechere, 1993; Sternberg, 1995; Voss *et al.*, 1995). This frequency is thought to be influenced by the patient population, the various treatment regimes, and the administration of antibiotics or other supportive care (Pfaller, 1996). The gastrointestinal tract in particular must be considered an endogenous source of *Candida* species (Reagan *et al.*, 1990; Wingard, 1995).

In addition to endogenous sources, *Candida* species are frequently found in the hospital environment, i.e. in food (lemon juice) (Berger *et al.*, 1988), in the air (Davidson and Mould, 1978), on floors and other surfaces of the hospital wards (Maki, 1982; Sanchez *et al.*, 1993; Vazquez *et al.*, 1993; Rangel-Frausto *et al.*, 1994). Cases of isolation of *Candida* species from dummies and droppers (Phleps *et al.*, 1986), retrograde syringe fluids (Sherertz *et al.*, 1992), and blood pressure transducers (Weems *et al.*, 1987) have also been documented.

C.albicans survives in hand creams (France, 1968) and other oil-in-water emulsions (Brown *et al.*, 1986), as well as in eye-cosmetics (Wilson *et al.*, 1971). It is also able to persist on cotton and nylon fabric, and therefore on fomites such as underclothing (Rashid *et al.*, 1984). However, it survives poorly on dry surfaces such as skin or glass (Kashbur *et al.*, 1980). One major source for the transmission of *Candida* (Bauer *et al.*, 1990) is considered to be hands of health

care workers (Casewell and Phillips, 1977; Phleps *et al.*, 1986; Sherertz *et al.*, 1992; Vazquez *et al.*, 1993; Strausbaugh *et al.*, 1994).

The evidence of exogenous acquisition of *Candida* species continues to increase (Isenberg, 1989; Doebbeling *et al.*, 1991; Pfaller, 1995). Thus it seems reasonable to predict that transmission of *C.albicans* from one individual to another via dry fomites is also likely as a route of transmission . However, it should be noted that many of these epidemiological studies suffered from problems regarding the discriminatory power and reproducibility of the typing methods used (Hunter, 1991; Schmid *et al.*, 1995). These problems are exemplified by the results of multiple studies of the so-called "London outbreak". Repeated analyses of this "outbreak" resulted in contradictory data not only when different typing methods were used, but also when the same method was used twice for the same isolates (Hunter, 1991; Schmid *et al.*, 1995). In contrast, the success of Ca3 fingerprinting in recent typing studies of *C.albicans* strains prompted the development of a nonradioactive Ca3 fingerprint protocol and the use of it for pilot studies investigating strain replacement and also for investigating nosocomial sources of infection.

3.2 Nonradioactive Ca3 Fingerprinting

A nonradioactive approach was sought for helping to visualize Ca3 DNA fingerprint profiles. The following describes radioactive and nonradioactive detection using different transfer and hybridisation conditions. The comparison between nonradioactive and radioactive Ca3 fingerprinting has been published on the CD Non-Radioactive Labeling and Detection (Roche).

Two samples of *Eco*RI digested DNA from three different clinical *C.albicans* isolates (see Figure 3-1, 1-3) were separated electrophoretically on the same 0.8% agarose gel. After depurination in 0.25 M HCl for 10 min, denaturation (twice

for 15 min in 1.5 M NaCL/0.5 M NaOH), and neutralisation (twice for 15 min in 0.5 M Tris-HCl, pH 7.5; 3 M NaCl), the gel was cut in half (half A: 1-3 and half B: 1-3) and each half was hybridised under different conditions. Southern blot hybridisation of half A was performed under conditions as described previously for radiolabelled Ca3 fingerprinting (see sections 2.5.6.3 and 2.5.6.4) and compared to Southern blot hybridisation of half B, which was performed under modified standard conditions as recommended by the DIG System User's Guide for Filter Hybridisation (Boehringer Mannheim/Roche).

Briefly, DNA from each half (A+B) was transferred onto positively charged nylon membranes (Boehringer Mannheim/Roche) using different transfer buffers. Transfer buffer for blot A contained 10 x SSPE (see section 2.5.6.1), while transfer buffer for blot B consisted of 20 x SSC (see section 2.5.6.1). Hybridisation was performed overnight at 65°C using 5 x SSPE {5% dextrane sulfate, and 0.3% SDS} hybridisation buffer for blot A, and 5 x SSC {0.1% N-laurosylarcosine (w/v), 0.02% SDS, pH 7.0} hybridisation buffer for blot B. Both blots were prehybridised for 2 h at 65°C with 30 ml of their respective hybridisation buffer supplemented with 3 µg calf thymus DNA as blocking reagent. Hybridisation took place using approximately 15 ng of DIG labelled Ca3 DNA per ml hybridisation buffer (see section 2.5.6.6). Four stringency washes were performed at 45°C for 30 min each using 2 x SSPE, 0.2% SDS for blot A, and 2 x SSC + 0.2% SDS for blot B.

Detection was performed as described earlier (see section 2.5.6.7). As shown in Figure 3-1, the SSPE hybridisation/transfer buffer system (Blot A) lead to strong background. Less background was detected using the SSC buffer system, but signal intensity was not satisfactory.

For the evaluation of optimal hybridisation conditions further experiments using the previously described SSC hybridisation buffer with an increased probe concentration (20-40 ng/ml) and supplementation of various blocking buffer

concentrations (0.1-0.5%) were performed (data not shown). Blots resulting from these experiments showed dramatically lowered signal intenisty or high background (data not shown). Optimal results (see Figure 3-2) were gained when using DIG Easy Hyb hybridisation buffer (Boehringer Mannheim/Roche) under conditions as descibed earlier (see section 2.5.6.5).

3.2.1 Comparison of Nonradioactive (DIG) Labelled and Radioactive [^{32}P]-Labelled Probe in Ca3 Fingerprinting

Southern blot hybridisations (see section 2.5.6.4 and 2.5.6.6) of 2 standard strains (3153A and AU52) using radioactive labelled (see section 2.5.6.3) and nonradioactive labelled (see section 2.5.6.5) Ca3 probe were compared.

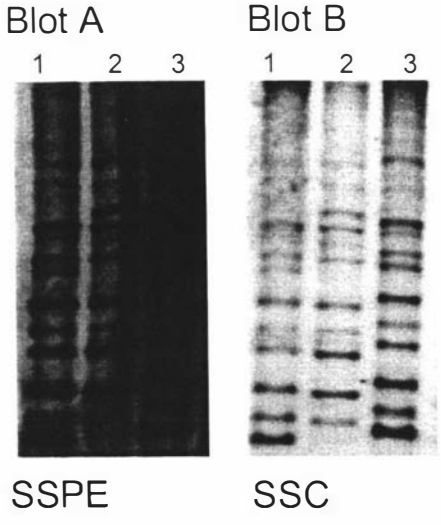


Figure 3-1: Comparison of transfer/hybridisation buffer systems to establish Ca3 Southern blot hybridisation using nonradioactive labelled (DIG labelled) probe. Blot A: transfer/hybridisation conditions as described for radiolabelled Ca3 fingerprinting protocol. Blot B: modified standard conditions as described by the manufacturer (Roche/Boehringer Mannheim).

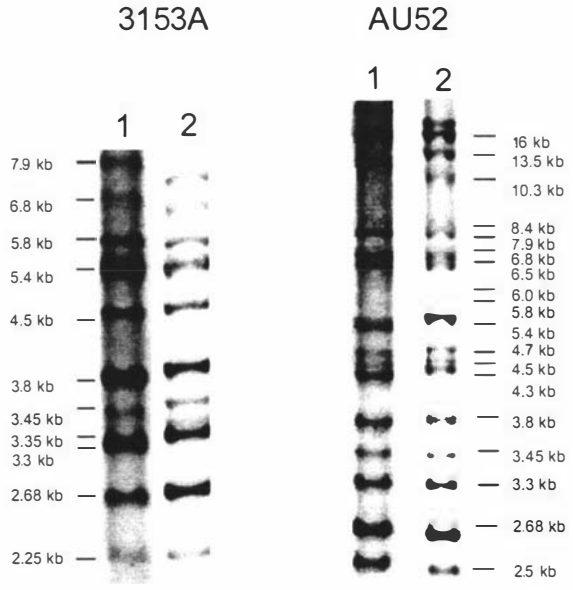


Figure 3-2: Southern blot hybridisation patterns (Ca3 fingerprints) of *C.albicans* standard strains 3153A and AU52 generated with [^{32}P] labelled (1) and DIG labelled (2) Ca3 probe. Both methods resulted in equivalent signal intensities. The background of radiolabelled Southern blots (1) was higher and bands were not as sharp as on nonradioactive labelled blots.

3.2.2 Discussion

Previous studies comparing Southern blot analysis of genomic DNA using radioactive and nonradioactive labelled probes have shown that both methods give equivalent banding profiles (Neuhaus-Url, 1993). To test whether this was also true for Ca3 fingerprinting a comparison between hybridisation patterns of two *C.albicans* reference strains (see Figure 3-2) using digoxigenin (DIG) labelled and

[^{32}P] labelled probe was made. Equivalent banding patterns were obtained for both methods. The nonradioactive approach produced lower background and sharper bands. Handling conditions for nonradioactive probe preparation and hybridisation were also more convenient than for radioactive experiments, making the nonradioactive detection preferable.

The Ca3 methodology was used to investigate the genetic diversity of strains harboured within and between individuals. The data obtained using this approach was analysed using split decomposition (Bandelt and Dress, 1992), an analytical approach previously used in viral studies (Dopazo *et al.*, 1993; Holmes *et al.*, 1999).

3.3 Range of Genetic Variability of Commensal *C.albicans* Strains Within and Between Individuals

3.3.1 Genetic Variability of C.albicans Isolates Within an Healthy Individual

Ca3 fingerprinting was used to investigate the genetic variability of *C.albicans* strains found in an healthy individual from Palmerston North, New Zealand (see section 2.4.1 for explanations of criteria used to select individuals). To investigate the extent of diversity of strains harboured at a single site nine *C.albicans* clones were isolated from a saliva sample. The genetic similarity between theses isolates was expressed in S_{AB} values (see section 2.6.1). The values appeared bimodally distributed. One peak was found at S_{AB} values between 0.61-0.70 and a second peak between 0.91-1.0, as shown in Figure 3-3. Of all calculated S_{AB} values 57% were below 0.8 and 43% above 0.8 with an average S_{AB} value of 0.78.

Analysed under split decomposition (see Figure 3-4a; and section 2.6.1.3 for description of this method), the same discrete data used for S_{AB} calculations

also showed two clusters. The histogram plot and splitsgraphs have been interpreted as suggesting the presence of two distinct strains and their sub (or quasi) strains (Eigen, 1993). The average genetic distance between these strains was calculated using the method of Steel *et al.*, 1996, which estimates the genetic dissimilarity between two clusters (Steel *et al.*, 1996). This distance was 0.27+/-0.05. The average pairwise genetic distance within both clusters was calculated and these were found to be 0.06+/-0.03 for cluster A and 0.13+/-0.06 for cluster B (see Figure 3-4a). The average pairwise genetic distance between all isolates from individuals A.26 was 0.2+/-0.09.

An interesting observation was that the treelike quality of the Ca3 fingerprinting data under split decomposition greatly improved when band intensity was excluded as a character in the original data matrix (compare Figures 3-4a and 3-4b). In this case was the average pairwise genetic distance within cluster A 0.05+/-0.03 and 0.06+/-0.05 within cluster B. Between all isolates was this diversity 0.16+/-0.09. The genetic distance between cluster A and B was 0.24+/-0.05 when band intensity was omitted. In all cases the fit statistic on the splitsgraphs was > 84% indicating that these networks provided a good representation of the observed genetic diversity.

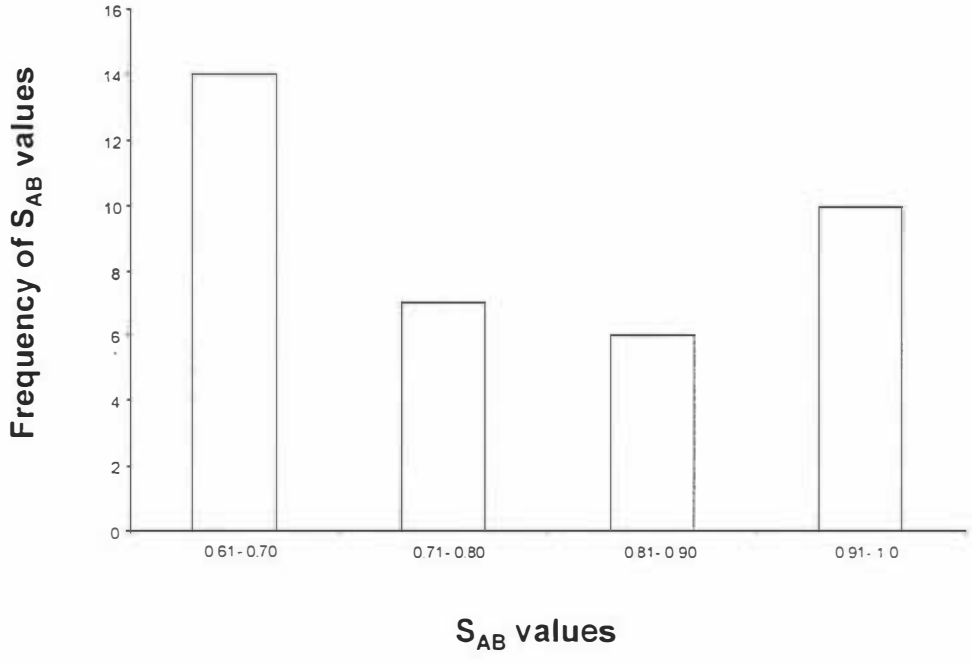


Figure 3-3: Bimodal distribution of S_{AB} values derived from pairwise comparisons between 9 *C.albicans* isolates from an healthy individual (A.26). One peak was found between 0.61 and 0.7 and a second peak was found between 0.91 and 1.

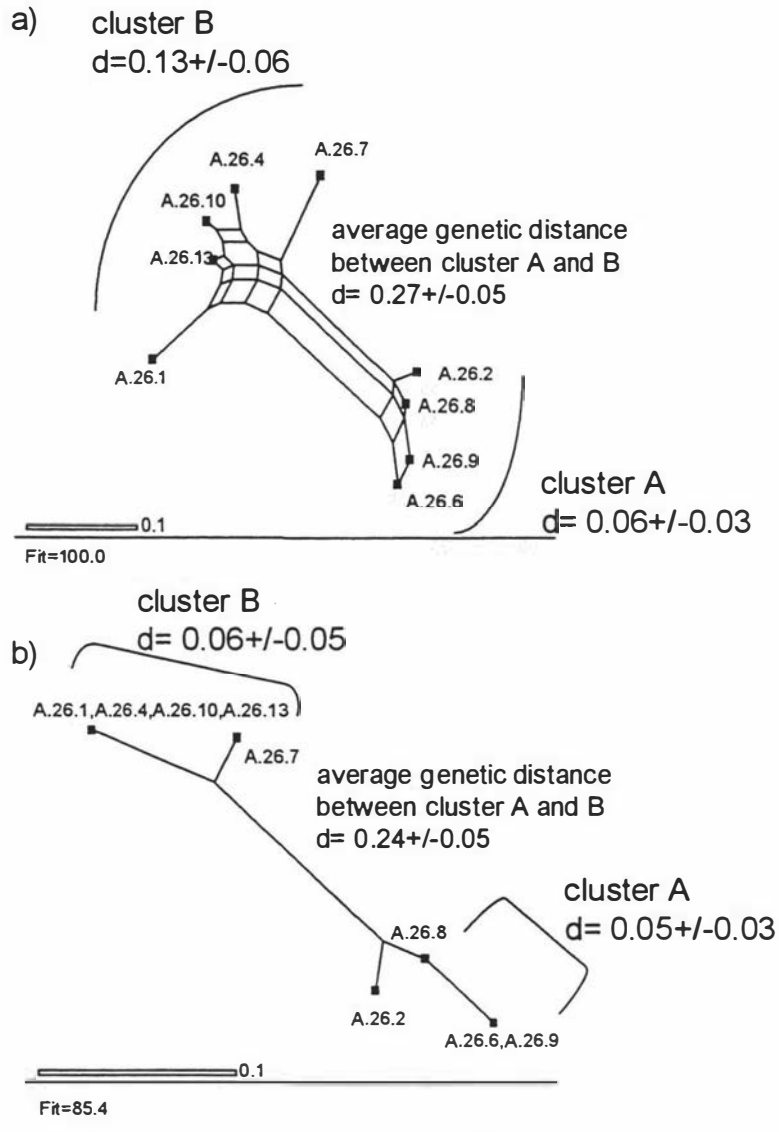


Figure 3-4: Analysis of Ca3 fingerprinting scoring data of nine *C.albicans* clones obtained from a saliva sample of an healthy individual under split decomposition. Splitsgraphs displayed were contructed using a) scoring data assessing band intensity and band size. The average pairwise genetic diversity (d) was 0.06+/-0.03 within cluster A and 0.13+/- 0.06 within cluster B.

Splitsgraph shown in b) was constructed when only presence or absence of bands was scored and the band intensity was eliminated. The average pairwise genetic diversity (d) within cluster A was 0.05+/-0.03 and 0.06+/- 0.05 within cluster B.

3.3.2 Genetic Variability of *C.albicans* Isolates Within an Highly Predisposed Individual

The extent of genetic variability between six strains harboured at a single site of an highly predisposed individual (criteria 2, see section 2.4.1.2) was investigated. Isolates were obtained from saliva samples (see section 2.4.2.1) for details on sampling methods) from a multiple myeloma patient (C1) and analysed by Ca3 fingerprinting. The average pairwise genetic distance of the *C.albicans* clones was 0.1+/-0.04. That is, a genetic diversity less than that found between the two saliva strains of a healthy individual. The relationship between isolates from this patient are shown in the splitsgraph of Figure 3-5. These data also show a good fit to the splitsgraph representation.

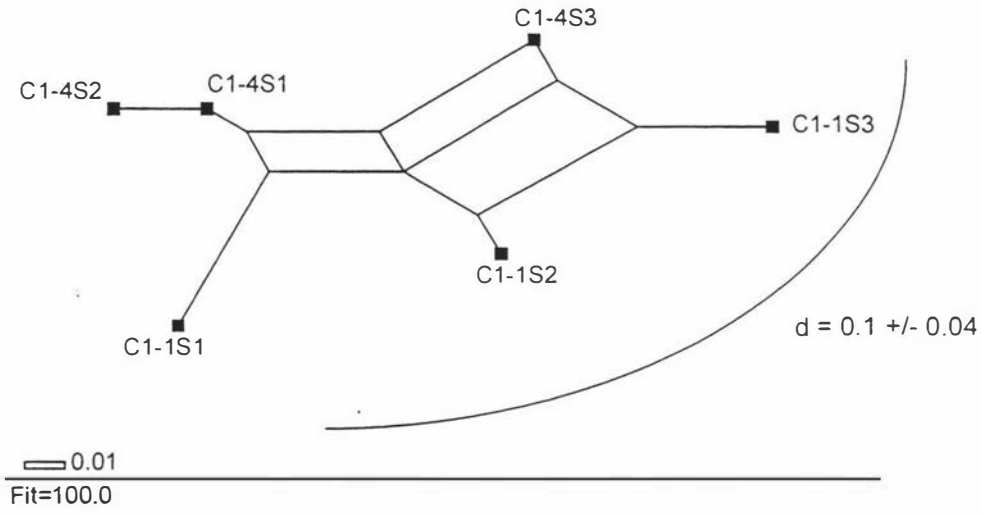


Figure 3-5: Relationship between *C.albicans* isolates which were obtained from saliva samples of a highly predisposed patient (C1). The average pairwise genetic diversity of the isolates was 0.1+/- 0.04.

3.3.3 Genetic Variability of *C.albicans* Isolates Between Healthy Individuals

One *C.albicans* clone (isolated from saliva, see section 2.4.2.1) from 13 healthy individuals (criteria 1, see section 2.4.1.1 for explanation of criteria) was analysed by Ca3 fingerprinting. S_{AB} values ranged between 0.46 and 0.89 (n=91). The average S_{AB} value was 0.63 +/- 0.07 and these data were normally distributed (please refer also to Figure 3-10). The majority (68%) of S_{AB} values were below 0.6.

The relationships between isolates analysed under split decomposition is shown in the splitsgraph of Figure 3-6. The average pairwise genetic diversity between these isolates was 0.36+/-0.07. The splitsgraph and distribution of values suggest a diversity of strains was harboured between healthy individuals. In this case the splitsgraph does not provide a good fit of the data to the graph representation, indicating many non tree-like patterns in the data.

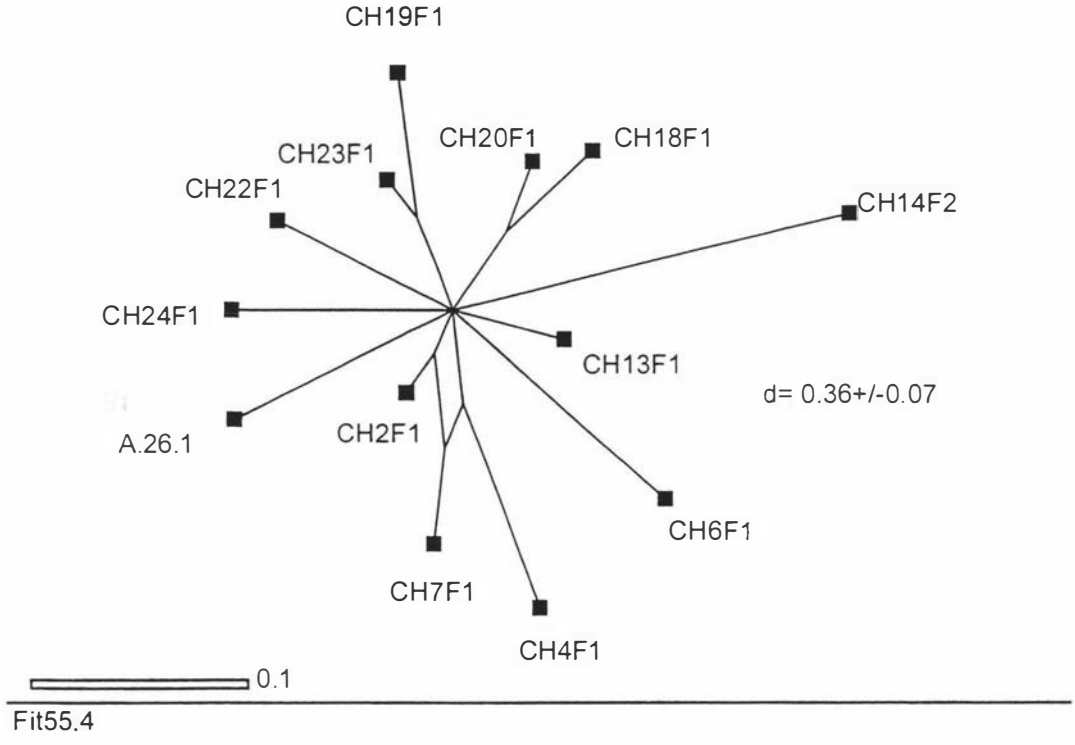


Figure 3-6: Analysis of the relationship between *C.albicans* isolates from saliva samples of 14 healthy individuals under split decomposition. The average pairwise genetic diversity between these isolates was 0.36+/-0.07.

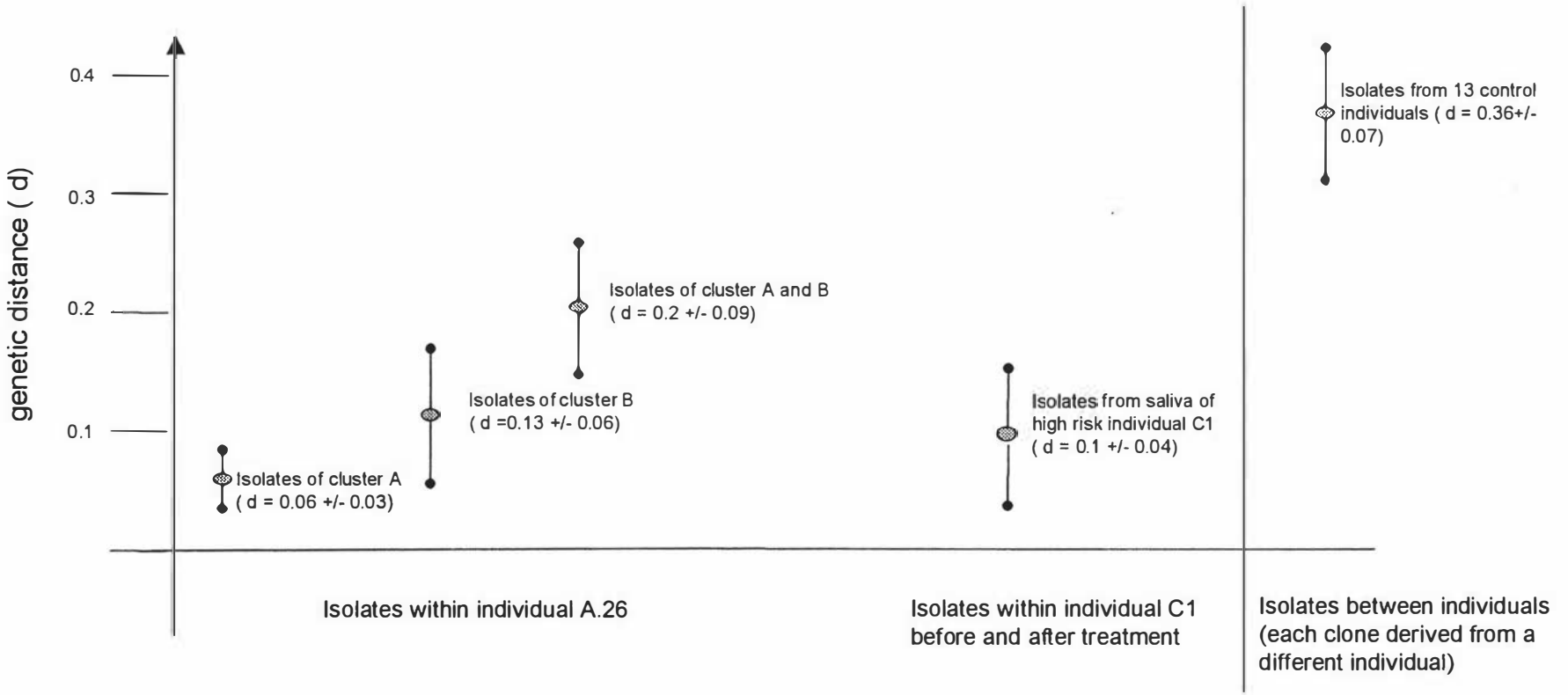


Figure 3-7: Average genetic distances (d) between *C.albicans* isoates from within (isolates from A.26 and C1) and between individuals (criteria 1). Distances within individuals were lower on average than between individuals.

3.3.4 Summary of Genetic Variability of *C.albicans* Isolates Within and Between Individuals

The genetic variability between isolates obtained from saliva samples within individuals (A.26, see also Figure 3-4; and an highly predisposed individual C1, see also Figure 3-5) and between individuals (see Figure 3-6) were analysed under split decomposition (see section 2.6.1.3) and results displayed in Figure 3-7. If conclusions are based on mean estimates of divergence then the results suggest the following:

- the genetic diversity between two strains harboured in a healthy individual differed from the genetic diversity between strains harboured in different individuals
- the distinctiveness of strains harboured within an individual was less than that between some strains harboured in different healthy individuals
- the genetic diversity of the strain colonising a predisposed individual was less than that of individual strains sampled from a healthy individual

3.3.5. Discussion

3.3.5.1 Ca3 Typing and the Study of Molecular Epidemiology

Recent studies have begun to investigate the molecular epidemiology of *C.albicans* using the Ca3 fingerprinting approach (Soll *et al.*, 1988; Soll *et al.*, 1991; Schmid *et al.*, 1992; Schröppel *et al.*, 1994; Schmid *et al.*, 1995). This method, which profiles the *EcoRI* banding patterns of repetitive DNA in the *C. albicans* genome, has been found to have high discriminatory power for microevolutionary changes (Pujol *et al.*, 1997), suggesting its suitability for population studies of commensal flora. Nevertheless, most Ca3 studies to date

have not yet analysed more than one clone from commensal *C.albicans* samples taken from individuals (one exception being a study by (Soll *et al.*, 1991) who sampled a maximum of two or, if vulva and vagina are considered as one location, three clones from each body location). Thus little information is known about the genetic variability of commensal *C.albicans* flora. The few studies which have investigated the similarity of commensal isolates within healthy individuals have found that commensal isolates obtained from the oral cavity were highly similar or identical, as reflected by S_{AB} values above 0.9 (Soll *et al.*, 1991; Hellstein *et al.*, 1993). Dissimilar commensal strains with S_{AB} values below 0.74 were found only when different body locations were sampled (Soll *et al.*, 1991) and these findings were explained by suggesting colonisation by different strains at these sites.

In the present study, up to nine clones per site were sampled for any one body location. Both the bimodality of the observed distribution of S_{AB} values (refer to Figure 3-3), and also the splitsgraph (which shows genetic relationships between isolates) suggest the existence of two strains and substrains (quasi strains; (Eigen, 1993)) in the oral cavity of healthy individual A.26 (see Figure 3-4). This interpretation of the results also suggests that the commensal substrains are distinguished by considerable genetic diversity as detected by the Ca3 fingerprint method.

When isolates were compared between 13 healthy individuals, an average pairwise S_{AB} value of 0.63 (one clone per person was analysed) was observed. This result was consistent with values reported elsewhere for similar comparisons (S_{AB} values of 0.69 (Schmid *et al.*, 1990), 0.67 (Soll *et al.*, 1991), 0.65 (Schmid *et al.*, 1992), and 0.68 (Lockhart *et al.*, 1995)). This finding supported the interpretation of strains and substrains from the isolates characterised from the oral cavity of a healthy individual and suggested that study group was also representative of a larger population.

The average pairwise genetic distance found between isolates, obtained from saliva samples from a patient at high risk to develop Candidosis, was 0.1+/-0.04. This suggested a genetic diversity similar to the diversity found within the two strains observed in a healthy individual. This result was interpreteted to suggest that this patient had been colonised by only one strain which existed with its quasi strains throughout chemotherapy. That is, there was no suggestion of strain replacement over the course of treatment in this individual. The extent of the substrain diversity detected by Ca3 fingerprinting poses a problem for investigation of putative strain replacement in individuals since conclusions would need to consider sufficient sampling of substrains to be confident of detecting interstrain differences. This poses a problem for the implementation in this context of Ca3 fingerprinting since it is both labor intensive and time consuming.

3.3.5.2 Ca3 Fingerprints Analysed Using Split Decomposition

Most analyses showed that, at least for small numbers of isolates, splitsgraphs provided a very good representation of the genetic diversity contained in Ca3 fingerprint profiles. This is also expected to be true for larger numbers of closely related strains. However, the observation of a poor fit when diverse strains were compared from 13 healthy individuals cautions phylogenetic interpretation of Ca3 profiles for more distantly related *C. albicans* strains. In such cases noise in the data due to misidentifcation of homologous bands and perhaps also nonindependence of bands could lead to contradictory patterns of relationship. In this situation split decomposition and other tree building procedures may not perform well (Swofford *et al.*, 1996).

3.3.6 Monitoring of Differently Predisposed Patients For Strain Replacement

A pilot study was undertaken to determine whether direct evidence could be found for strain replacement in individuals predisposed to Candidosis. Split

decomposition analysis with reference to individual substrain diversity was used to provide an analytical framework for determining the significance of observed genetic diversity between isolates sequentially sampled from individuals.

One dissappointing aspect of the study undertaken was that individuals in the study group investigated did not develop Candidosis as a result of the hospital treatments they received. As will be discussed, this posed a problem for analysis of results. For this reason only methodological conclusions have been drawn from the study presented here. Nevertheless, in this section (and the subsequent section on AFLP derived markers) the findings described provide a basis for future DNA investigations into strain replacement in individuals predisposed to Candidosis.

3.3.6.1. *C.albicans* Isolates Within an Highly Predisposed Individual Before and After Introduction of a Predisposing Factor

Isolates were obtained from saliva samples (see section 2.4.2.1) and rectal swabs (see section 2.4.2.3) from a multiple myeloma patient (C1) and analysed by Ca3 fingerprinting. Due to the time required to prepare and process Ca3 profiles only a relatively small number of isolates could be screened at once.

Six clones of *C.albicans* from patient (C1) were isolated from samples taken prior to the first course of neutropenia-inducing chemotherapy (when the patient was at low risk of acquiring Candidosis), and then subsequently five isolates were taken after chemotherapy (when the patient was at a higher risk). At the same sampling time *C.albicans* clones were also taken from saliva samples from the patient's partner FM1 (3 before and 2 after chemotherapy treatment of C1).

S_{AB} values for 16 isolates, obtained from saliva samples and rectal swabs from patient (C1) and saliva samples from patient's partner (FM1), remained

relatively high throughout the trial, which was reflected by S_{AB} values ranging between 0.69 and 1 and an average S_{AB} value of 0.87+/-0.06. The distribution of isolates displayed in a dendrogram (data not shown) did not show a clustering of isolates that derived either from one individual (C1 or FM1) or from a specific body location (mouth or rectum). These data indicate that both individuals were colonised with the same strain and its substrains, and that no evidence for strain replacement was obtained during the individual's first course of chemotherapy.

It was noted that patient C1 showed a 3.7-fold increase (510 c.f.u/ml to 1.9 x 10^3 c.f.u/ml saliva) in the colonisation intensity in the saliva during chemotherapy treatment. However, this increased cell count after chemotherapy was still in the range of cell counts found in healthy individuals (up to 10^4 c.f.u./ml saliva) (Odds, 1988). Neither patient nor partner showed any signs of Candidosis. Patient P1 was the only highly predisposed patient (criteria 2, see section 2.4.1.2) that could be recruited for this trial, therefore patient criteria were expanded in developing the study.

3.3.6.2 *C.albicans* Isolates Within Lower Predisposed Individuals Before and After Introduction of a Predisposing Factor

Patients at lower risk were more common than high risk patients and thus results could be obtained from a larger number of individuals. Individuals fitting three separate criteria (criteria 3, 4, 5 see also section 2.4.1.3) were considered part of a lower risk group. Table 3-2 gives an overview of underlying disease and predisposing factors for each criteria.

Table 3-2: Patients of lower risk group, their criteria and predisposing factors for potential strain replacement (for more details see section 2.4.1 Patient Selection)

patient criteria	predisposing factor	potential time for replacement
criteria 3	hospitalisation, antibiotics, surgery (HAS)	hospitalisation; isolates obtained before HAS were used previously as part of the group of healthy (criteria 1) individuals, see section 2.4.1.3.1
criteria 4	solid cancer, first course of chemotherapy	immunosuppression due to chemotherapy, see section 2.4.1.3.2
criteria 5	haematological cancer, first course of chemotherapy	immunosuppression due to chemotherapy, see section 2.4.1.3.3

Patients in the lower risk groups showed similar colonisation rates and intensities (see Table 2-2) and no signs of Candidosis before and after impact of predisposing factor (see criteria 2-5, section 2.4.1.3).

To use Ca3 fingerprinting for this larger study group a change in sampling strategy was made to screen more individuals. However, fewer isolates could be characterised (1 clone per person before and after predisposition).

The genetic relatedness between sets of isolates from members of this group was profiled by Ca3 fingerprinting and analysed under split decomposition (each set consisted of pre (patient no.+ F) and post (patient no.+ L) treatment isolates). The rationale behind the experimental design was that in the event of the analysis suggesting a significant difference between isolates in any given set, then culturing and analysis of additional isolates from the pre- and post treatment swabs could be made for that individual.

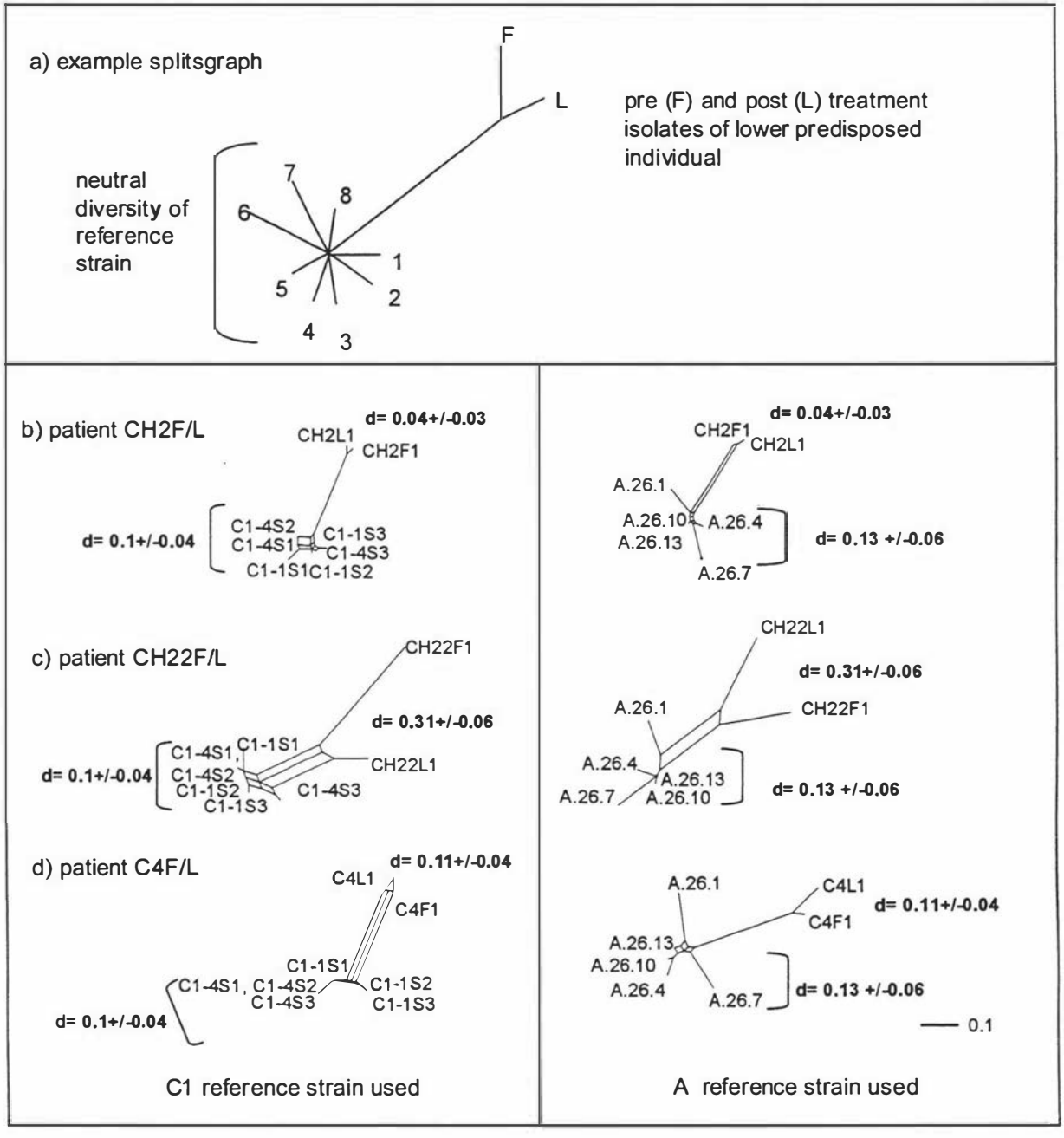


Figure 3-8 a-d: Analysis of genetic distances between isolates obtained before (F) and after (L) introduction of a predisposing factor. a) shows an example splitsgraph. To help interpret the significance of genetic distances between pre and post treatment isolates, comparisons were made to reference strains. b-d) Split decomposition analysis of isolates, obtained from saliva samples from predisposed individuals (C = criteria 3 and 4, CH = criteria 2) before (F) and after (L) their respective treatment. Genetic distances were calculated and compared either to cluster B isolates, obtained from saliva samples from an healthy individual (A.26.1 to 9), or isolates obtained from saliva samples from an highly predisposed patient (C1-1S1 to 3 and C1-4S1 to 3). This analysis was performed for all criteria 2,3, and 4 patients (not all data shown). The most diverse set of isolates (d = 0.31+/-0.06) was obtained from patient CH22.

Splitsgraphs were constructed, each containing a pre and a post treatment isolate (set) from the lower risk group as well as a group of isolates previously detected as strain and substrains (see Figure 3-8). Isolates previously interpreted as strain and sub or quasi strains were used to provide a framework for the natural variation of genetic distances within substrains. Thus, for interpreting the significance of the genetic difference between pre and post treatment isolates of this lower risk group, either (i) cluster B isolates from an healthy individual (refer to Figure 3-4a and b) or (ii) isolates obtained from saliva samples from an highly predisposed individual (see Figure 3-5) were included in the analysis under split decomposition. With one exception, the genetic distances between isolates from all sets were within the range of quasi strains, indicating no evidence for replacement.

The most diverse set of isolates characterised were from patient CH22 with a genetic distance of d= 0.311 +/- 0.06 between the pre and post treatment isolates. Further analysis revealed that post treatment isolates from two different patients (CH22L and CH23L) had a S_{AB} value of 0.96. The Ca3 fingerprints (Southern blot hybridisation pattern) of the set isolates from patients CH22 and 23 are shown in Figure 3-9. A comparison between the two pre treatment isolates (CH22F and CH23F) showed a S_{AB} value of 0.56 which was below the average S_{AB} value of 0.63 for isolates obtained from different healthy individuals (see section 2.4.1.1). The high similarity between isolates from patients CH22 and CH23 may indicate a change in *C.albicans* colonisation in one of these patients. Considering that the S_{AB} value between sequential isolates of patient CH22 (CH22F/L=S_{AB} 0.66) was higher than between isolates from patient CH23 (CH23F/L= S_{AB} 0.8), it was more likely that this change occurred in patient CH22. Although, this finding might be an indication that, following treatment, strain replacement occurred in patient 22, additional isolates from pre and post treatment swabs were not cultured to test this possibility. Investigation of this possibility would have been more interesting in the event that patient 22 had developed

Candidosis. Rather than characterise the *C.albicans* flora of patient 22 in more detail with Ca3 fingerprinting, further efforts, as described in the following chapter, were directed at obtaining a more rapid screening procedure for use in strain replacement studies.

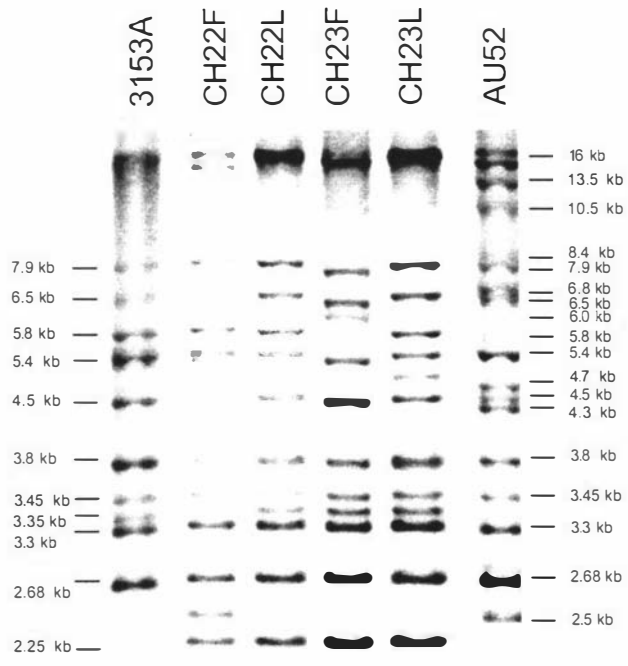


Figure 3-9: Ca3 fingerprinting patterns for sets of isolates from criteria 3 patients CH22 and CH23. Isolates F were obtained prior treatment and isolates L were obtained after treatment.

3.3.6.3 Comparison Between Genetic Variability of *C.albicans* Isolates Derived From Healthy and Predisposed Individuals

To test whether variations in the *C.albicans* flora between individuals of different criteria exist, similarity values of healthy individuals (criteria 1, refer also to section 2.4.1.1) and cancer patients (criteria 3, 4, 5 see section 2.4.1.3) were compared and these are presented in Table 3-3.

Table 3-3: Average and range of similarity values between isolates from healthy individuals and cancer patients

no. of group	group of individuals (n = number of individuals)	range of S_{AB} values from matrix	average S_{AB} (n = number of S_{AB} values)
A	healthy individuals (see criteria 1) (n=13)	0.46 - 0.89	0.63 +/- 0.07, n = 91
B	cancer patients (see criteria 2, 4, 5) prior to chemotherapy, (n = 15)	0.29 - 0.79	0.56 +/- 0.08, n = 105

As indicated in Table 3-3, isolates from cancer patients (group B) showed a wider range of S_{AB} values (0.29 to 0.79), suggesting that this group comprised more diverse isolates than the criteria 1 isolates. To test whether this difference was significant the distribution of S_{AB} values was plotted in a histogram which was displayed in Figure 3-10. This analysis revealed that 43% (40/91) of the S_{AB} values between isolates from healthy individuals were below 0.6 compared to 68% (76/105) of S_{AB} values between isolates from cancer patients. The null hypothesis (no difference in genetic similarity between isolates of cancer and control group) was rejected in a *z*–test with P value < 0.001.

Analysis under split decomposition (see section 2.6.1.3), shown in Figure 3-11, revealed that the distinct diversity did not derive from atypical strains within the group of cancer isolates. The reason why isolates from cancer patients were more diverse than isolates from healthy individuals remains unclear.

To test whether *C.albicans* isolates from healthy individuals and cancer patients are genetically distinct groups group A and group B isolates (see Table 3-3) were combined and analysed under split decomposition (data not shown). No distinct difference between isolates from cancer patients and healthy individuals was observed.

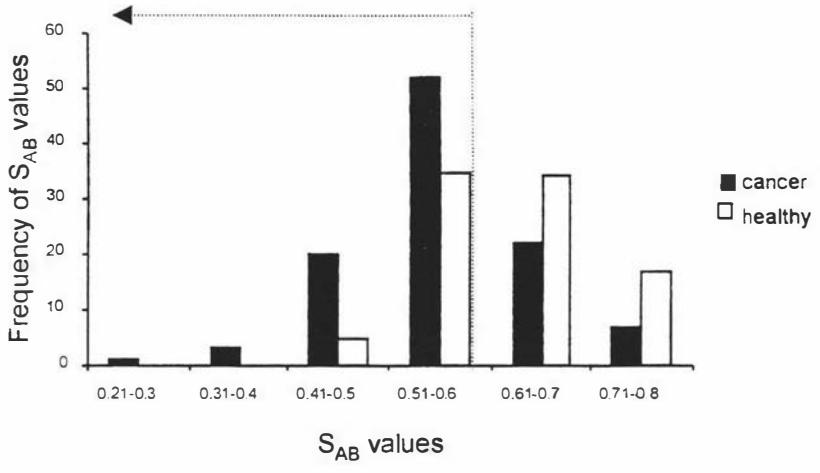


Figure 3-10: Distribution of S_{AB} values between healthy (criteria 1) individuals (□) and cancer (criteria 2, 4, 5) patients (■). Data were normally distributed. In the group of cancer patients were 68% (76/105) of all pairwise comparisons below 0.6. In comparison, in the group of healthy individuals were 43% (40/91) of S_{AB} values below this value.

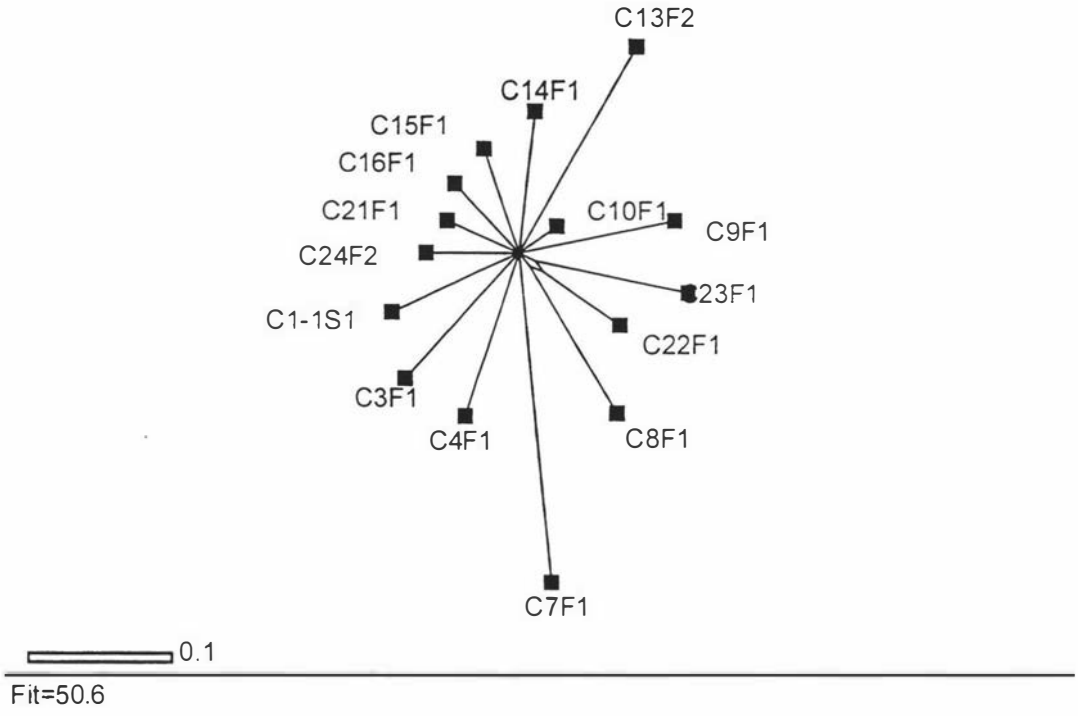


Figure 3-11: Split decomposition analysis of Ca3 fingerprinting data from cancer patients (criteria 2, 4, and 5) to deterime whether atypical strains may be responsible for the higher diversity. As shown above no significantly atypical strains have been found.

3.3.6.4 Discussion

Two observations have led to the hypothesis of strain replacement. These are that clinical isolates from hospitalised patients are more similar than isolates from nonhospitalised patients and that HSP strains are relatively more common in Candidosis patients (clinical isolates) than in healthy individuals (see also section 1.4). The present study has highlighted the considerations needed to test this hypothesis. These have been outlined under following subheadings.

3.3.6.4.1 Study Groups

In the present work the patients considered closest to ideal for strain replacement studies were multiple myeloma patients undergoing the first neutropenia-inducing chemotherapy (see section 3.1.2). However, very few such patients were available for the present study. For this reason the patient selection criteria was relaxed to include lower predisposed patients.

However, none of these patients went on to develop Candidosis, making it difficult to determine whether predisposition really occurred. This finding, whilst in some respects not surprising (see below), identifies a problem in testing the relationship between strain replacement and predisposition to Candidosis. That is, while Candidosis may have higher incidence in individuals receiving certain hospital treatments this says little as to whether specific individuals actually become predisposed to the disease as a result of that treatment. Evidence of this only follows if individuals subsequently develop the disease. In this case interpretation of whether strain replacement has or has not occurred is straightforward. However, when a specific individual does not contract the disease, interpretation of the occurrence or not of strain replacement and it's putative relationship to predisposition is less clear.

Hence, selection of a larger study group may help in future studies. A feature that limited the present study was that at the time this work began, no documented epidemiological data was available for the incidence of nosocomial transmission at Palmerston North hospital where the present study was based. Thus it may not have been the best place to undertake the strain replacement work.

Worldwide the highest nosocomial incidence of fungal infections (with *C.albicans* being the most frequent pathogen, (Fridkin and Jarvis, 1996)) was

reported for immunosuppressed patients such as AIDS patients (Fisher-Hoch and Hutwanger, 1995) and cancer patients undergoing neutropenia-inducing chemotherapy treatment (Meunier *et al.*, 1992), bone-marrow or solid organ transplants patients (Pfaller *et al.*, 1987; Odds *et al.*, 1989), severely ill burn patients, and patients in neonatal intensive care units (Prasad *et al.*, 1987; Butler and Baker, 1988). Candidemia is the most extensively studied nosocomial invasive fungal infection (Fridkin and Jarvis, 1996). The ESR (Environmental Science and Research Ltd, New Zealand) reported approximately 5 cases of candidemia in 1988, which has risen to approximately 73 in 1998. In a previous study conducted at the Wellington hospital in New Zealand indirect evidence for nosocomial transmission of hospital-specific strains has been obtained. In contrast, no indication for such transmission was obtained for Auckland hospital, New Zealand (Schmid *et al.*, 1995). These observations may suggest Wellington hospital as a potential site for further studies.

However, consideration should also be given to statistical data from the CDC (Center for Disease Control) and NNIS (National Nosocomial Infections Surveillance System) on nosocomial infections which suggest low chances of success for monitoring the transition of a healthy to an infectious state in hospitalised patients developing Candidosis. These studies showed that the incidence of nosocomial fungal infections in the States in 1990 was on average 3.8 per 1000 discharges, of which up to 78.3 % were *Candida* infections and *C.albicans* being the most frequent infection causing species (Fridkin and Jarvis, 1996). This means that the incidence for nosocomial *C.albicans* infections was approximately 0.3 % or in other words, 3 cases of *C.albicans* infections per 1000 discharges. Although the numbers of Candidosis incidences in risk groups (see above) were much higher, e.g., in burn/trauma patients (16.1/1000 discharges), oncology (8.6/1000 discharges), and high risk nursery (7.6/1000 discharges) (Beck-Sague *et al.*, 1993), suggesting that a large number of patients needs to be

analysed to even follow one patient through the transition from a healthy to a infectious state.

In 1998 a total of 108 Candidosis cases were documented in New Zealand, of which 57% (62/108) were caused by *C.albicans* (personnel correspondence, Environmental Science and Research Ltd, New Zealand). The relatively low incidence of Candidosis and the low prevalence of HSP cluster strains in New Zealand (Schmid *et al.*, 1999) provide an obstacle to study replacement.

3.3.6.4.2 Sampling Protocols

A second issue needing consideration for future studies concerns sampling of patients flora, although the present data do not provide information for optimal sampling protocols. To detect replacement, it is possible that samples may need to be monitored from an earlier stage than was done in the present work since it is not known at what moment -if at all- during the transmission from commensal to pathogenic state the hypothesised replacement is likely to occur. Another potential problem for detection concerns sampling from malignant patients, as studied in the present study. These often develop systemic infections termed candidemia (when *Candida* cells can be detected in the bloodstream), and oral Candidosis (Odds, 1988). The patient's colonized gastrointestinal tract is most likely the route of entry into the bloodstream (Anaissie, 1992). Thus blood and saliva samples should be obtained for analysis of such patients. In the present study only saliva samples and some rectal swabs could be obtained from cancer patients and this type of sampling could be inadequate to detect replacement. Future studies need to consider sampling protocols that will optimize the success of detecting strain replacement.

3.3.6.4.3 Substrain Diversity and Assessment

In the present study, in order to implement the Ca3 method on a larger study group, the number of initial clones screened was reduced to two sequential isolates per individual. However, the results from the analysis of multiple isolates sampled from a healthy individual, soon made it apparent that interpretation of Ca3 profiles for detecting replacement would be problematic. This was indicated by the range of S_{AB} values or genetic distances found between substrains of commensal *C.albicans*. In the present pilot study, the genetic diversity of substrains occurring at a single site in individuals was used as a reference to interpret the significance of genetic difference observed between *C.albicans* isolates from pre and post treatments (see Figure 3-8). In all cases, but one, the genetic distances between isolates from these sets were within the range of quasi strains, indicated no detection of replacement.

In the case of pre and post treatment isolates, S_{AB} values/genetic distance data for individual CH22F/L showed a significant difference and high similarity to an isolate obtained from patient CH23. The Ca3 fingerprints (see Figure 3-9) also clearly indicate the similarity between these isolates from patient CH22 and CH23. To test for possible strain replacement in this case (against a possible explanation that two distinct strains were present on the individual before treatment; as was the case for *C.albicans* found in the mouth of individual A.26) further isolates from pre and post treatment swabs form each individual could be examined. However, in this case retrospective isolate sampling from the pre and post treatment swabs was not made due to the patient not showing signs of Candidosis. Rather efforts were directed at obtaining a more rapid screening procedure for use in strain replacement studies. This decision was taken since it had become clear by this stage of the project that a faster method to genetically characterise strains, would allow analysis of a larger study group, improved pre and post treatment sampling and also better estimates of commensal flora population diversity.

3.3.6.4.4 Commensal Flora of Cancer Patients

Until the present study there existed no Ca3 fingerprinting data for commensal isolates from cancer patients. The results reported here suggested a more diverse flora in cancer patients prior to chemotherapy. It is unknown how general these findings may be or what the changes in diversity represent. Atypical strains may be a possible explanation for this phenomenon. Such strains which were primarily found in patients infected with the immunodeficiency virus lead recently to the discovery of the new species *C.dublinensis* (Sullivan *et al.*, 1995). However, this possibility was tested under split decomposition and it was shown that no atypical strains were responsible for the high diversity in this sample (see Figure 3-11). Another possible explanation for the higher diversity of *C.albicans* isolates in cancer patients may be that conditions in an early stage of the malignant disease, i.e. minor changes in host defence mechanisms, may allow more diverse strain types to colonise the host or possibly to replace previously present strains. This hypothesis needs to be tested in future studies.

HSP strains have been described as the most frequent strain type, and responsible for the lower similarity of clinical isolates compared to commensal isolates in one geographical locale (Schmid, 1993). An example for clinical isolates being genetically more similar than isolates from healthy individuals has been reported for AIDS patients (Schmid e*t al.*, 1992). In contrast, clinical isolates obtained from vaginitis patients showed the same genetical diversity as isolates from healthy individuals, a result consistent with no replacement having occurred in this patient group (Schmid *et al.*, 1993). In the present study the suggestion that isolates from some cancer patients were genetically more diverse than isolates from healthy individuals, is also consistent with a conclusion that no HSP strain replacement occurred in the individuals studied. It would be very interesting to investigate strain diversity in individuals developing Candidosis.

3.4 Ca3 Typing of Potential Environmental Sources of *C.albicans*

Potential environmental sources of nosocomial infection in the Palmerston North Haematology Ward were examined using Ca3 fingerprint typing (see section 2.5.6.4).

3.4.1 Occurrence of *C.albicans* in the Environment of High Risk Patients

Attempts to culture 144 swabs (see section 2.4.2.4) showed the presence of the yeast on three moist surfaces (the sinks in patient rooms no. 7 and 2, and from the bath tub in the male bathroom).

3.4.2 Discussion

Various studies have reported the transmission of *Candida* species to high risk patients via infusates, biomedical devices (Moro *et al.*, 1990; Finkelstein *et al.*, 1993), or from the hands of health care worker (Casewell and Phillips, 1977; Phleps *et al.*, 1986; Vazquez *et al.*, 1991; Sherertz *et al.*, 1992; Strausbaugh *et al.*, 1994). However, many of these studies suffered from the low discriminatory power and reproducibility of the typing methods used (Hunter, 1991). Using the Ca3 fingerprinting, a method with high discriminatory power, it was shown that isolates obtained from hospitalised patients were genetically significantly less diverse than isolates from nonhospitalised patients indicating indirect evidence for nosocomial transmission of *C.albicans* (Schmid *et al.*, 1995). We have successfully isolated and typed *C.albicans* from moist surfaces of the haematology ward of Palmerston North Hospital, the locale of patients at high risk. Thus the methods described here would provide a useful approach for future studies to obtain direct evidence for sources of nosocomial transmission.

CHAPTER 4 - TESTING THE CA3 PHYLOGENY AND DEVELOPMENT OF HSP SPECIFIC PCR MARKERS

4.1 Introduction

Given the epidemiological significance of the finding from Ca3 fingerprint data that there exists a closely related group of *C.albicans* strains which are prevalent in patients developing Candidosis (Schmid *et al.,* 1993; Schmid *et al.,* 1999), an independent test of this hypothesis was sought using other molecular markers. These markers were obtained using the amplified fragment length polymorphism (AFLP) method.

The AFLP technique (Vos *et al.*, 1995; Lockhart and McLenachan, 1997) is a fingerprinting method that combines the advantage of finding polymorphic genome regions by RFLP (restriction fragment length polymorphism) with the efficiency of DNA amplification by PCR. The AFLP technique uses restriction digestion, ligation of common adaptors to the restriction fragments, and PCR amplification using primers based on the adaptor sequences. To select a manageable number of all fragments, the primers have one, two or three arbitrary base extensions at their 3' end. The selectively amplified restriction fragments can be resolved on denaturating polyacylamide gels and visualised by silverstaining (refer to section 2.5.7.3 for detailed method description).

Large numbers of genetic markers can be generated with AFLP. These markers result from genomic differences between strains due to insertion/deletions and the loss/gain of restriction sites. Under some conditions these data provide sufficient levels of genetic diversity to evaluate phylogenetic relationships between species (Majer *et al.*, 1996; Rosendahl and Taylor, 1997). AFLP is a time-and cost-efficient method with high reproducibility and resolution, which makes it equal or superior to those of other markers (e.g. alloenzymes, random amplified

polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), and microsatellites (Mueller and LaReesa Wolfenbarger, 1999). An additional advantage of AFLP, which is exploited in the current work, is that this marker system also readily allows the isolation, cloning and sequencing of polymorphic bands for the generation of novel genetic markers (e.g. Lockhart and McLenachan, 1997). Its use in the present study was also encouraged by recent results published on the analysis of genetic variation in yeast (de Barros Lopez *et al.*, 1999). In this, AFLP was shown to be effective for discrimination at the inter and intraspecific levels in *Saccharomyces cerevisiae* (de Barros Lopez *et al.*, 1999). It has also been useful in studies on other fungal organisms (Majer *et al.*, 1996; Mueller *et al.*, 1996; Rosendahl and Taylor, 1997).

4.2 Preliminary Investigation of Usefulness of AFLP

To investigate the potential of the method, a preliminary AFLP analysis with a small number of *C.albicans* strains was first performed. This was done to investigate (i) the reproducibility of AFLP profiles, (ii) the variability of AFLP fingerprints, and (iii) whether the AFLP data would give tree-like data useful for testing phylogenetic hypotheses (Page and Holmes, 1998).

(i) Reproducibility of AFLP fingerprints

Multiple DNA extractions of isolates CLB43, YSM1, and AU47 were prepared (see section 2.5.1.1) and selectively amplified (see section 2.5.7) using 4 different AFLP primers (*Mse*I -CAC, -CAG, -CTG, and -CTC; see Table 2-7). Figure 4-1 displays AFLP profiles derived from multiple DNA extractions of these three isolates.

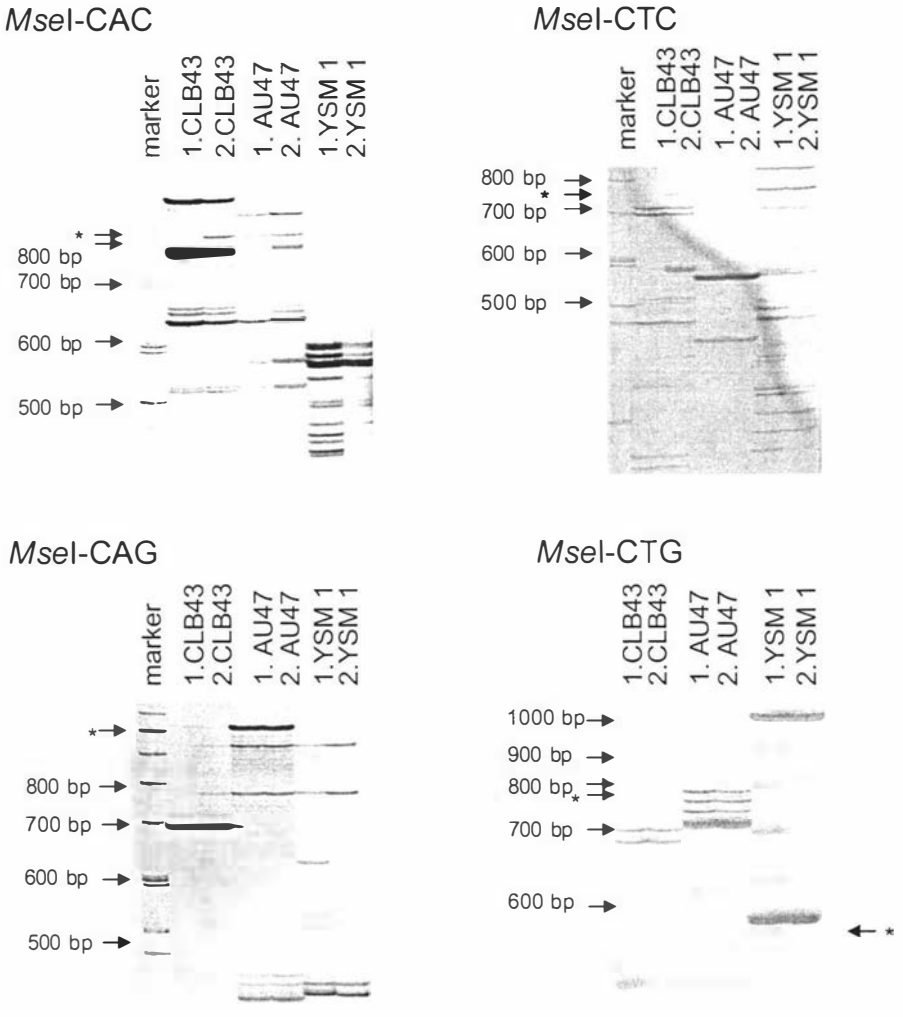


Figure 4-1: AFLP profiles that derived from selective primers *Mse*I-CAC, -CTC, -CAG, and -CTG from multiple DNA extractions of three *C.albicans* isolates (CLB43, AU47, and YSM1). Bands that were not present in both DNA extractions were marked by *, and were ignored for further analysis.

Profiles were scored (see section 2.5.5.2), and the reproducibility was evaluated according to reproducibility coefficients, that were calculated by the following equation:

$$\frac{\text{no. of bands present in both DNA extraction}}{\text{total no. of bands}} = \text{reproducibility coefficient}$$

Reproducibility coefficients calculated from AFLP banding patterns, displayed in Figure 4-1 indicate a high degree of reproducibility under the experimental conditions used (Table 4–1).

Table 4-1: Reproducibility coefficient of multiple DNA extractions analysed by four primer combinations

Primer → Isolate ↓	*Mse*I-CAG	*Mse*I-CAC	*Mse*I-CTC	*Mse*I-CTG
CLB43	1(11/11)	0.81 (9/11)	0.9 (20/22)	0.88 (15/17)
YSM1	1 (28/28)	0.86 (13/15)	1 (9/9)	1 (15/15)
AU47	0.94 (17/18)	0.92 (13/14)	0.75 (12/16)	1 (19/19)

(ii) Variability of AFLP fingerprints

To test whether the variability of AFLP fingerprints would allow intraspecific discrimination, 10 *C.albicans* isolates from different individuals were analysed using three different AFLP primers (*Mse*I-CTC, -CTG, and –CAC, see Table 2-7). Two additional isolates (with a S_{AB} value of 0.97) from the same individual were also studied. Figure 4-2 shows that the AFLP profiles for the first 10 isolates were variable and that the profiles for isolates from the same individual (C1-1R2 and C1-1S2) were highly similar.

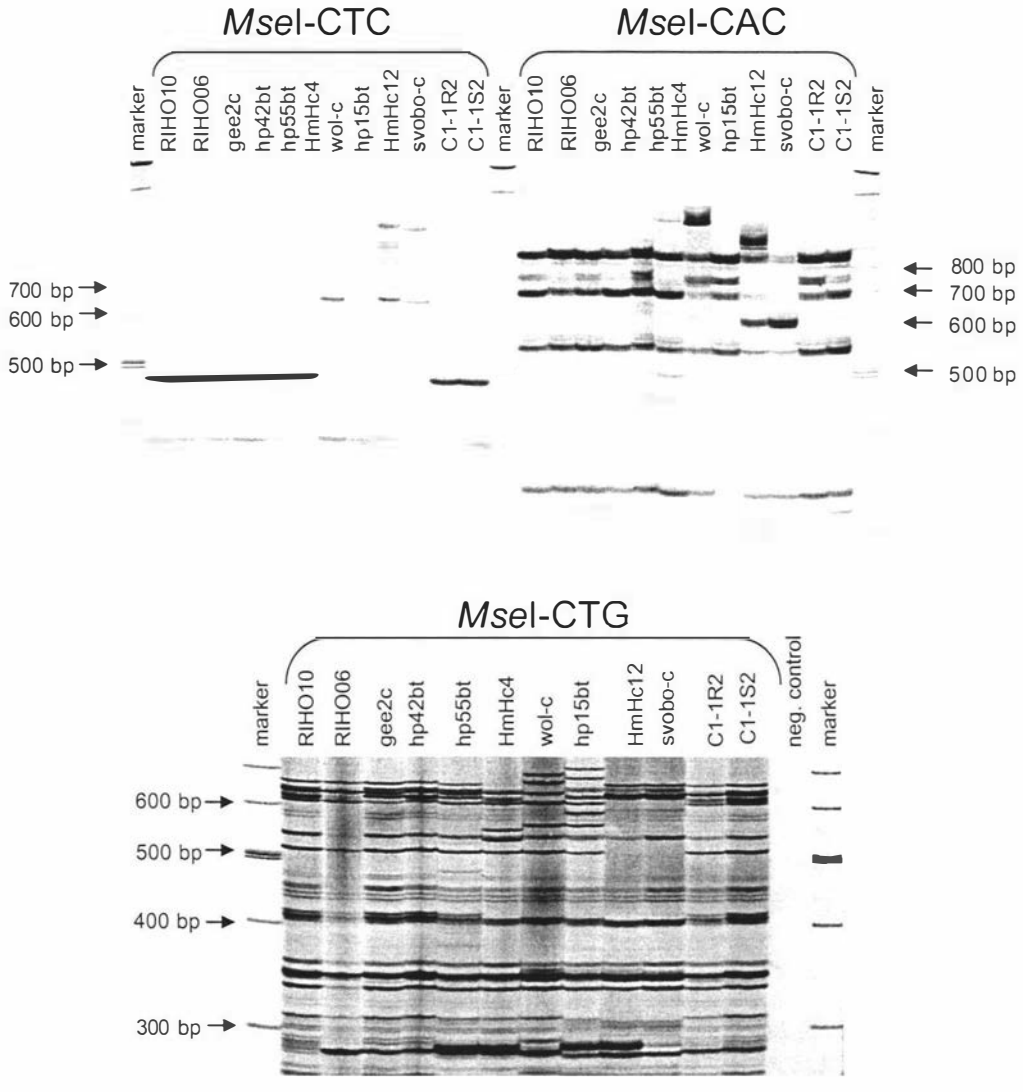


Figure 4-2: Test of AFLP fingerprinting for intraspecific discrimination between different *C.albicans* isolates. AFLP profiles were amplified using selective primers *Mse*I-CTC, -CAC, and -CTG from 10 different *C.albicans* strains. Isolates from the same individual (C1-1S2 and C1-1R2) were highly similar, while banding patterns from isolates from different individuals (remaining isolates) differed considerably.

(iii) Usefulness of data for testing phylogenetic hypotheses

A splitsgraph (see section 2.6.1.3) was constructed from a data matrix containing information from recoded bands (presence and absence) in 3 different AFLP profiles (*Mse*I-CTG, -CTC, and -CAC, see Table 2-7). The splitsgraph shown in Figure 4-3 is tree-like and confirms the clustering of isolates associated with the HSP group. The split between HSP and noncluster isolates was supported by bootstrapping at a level of 91%.

Taken together the results described under (i-iii) were encouraging that a similar study looking at a larger number of isolates would help provide a good test of the earlier Ca3 derived HSP phylogeny.

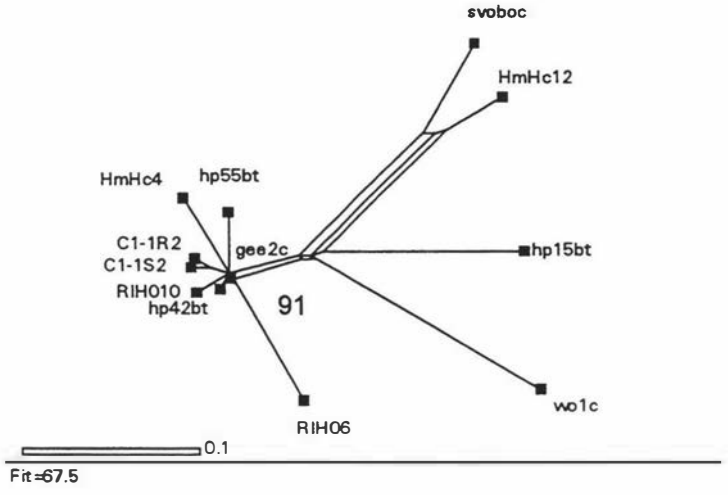


Figure 4-3: Analysis of AFLP fingerprinting patterns from 12 *C.albicans* isolates under split decomposition. The treelike splitsgraph confirmed the clustering of strains associated with the HSP group as demonstrated by earlier Ca3 fingerprinting studies (HSP strains: RIHO6, RIHO10, hp42bt, hp55bt, HmHc4, gee-2c). The HSP/non HSP split was supported by bootstrapping at a level of 91%.

4.2.1 A Comparative Study of 36 *C.albicans* Isolates Using Ca3 Profiles and AFLP

AFLP analysis was expanded to a selection of 36 *C.albicans* isolates from a worldwide collection. The 36 selected isolates included 27 samples from infection sites of Candidosis patients (pathogenic isolates) and 9 samples isolated from the commensal flora of healthy individuals (commensal isolates). In previous Ca3 fingerprinting analysis these isolates showed S_{AB} values (see section 2.6.1.2) ranging between 0.32 and 0.94.

AFLP profiles that derived from 4 different selective amplifications (*Mse*I-CTC, -CTG, -CAC, -CAG, see Table 2-7) are shown in Figure 4-4. Profiles were scored for the presence and absence of bands, and scoring data entered into a datafile (see Appendix 6)

4.2.1.1 Data Structure and Tree Building From Ca3 and AFLP Profiles

Quartet puzzle trees were reconstructed for 36 *C.albicans* isolates using a parsimony criteria as implemented under PAUP 4.0 (Swofford, 1998). AFLP (see Figure 4-4, please refer to Appendix 9 for scoring data) and Ca3 profiles (typed in previous studies, please refer to Appendix 7 for scoring data) were scored for the presence of bands and in the case of the Ca3 data, also for band intensity (as done in earlier Ca3 tree building studies). Trees are displayed in Figure 4-5. Numbers on trees are puzzling support values ("reliability values") and measure how often groupings in the final tree were also found in intermediate puzzling trees. High numbers reflect strong support for taxa partitions (particular groupings of strains). Low values result from patterns indicating contradictory relationships. Strains in the HSP group of both AFLP and Ca3 trees have been highlighted (RHO6, hp10bt, hp38an, RIHO1, CLB42, YSU649, var1.10, AU25, jam2c, HUN125, HUN121, HmHc4, OD8826, HUN92, Otg16, and Fj11). A few strains occurred in the HSP cluster only in the Ca3 tree (e.g., CLB43, Figure 4-5a) or only in the AFLP tree (e.g., C1-1S2, Fj10 and YSM1, Figure 4-5b). The support for the HSP group was lower (12%, see Figure 4-5a) in the tree that derived from Ca3 data than in the tree that derived from AFLP profiles (69%, see Figure 4-5b). In general, lower puzzling support values for internal branches in the Ca3 tree suggested these Ca3 data may not have as much phylogenetic resolution as the AFLP data.

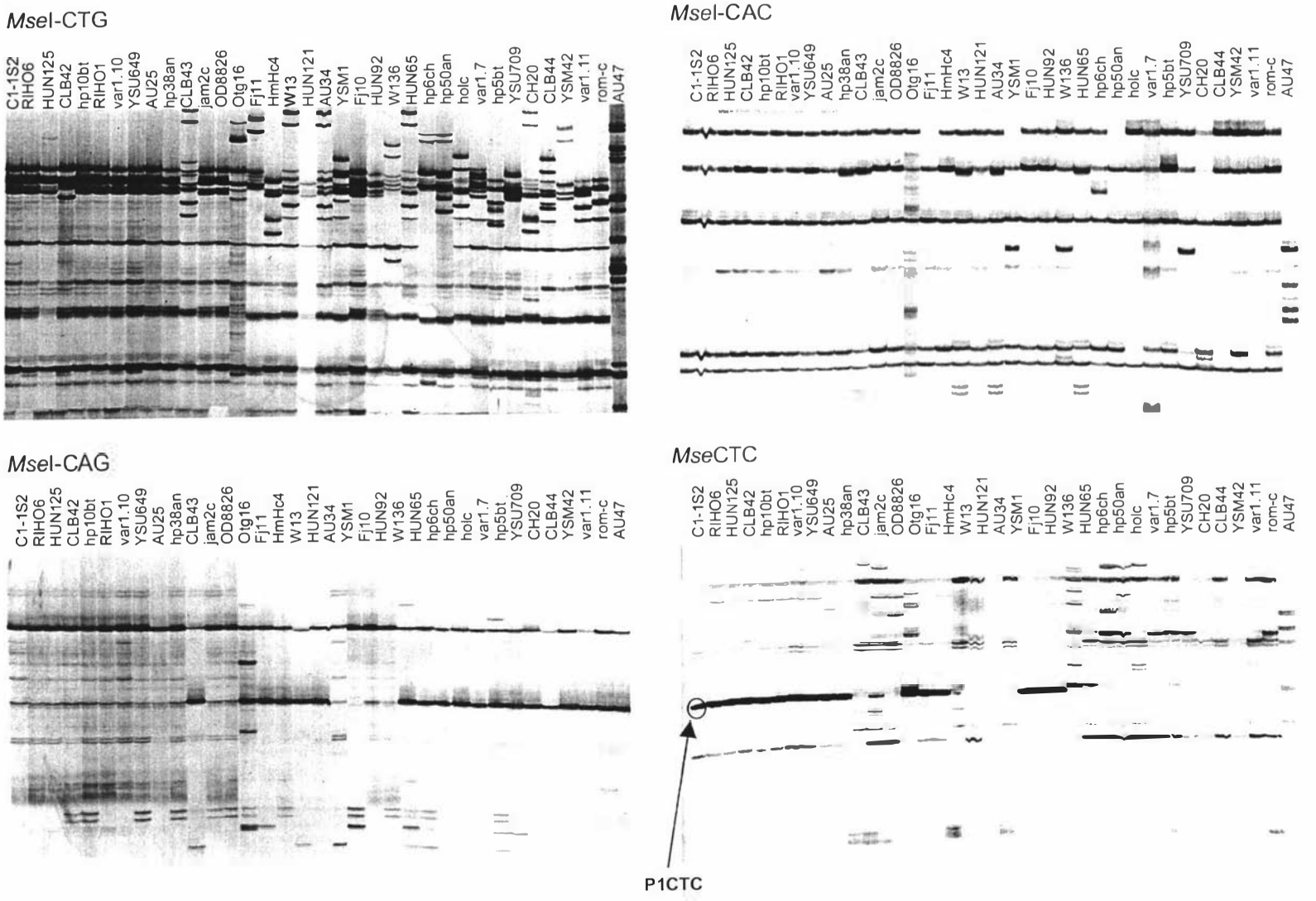


Figure 4-4: AFLP profiles of 36 *C.albicans* isolates generated with selective amplification using primers *Mse*I-CTG, *Mse*I-CAC, *Mse*I-CAG, and *Mse*I-CTC. All four primers generated considerably different banding patterns. Polymorphic bands, e.g. P1CTC were isolated and sequenced.

To see if the tree-like quality of the Ca3 data could be improved, the contribution of band intensity as a character was investigated. This was done because split decomposition analysis of Ca3 data from preliminary strain replacement studies (see section 3.3.1) suggested that the data quality could be improved by omitting band intensity as a character. Band intensity was examined in respect of how it influenced the Ca3 tree topology and, specifically whether strains were present or not in the HSP cluster. A quartet puzzle tree made using a parsimony criteria was reconstructed using data matrices from Ca3 profiles coded for presence and absence of bands and band intensities (the matrix is given in Appendix 7). This tree is shown in Figure 4-6a, and is the same as previously shown in Figure 4-5a, but is included again for direct comparison with a Ca3 puzzle tree obtained when band intensity was excluded as a character (Figure 4-6b; see Appendix 8 for this data matrix). A comparison of the two trees shows that when band intensity was excluded as a character some strains no longer associated with the HSP cluster. Excluding band intensity as a character also did not generally improve the puzzling support values in the reconstructed tree.

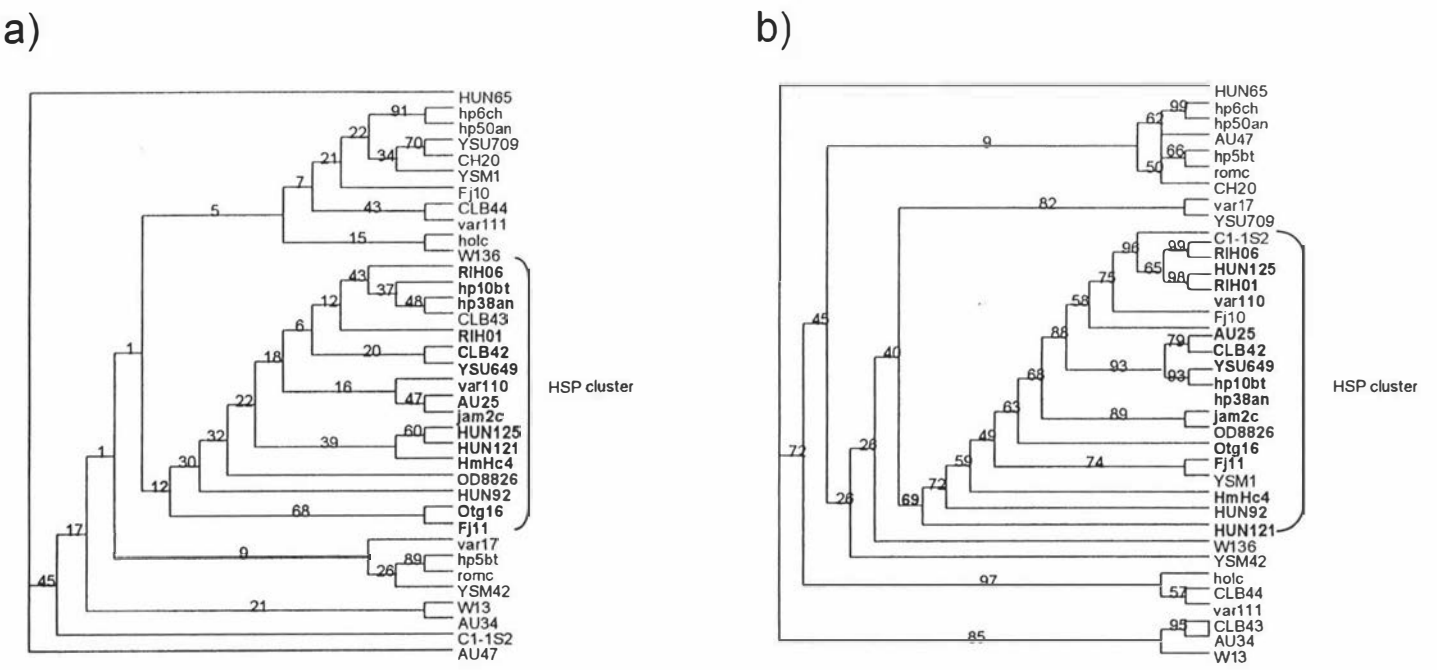


Figure 4-5: Quartet puzzle trees were constructed using a parsimony criteria. The quartet puzzle tree in (a) was build using Ca3 scoring data, while the quartet puzzle tree in (b) was build using AFLP scoring data. Numbers on trees are puzzling support values. High numbers reflect strong support for the particular partition. Support for the HSP group was lower in the tree that derived from Ca3 data (12) than in the tree that derived from AFLP data (69). Strains in the HSP group of both tree have been highlighted.

The degree of compatibility of patterns within different data sets was next investigated using quartet puzzling and a tree comparison metric. The partition metric was chosen since it is particularly useful when comparing trees which are highly similar and differ only in respect of local rearrangements of some taxa between the compared trees (Steel and Penny, 1993).

Because puzzling steps in quartet puzzling tree building are heuristic, the similarity of trees obtained from independent tree building runs for different data sets can be used to investigate the relative tree-like qualities of these data sets. For this comparison, the partition metric difference was calculated for three independent tree building runs on each data set. The rationale for doing this was that if a given data set showed good tree-like properties then it would be expected to give resolved similar or identical trees in independent runs. If trees were identical between runs then this would give an average partition metric difference between the resulting trees close to or equal to zero. However, this result would not be expected in the absence of hierarchical structure in the data. In this case, ambiguity in phylogenetic structure could result in average difference estimates much greater than zero. Results from this study are shown in Table 4-2.

Table 4-2: Partition metric tree comparison on parsimony quartet puzzling trees

data set	n= no. of sites p= no. of parsimony-informative sites	average difference
Ca3 data including intensities	n= 46 p= 37	4 +/- 3.4
Ca3 data without intensities	n= 46 p= 37	2.6 +/- 1.15
AFLP	n= 73 p= 66	0
AFLP & Ca3 with intensities combined	n= 119 p= 103	0
AFLP & Ca3 without intensities combined	n= 119 p= 103	0

The same analysis was performed using distance criteria under quartet puzzling. Quartet puzzle trees were reconstructed from the same data sets (Ca3 with and without intensities, AFLP, and combined AFLP & Ca3 with and without intensities) using distance criteria. Results are shown in Table 4-3.

Table 4-3: Partition metric tree comparison on distance quartet puzzling trees

data set	n= no. of sites p= no. of parsimony-informative sites	average difference
Ca3 data including intensities	n= 46 p= 37	2.66 +/- 2.3
Ca3 data without intensities	n= 46 p= 37	6 +/- 3
AFLP	n= 73 p= 66	1.33 +/-1.15
AFLP & Ca3 with intensities combined	n= 119 p= 103	5.3 +/- 3
AFLP & Ca3 without intensities combined	n= 119 p= 103	0

These results suggest that the AFLP plus Ca3 profiles (no intensities) gave data with better tree building qualities than did the Ca3 profiles alone, AFLP profiles alone or combined AFLP and Ca3 profiles which included Ca3 intensities.

Interestingly, although the results using distance criteria were similar to those obtained for quartet puzzling under parsimony, the AFLP data appeared to show more internal incompatibility when profiles were coded as distances rather than examined as site patterns (as used in parsimony quartet puzzling).

The quartet puzzle trees made using parsimony and distance criteria were also compared. Results are shown in Table 4-4.

Table 4-4: Results of partition tree comparison metric as implemented in PAUP 4.0 using distance and parsimony criteria.

data set	n= no. of sites p= no. of parsimony-informative sites	average symmetric-difference
Ca3 data including intensities	n= 46 p= 37	39.3 +/-1.1
Ca3 data without intensities	n= 46 p= 37	24 +/-3.46
AFLP	n= 73 p= 66	2.66 +/- 1.1
AFLP & Ca3 with intensities combined	n= 119 p= 103	24.6 +/- 1.1
AFLP & Ca3 without intensities combined	n= 119 p= 103	2

These results suggest that significant differences between reconstructed trees occurred when different tree building criteria were used on most data sets. However, this was not the case when AFLP and Ca3 (no intensities) were combined. This again favoured the combined use of AFLP and Ca3 (no intensities) data for reconstructing a phylogeny for the 36 strains of *C.albicans*.

4.2.1.2 An HSP Phylogeny From Combined AFLP and Ca3 Profiles

Quartet puzzle trees made using a parsimony criteria for combined data sets are shown in Figure 4-7. The quartet puzzle trees constructed using combined AFLP and Ca3 excluding band intensities (see Fig 4-7b) showed the same HSP group as a quartet puzzle tree constructed with AFLP and Ca3 with band intensities (see Fig 4-7a), and AFLP data alone (see Figure 4-5b).

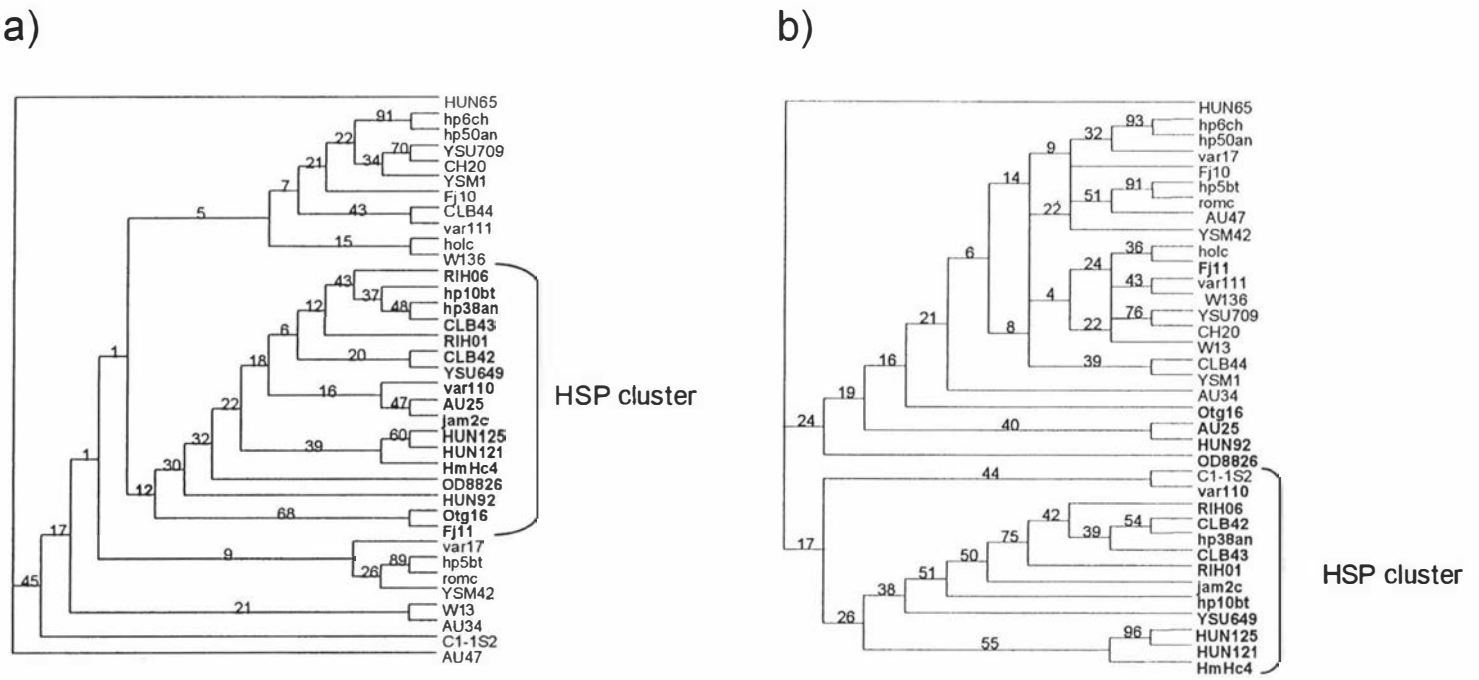


Figure 4-6: To show the effect of band intensity as a character in Ca3 profiles, quartet puzzle trees were made using parsimony criteria. (a) Quartet puzzle tree constructed using Ca3 scoring data in which band intensity was included as a character. (b) Quartet puzzle tree was built using Ca3 scoring data, in which band intensity was excluded as a character; thus only presence or absence bands were scored. Strains that were present in the HSP group when intensity was included (tree 4-6a) have been highlighted. Trees show significant differences and relatively low puzzle support values.

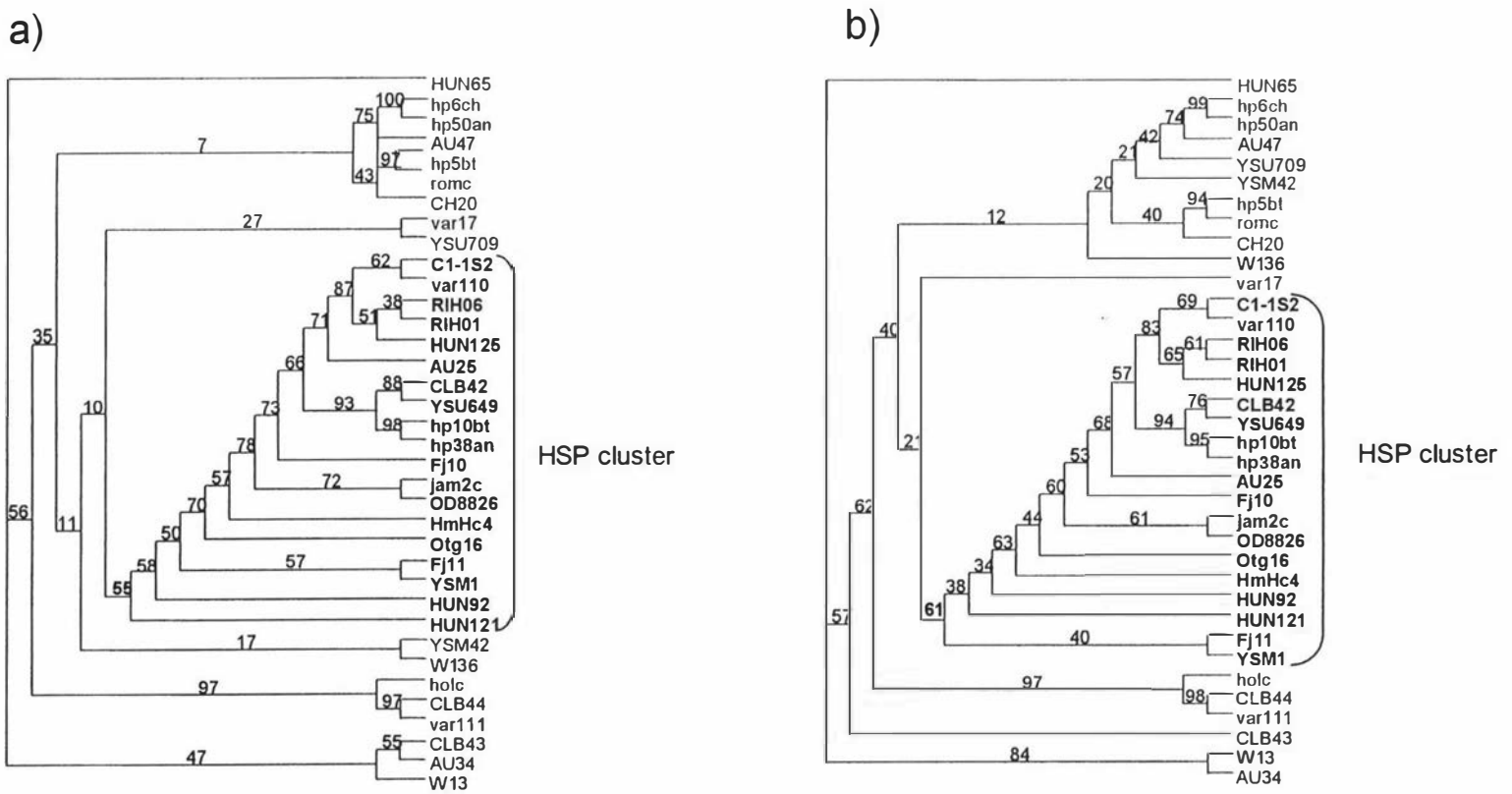


Figure 4-7: Quartet puzzle trees constructed with a) combined AFLP and Ca3 (including intensities) scoring data and b) combined AFLP and Ca3 (excluding intensities) scoring data using parsimony criteria or both trees. Both trees showed the same HSP group as a quartet puzzle tree constructed with AFLP data alone (see Figure 4-5b). The puzzling support of the HSP cluster in the puzzle tree constructed with the combined data set was 55% when Ca3 band intensities were included and 61% when intensities were excluded.

The puzzling support of the HSP cluster in the puzzle tree constructed with the combined data set (with Ca3 band intensities) was 55% (see Fig 4-7a) and 61% (see Fig 4-7b) when the Ca3 band intensities were excluded.

The HSP split occurred at 60.5 % when band intensities were excluded from the combined data set. The biggest incompatible split in respect of the HSP cluster split occurred at 5%. Most of the incompatible splits involved noncluster isolates var1.7, CLB43, and W136 that tended to group with HSP strains (see Appendix 14). When these strains were eliminated from the combined AFLP & Ca3 (no intensities) data set, the puzzling support value for the HSP cluster increased to 91% (data not shown). Therefore these strains were thought to be ambiguous in respect to the HSP cluster. However, this analysis suggests that the phylogenetic signal identifying this group is relatively strong given that the tree is reconstructed from restriction fragment data and not restriction site data (see discussion 4.2.2.2).

In summary, the results obtained from AFLP analysis confirm the existence of a closely related group of "HSP" strains. Using quartet puzzling as a tree building approach, the best estimate of HSP cluster appears to be derived from a combined data of AFLP and Ca3 (no intensities) data (see Figure 4-7b). In this tree the HSP cluster comprised the following strains: C1-1S2, var1.10, RIHO6, RIHO1, HUN125, CLB42, YSU649, hp10bt, hp38an, AU25, Fj10, jam2c, OD8826, Otg16, HmHc4, HUN92, HUN121, Fj11, and YSM1.

4.2.2 Discussion

4.2.2.1 Usefulness of AFLP

Preliminary AFLP analysis undertaken in the present study showed that AFLP profiles for *C.albicans* isolates were highly reproducible. This was demonstrated by obtaining highly similar banding patterns from different DNA extractions of the same strain, and by obtaining high similarity between banding patterns of HSP cluster strains (a group of closely related strains as suggested by earlier Ca3 studies, see section 1.4). No artefactual variation was found as reported for RAPDs (Ellsworth *et al.*, 1993). AFLP fingerprinting patterns differed between *C.albicans* isolates derived from different individuals, while profiles derived from isolates from the same individual were highly similar. Additionally, it was shown by analysis under split decomposition that AFLP profiles provided data with tree-like properties. These preliminary results indicated that AFLP profiles would be useful for studying intraspecific variation in *C.albicans* and for testing HSP phylogeny.

4.2.2.2. Assessment of HSP Phylogeny Using Ca3 and AFLP Fingerprinting Data Separately

The present study has found concordance between phylogenetic analysis of *C.albicans* isolates by Ca3 and AFLP fingerprinting. Quartet puzzle trees (using either distance or a parsimony criteria) that were built either with Ca3 or AFLP fingerprinting data showed a similar clustering of strains associated with the HSP phenotype. The phylogenetic signal identifying this group is relatively strong given that the tree was reconstructed from restriction fragment data and not restriction site data. The tree building properties of the former are known to be less desirable than the latter. The reason for this is that detected restriction fragments are not

necessarily independent, which may cause problems for the estimation of genetic or phylogenetic relationships among individuals (Tayler *et al.*, 1999). For example, in the case where a new site evolves between two preexisting sites, two new short fragments appear and one long fragment disappears. Thus, while two strains may share two of these three restriction sites there will be no indication of this from restriction fragment patterns. Whilst it was not a subject of study in the present thesis it is expected that these potential problems may have resulted in the relatively low support values for the HSP group.

Also, in the case of both restriction site and restriction fragment data, the loss of sites and fragments can occur in more ways than gain, and the processes that give rise to such events can differ (e.g., there can be loss of the recognition sequence by point mutation or by the result of deletion or intramolecular recombination etc.). Thus these properties may also contribute to poorer quality data for tree building.

Although both AFLP and Ca3 fingerprinting reveal restriction fragment polymorphisms it also has to be considered that AFLP produced a library of restriction fragments that may be more representative of the entire genome (Vos *et al.*, 1995), whereas Ca3 fingerprinting targeted only restriction fragments containing the repetitive sequence Ca3. Tayler *et al.* (1999) have noted that in fingerprinting methods with repetitive DNA sequences, like Ca3, the problems of identifying homology of positive alleles (presence of band) and of determining the significance of null alleles (absence of bands) are magnified due to the complex patterns obtained. If this is true for the Ca3 profiles studied here, it may be a reason for the few strain discrepancies between the HSP cluster in the Ca3 tree (i.e., CLB43, see Figure 4-5a) and the AFLP tree (e.g., C1-1S2, Fj10, and YSM1, see Figure 4-5b). This may have also contributed to the observed inconsistency (see Table 4-4) in the repeated reconstruction of quartet puzzle trees when built using Ca3 fingerprinting data.

Another possible explanation for the differences between HSP groups when analysed by Ca3 and AFLP separately may be the different rates of evolution for Ca3 patterns. Pujol *et al.* (1999) have suggested that the high molecular bands (higher than 7.9 kb) in a Ca3 fingerprinting pattern may represent microevolution within a strain and may therefore undergo genetic changes at a different (faster) rate than the remaining bands (Pujol *et al.*, 1999). If different parts of the Ca3 data matrix are evolving at different rates then this would be expected to affect phylogenetic estimates of the genetic distinctiveness between strains (Swofford *et al.*, 1996).

Despite such concerns, the results reported here are consistent with earlier findings that Ca3 fingerprinting is a method with considerable resolving power (Schröppel *et al.*, 1994; Lockhart *et al.*, 1995; Lockhart *et al.*, 1996). The concordance observed between Ca3 and AFLP profiles as to the existence of a similar cluster of HSP strains is also consistent with previously reported results obtained with methods such as RAPD (randomly amplified polymorphic DNA) and MLEE (multilocus enzyme electrophoresis) which also demonstrated high concordance with Ca3 fingerprinting results (Pujol *et al.*, 1997).

4.2.2.3 Combining Data in Analysis of the HSP Phylogeny

Different approaches are currently disputed for assessing the concordance between data sets (Huelsenbeck *et al.*, 1996). One is to compare the topological correspondence between results of separate analysis of data sets, as discussed in the last section. A totally different approach is the combination of different data sets, which is termed "total evidence" by Kluge (Kluge, 1989). In such cases, some authors have advocated the importance of first assessing the homogeneity of phylogenetic signal in the different data forms (Bull *et al.*, 1993; Huelsenbeck *et al.*, 1996). However, testing for this can be difficult to implement (Huelsenbeck *et*

al., 1996). In the present study the homogeneity of data was not assessed, however all data forms represented either restriction site or restriction fragment data. Thus, they were considered to be more or less similar, and combined in the present study. Joint Ca3 and AFLP data were shown by use of the partition metric to be internally consistent.

Using the phylogenetic framework provided by quartet puzzling tree building an evaluation could next be made of PCR markers that may be useful for rapid identification of HSP strains.

4.3. Identifying Molecular Markers Specific for HSP Strains

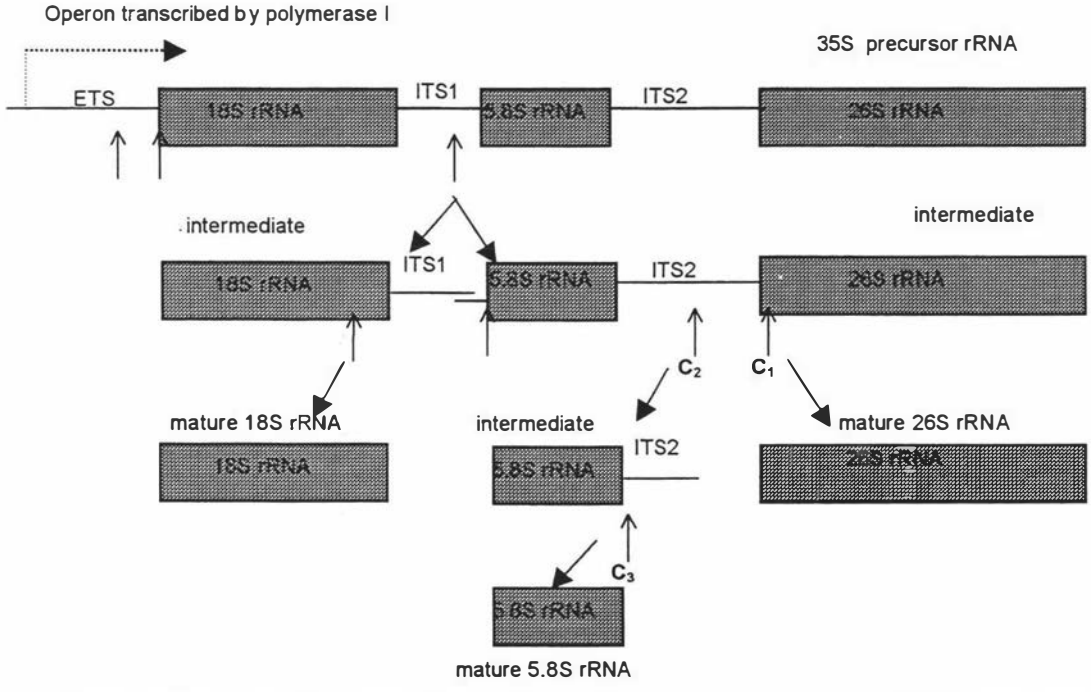
As discussed in an earlier part of this thesis, a present difficulty for implementing strain replacement studies for investigating a possible causal role of HSP strains in Candidosis is the absence of a rapid genetic screening procedure for identifying HSP strains. Sections 4.3.1 and 4.3.2 describe investigation of the nuclear ITS region and an AFLP derived region for the purpose of identifying molecular markers useful for rapid genetic screening.

4.3.1 Internal Transcribed Spacer Regions (ITS)

4.3.1.1 Introduction

Genes that encode nuclear ribosomal RNAs (rDNAs) have been used extensively to study genetic variations in fungi. Hence they were an obvious choice for study in the present investigation. The rRNA genes are evolutionarily highly conserved, whereas their spacer regions are more variable. In *C.albicans* they are found in a single tandem array of 50-100 copies per haploid genome of approximately 10 kb on chromosome seven (Magee *et al.*, 1987). Three of the rRNA genes, 18S, 5.8S and 26S rDNA, are lead by an external transcribed spacer

(ETS) and separated by two internal transcribed spacer (ITS1 and ITS2). These genes form part of a single operon, which is transcribed as 35S RNA by RNA polymerase I (Veldman *et al.*, 1981).



Three different cleavage sites (C1-C3) are present in the ITS2 sequence to correctly terminate the 5.8S and the beginning of the 28S subunits (Illustration 1; Yeh and Lee, 1990; van Nues *et al.*, 1995). Cleavage at sites C1 and C2 results in the mature 26S rRNA and an intermediate. Final cleavage at site C3 forms the mature 5.8S RNA (van Nues *et al.*, 1995). It is assumed that the secondary structure of the ITS2-intermediate plays a role in correct C1-C3 cleavage (Yeh and Lee, 1990; van Nues *et al.*, 1995). This intron is therefore called a "pseudointron". Based on mutation/deletion studies on *Saccharaomyces cerevisiae*, three essential domains (domains II, III, and V) that are highly conserved have been identified in the secondary structure of this ITS2 (van Nues *et al.*, 1995).

4.3.1.2 A Preliminary Study of Variability in the Nuclear ITS Region

Four representative strains were selected for a preliminary analysis to investigate the variability of ITS1 and ITS2 in *C.albicans* strains. These included two strains associated with the HSP cluster (RIHO1 and hp42bt, previously identified in an earlier Ca3 study) and two strains that were not part of the HSP cluster (noncluster strains hol-c and hp5bt).

Two primer pairs, shown in Figure 4-8a, were designed on the basis of published rDNA sequences of *C.albicans* (Mercure *et al.*, 1993). The two transcribed spacer regions, ITS1 and ITS2 (see section 4.3.1), were amplified by PCR (see section 2.5.7 for PCR protocol and conditions), purified (see section 2.5.3.3), and sequenced (see section 2.5.8.3). Primers pITS1 and pITS2 (see Table 2-7) amplified a 584 bp fragment containing spacer region ITS1, while primers pITS3 and pITS4 amplified a 441 bp fragment containing spacer region ITS2. The same primers were used to sequence both strands of amplified fragments.

ITS1 sequences from all four strains were 139 bp long and they were highly similar (data not shown). Sequences of ITS2, shown in Figure 4-8b, were 151 bp long and comparisons of the strains revealed an *Acc*1 restriction site present in the two noncluster strains, which was absent in the HSP strains.

4.3.1.3 *Acc*I RFLP Test on a Larger Number of *C. albicans* Isolates

This variation was used as basis for a restriction fragment length polymorphism (RFLP) study on a larger number of strains (essentially those strains suggested in the previous section). The ITS2 region was amplified from a total of 35 *C.albicans* strains by PCR (as shown in Figure 4-8a) and cleaved using restriction endonuclease *Acc*I.

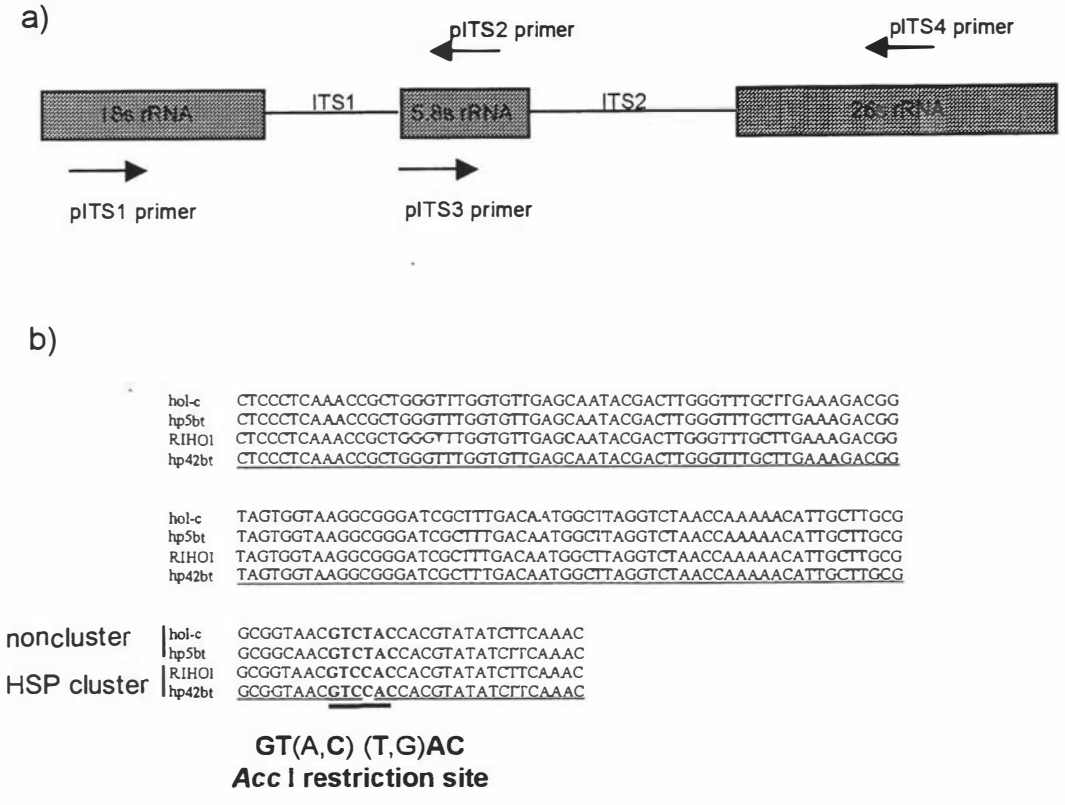


Figure 4-8a: Order of three rRNA genes and design of primers for the amplification of the internal transcribed spacer regions ITS1 and ITS2. PCR fragment containing the ITS1 spacer region was amplified using primers pITS1 and pITS2. The PCR fragment containing ITS2 was amplified using primers pITS3 and pITS4 (see Table 2-7).

Figure 4-8b: Alignment of ITS2 sequences from two HSP strains (hp42bt and RIHO1) and two noncluster strains (hol-c and hp5bt). A restriction site of endonuclease *Acc*I was present in the two noncluster strains and absent in the two HSP strains. All non spacer regions were eliminated for sequence alignments.

Sixty percent (21/35) of *C.albicans* strains were not cleaved by *Acc*1, and 40% (14/35) were digested by *Acc*I (see also Figures 4-9 and Figure 4-10).

To test whether the DNA templates from isolates not digested by *Acc*I could be cleaved by other restriction endonucleases, control digests were performed with *Sph*I. This enzyme was chosen since a cleavage site was present in the sequence of both the HSP and noncluster strains studied (this part of the sequence is not shown in Figure 4-8b). Examples of RFLP results from *Acc*I digests and control digests using *Sph*I are shown in Figure 4-9.

To evaluate the potential usefulness of ITS2 as a molecular marker for epidemiological studies, the results of RFLP analysis of HSP and noncluster ITS2 regions were superimposed onto the parsimony quartet puzzle tree made from a combined data set (AFLP & Ca3 without intensities) which is displayed in Figure 4-10. This comparison showed that most HSP strains were not cleaved by *Acc*I (exceptions being: HUN121, Fj11, and Otg16), while some noncluster strains were digested using this enzyme (exceptions being HUN65, hp6ch, hp50an, CH20, CLB43, and AU34).

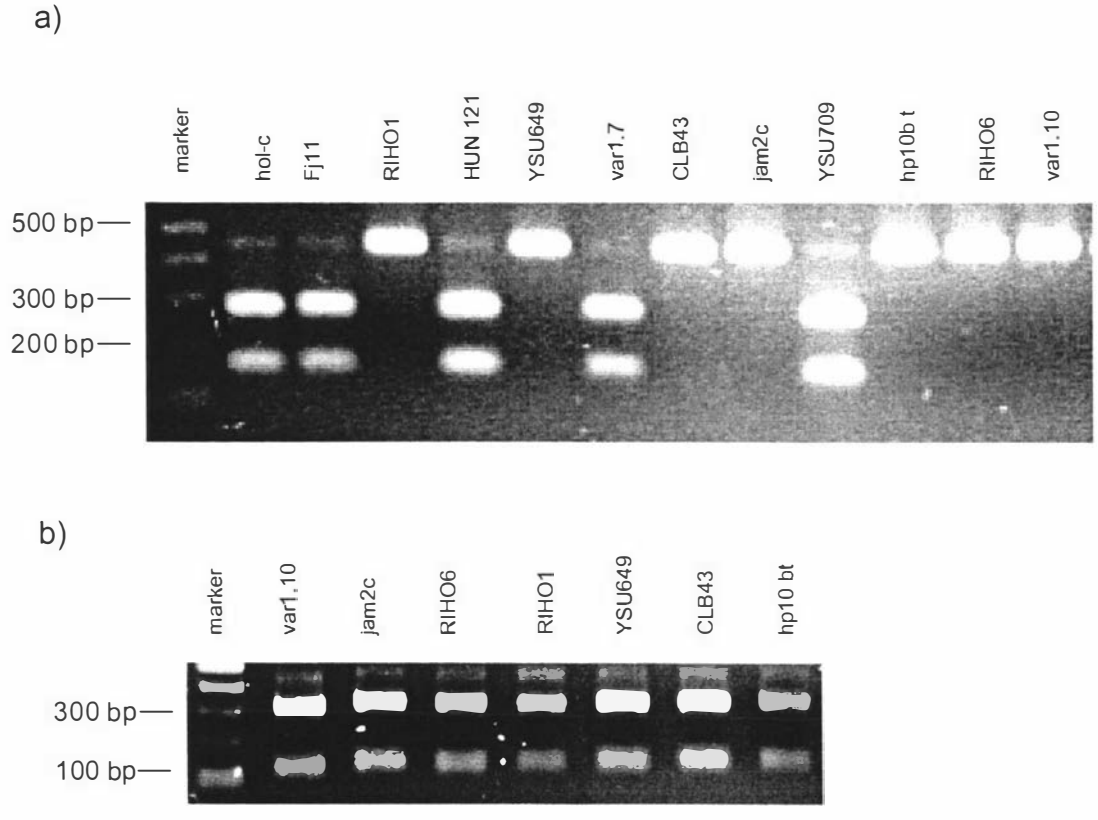


Figure 4-9a: Restriction fragment length polymorphism (RFLP) cleavage of amplified fragments containing ITS2 from different *C.albicans* strains by endonuclease *Acc*I .

Figure 4-9b: Control digest with *Sph*I was performed on strains not digested with *Acc*I.

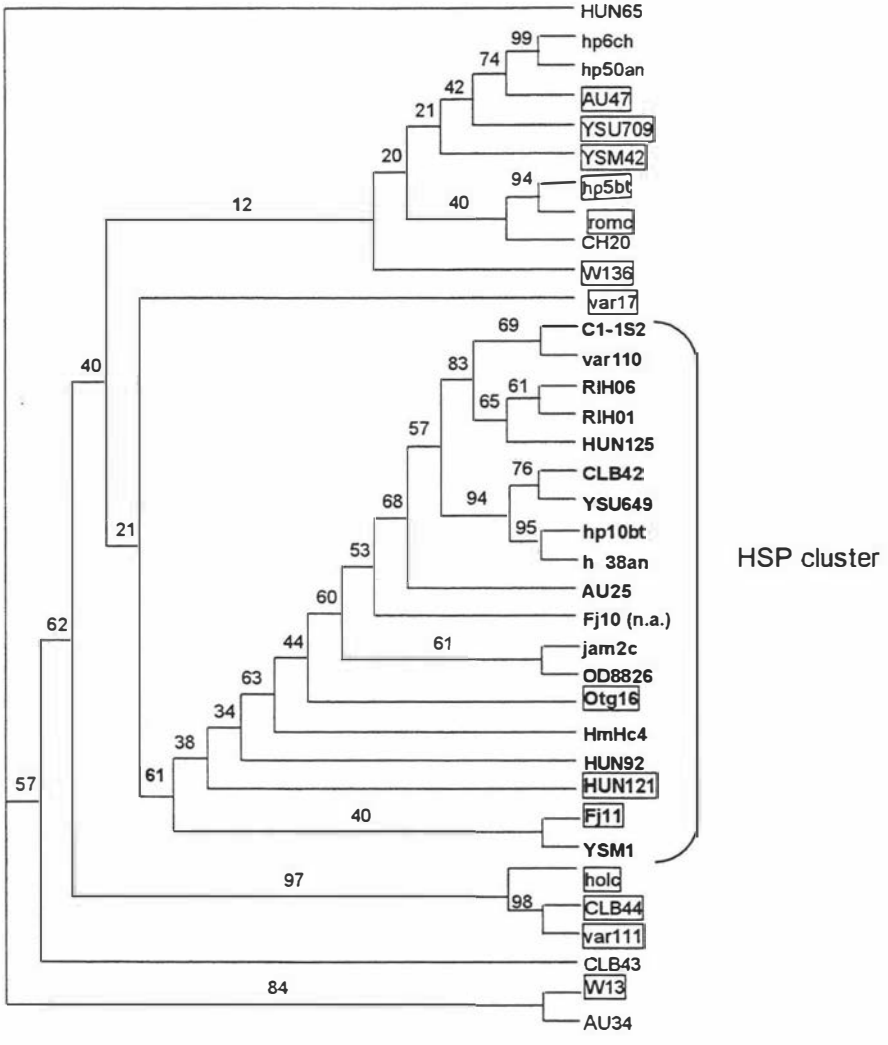


Figure 4-10: RFLP analysis results of the amplified ITS2 region using endonuclease *Acc*I were superimposed onto a Quartet Puzzle tree made from a combined data set (AFLP & Ca3 fingerprinting data excluding band intensities, see also Figure 4-7b). Strains that were cleaved by *Acc*I are marked with □. Most HSP strains (15/18) were not cleaved by *Acc*I, while most noncluster strains (11/17) were digested using this enzyme. (n.a. = not available)

4.3.1.4 Discussion

The nuclear ITS region has commonly been used to analyse intraspecific relationships (see section 4.3.1.1). In the present study it was investigated whether the region showed variation that may be used to develop a molecular marker for the rapid discrimination between HSP and noncluster strains (see section 1.4)

Other studies on *C.albicans* have reported a length of 137-139 bp depending on the strain, for ITS1 and 151 bp for ITS2 (Mercure *et al.*, 1993; Srikantha *et al.*, 1994). These findings stand in agreement with results obtained in the present study, in which ITS1 regions were 139 bp and ITS2 were 151 bp long.

However, unlike an earlier study, which examined a total of 10 *C.albicans* isolates (representing typical and morphologically (or physiologically) atypical strains), and found all *C.albicans* isolates to be identical at nucleotide level within both ITS regions (Lott *et al.*, 1998), in the present study minor sequence variations were found within ITS2 (see Figure 4-8b).

The spacer region ITS2 has been termed a "pseudointron" because it has been found that the secondary structure of the ITS2 RNA intermediate plays a role in splicing of rRNA genes (see section 4.3.1) (Yeh and Lee, 1990; Lott *et al.*, 1998). In *Saccharomyces cereviseae* this ITS2 RNA intermediate contains three essential conserved domains (II, III, and V) (Yeh and Lee, 1990; van Nues *et al.*, 1995). In the present study, none of the observed nucleotide exchanges occurred within these conserved regions. Thus it is expected that variations occur in a relatively fast evolving region of ITS2.

The observed sequence variations were the basis for an RFLP analysis using endonuclease *Acc*I. Since the *Acc*I restriction site is ambiguous, - GT (A,C) (T,G) AC -, it cannot be assumed that restriction sites within different strains are identical. Also, loss of an *Acc*I site can be due to different nucleotide exchanges. Thus, it is possible that some noncluster strains (e.g. CH20, HUN64, hp6ch, and hp50an; see Figure 4-10) that do not have an *Acc*I restriction site, have different sequence from those HSP strains which also do not have such a site. Sequencing of these strains would be needed to investigate this possibility. This ambiguity (and potential complexity) between HSP and noncluster strains suggests that cleavage of ITS fragments by *Acc*I cannot be used as a definite test for identifying HSP strains.

4.3.2 AFLP Derived Markers

4.3.2.1 Introduction

Previous studies have shown, that AFLP can be implemented to isolate and clone polymorphic bands, and subsequently to generate novel genetic markers (Lockhart and McLenachan, 1997; Lockhart *et al.*, 1999, McLenachan *et al.*, submitted). The AFLP profiles determined in the present work were examined for the presence of bands that might be unique to HSP strains. One band which appeared promising "P1CTC" has been partially characterised and examined for its distribution in *C.albicans* strains.

4.3.2.2 Analysis of AFLP Fragment P1CTC

The AFLP fragment P1CTC (Figure 4-7, AFLP profile *Mse*I-CTC), was found to correspond to a band present in most strains clustering with HSP isolates (17/19) in a quartet puzzle tree made from AFLP data (see Figure 4-5b) and also in a quartet puzzle tree that was made from AFLP data in which the P1CTC band

was excluded (data not shown). This band from the *Mse*I-CTC profile (see Figure 4-7, AFLP profile *Mse*I-CTC) was cloned (see section 2.5.8) and sequenced (see section 2.5.8.3). Although the extent of the difference between P1CTC loci in HSP and noncluster strains is still to be fully characterised at the sequence level, sections 4.3.2.2 - 4.3.2.6 describe progress made towards characterising the genetic basis of the difference between P1CTC sequences in cluster and noncluster strains. Sections 4.3.2.7-4.3.2.8 describes results concerning its potential as a marker for strain replacement studies.

(i) 5' region of P1CTC

After eliminating complementary nucleotides of *Mse*I linker sequence (see section 2.5.7.1), the P1CTC fragment was 541 bp long and had a GC content of 36%. A BLAST search, using default parameters, revealed a high percent sequence identity between the fragment and two previously characterised sequences. The first putative homology occurred between the 5' flanking region, from a gene encoding a high affinity corticosteroid binding protein gene (CBP1, (Malloy *et al.*, 1993)) of the GenBank database, and the first 65 bp of the AFLP fragment P1CTC. As displayed in Figure 4-11a, P1CTC revealed a 92% identity to 65 bp in the upstream region of CBP1, while the remaining 480 bp of P1CTC showed no similarity to CBP1 locus. To test for the potential significance of P1CTC sequence variants at the CBP1 locus, facing primers were designed in P1CTC and CBP1 (see Figure 4-11a).

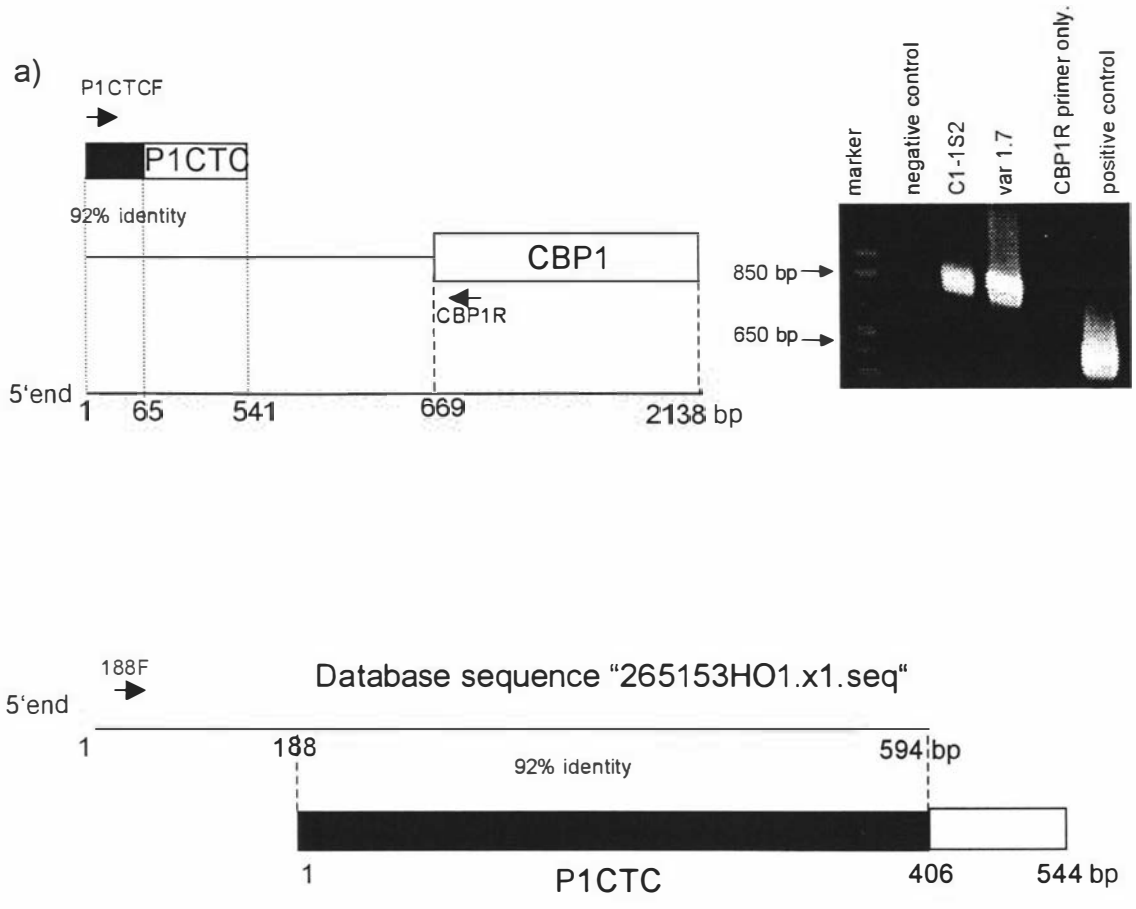


Figure 4-11a: Sequence identity (92%) between the first 60 bp of AFLP fragment P1CTC and the upstream region of a corticosteroid binding protein gene (CBP1). Primers P1CTCF and CBP1R were designed and this region amplified in one HSP (C1-1S2) and one noncluster strain (var1.7). Amplification in both strains yielded a strong approximately 800 bp large product.

Figure 4-11b: Sequence identity (92%) between the first 406 bp of P1CTC and an unpublished sequence termed "265153HO1.x1.seq" of the Stanford data base. Primer 188F was designed and used to amplify P1CTC and its flanking regions as described in section 4.3.2.9.3.

These primers amplified an 800 bp fragment in both an HSP (C1-1S2) and a noncluster strain (ambiguous noncluster strain var1.7, Figure 4-11b). Sequencing of these products confirmed that they were highly similar and that only the 5´region of P1CTC was located at the CBP1 locus. This result suggested that sequence differences in the region corresponding to the amplified fragment at the CBP1 locus could not account for the presence/absence differences observed in the AFLP profile for HSP and noncluster strains.

(ii) 3' region of P1CTC

Another high percentage sequence identity match (92%; 373/406 bp) for the 3'end of AFLP fragment P1CTC occurred with a sequence termed "265153H01.x1.seq" present in the Stanford database of unpublished *C.albicans* sequences. As shown in Figure 4-11b, a new primer (188F; see Table 2-7) was designed based on this sequence identity, and used to amplify a fragment comprising P1CTC and its 5' and 3' flanking regions as described in section 4.3.2.9.3 below.

4.3.2.3 Multilocus P1CTC-like Sequences in HSP and Noncluster Strains

To investigate the possibility that P1CTC related sequences exist in other parts of the *C. albicans* genome, Southern blots using P1CTC as a probe were carried out under moderately stringent conditions.

Figure 4-12a shows hybridisation results for *Mse*I digested genomic *C.albicans* DNA (see sections 2.5.1.1, 2.5.2, 2.5.6.5-7) and probed with the PCR fragment P1CTC under moderate stringency conditions. Individual banding patterns comprised 1-4 bands, with band sizes of >500 bp, 500 bp, 250 bp and <250 bp. A strong 500 bp band was present in HSP strains (CLB42, HUN125, HUN92), and absent in noncluster strains (var1.7, HUN65, CH20) investigated.

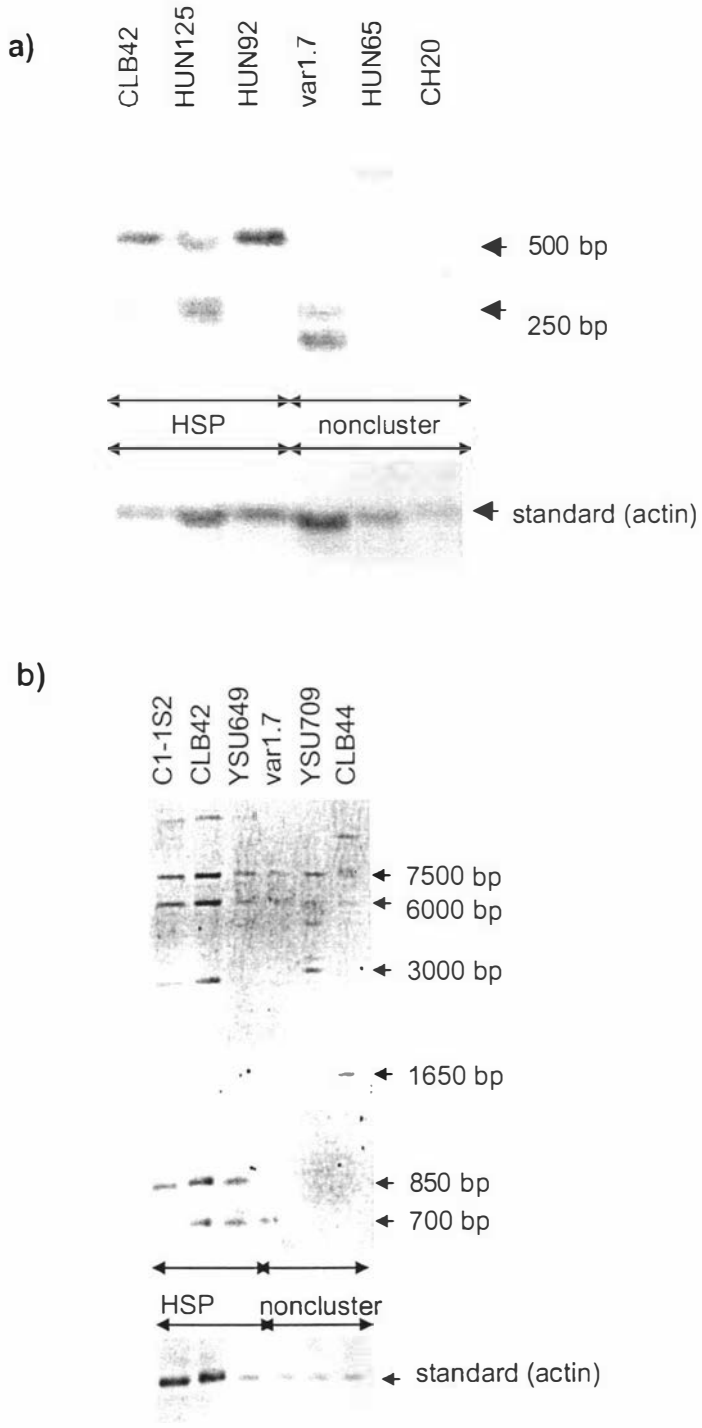


Figure 4-12a-b: Southern blot hybridisation experiments were performed to test for the presence of P1CTC or related sequences in HSP and noncluster strains. Hybridisations of a) *Mse*I and b) *Eco*RI digested genomic *C.albicans* DNA were performed under moderate stringency conditions using P1CTC as probe (top). As controls, blots were stripped and reprobed with part of the single copy gene actin (bottom).

However, a faint band with similar size was observed in the ambiguous (see section 4.2.1.2) noncluster strain var1.7. The 250 bp band was present in most strains, while bands >500 bp and <250 bp were only present in a few strains.

Figure 4-12b shows examples of the hybridisation results, when genomic DNA was restricted with *Eco*RI (see section 2.5.2) and probed using the PCR fragment P1CTC (see section 2.5.7.1) under moderate stringency conditions (see section 2.5.6.6). These conditions resulted in profiles of 4-6 bands, as shown in Figure 4-12b. Band sizes ranged between >12000 bp and 700 bp. An 850 bp band was exclusively present in isolates associated with the HSP phenotype (C1-1S2, CLB42, and YSU649), while a 700 bp band was present in HSP and one of the ambiguous noncluster isolates (var1.7, see section 4.2.1.2). Banding patterns in noncluster isolates were considerably variable, while the patterns for HSP isolates were highly similar except for the presence and absence of one band (3000 bp band in isolate YSU649).

Both Southern blots, *Eco*RI and *Mse*I, were stripped (see section 2.5.6.8) and reprobed with part of the single copy gene actin (see Table 2-7) as controls for complete digestion of genomic DNA (see Figure 4-12).

Since these results indicated that P1CTC like sequences occurred in multiple copies in the *C.albicans* genome, efforts were made to try and locate the specific loci at which the P1CTC AFLP gel polymorphism differed between HSP and noncluster strains. This was done by seeking to obtain further information on regions flanking the 541 bp P1CTC fragment.

4.3.2.4 Obtaining Flanking Regions of P1CTC

Two different approaches were employed to obtain the flanking regions of AFLP fragment P1CTC. The first approach attempted to obtain flanking regions

using ligation-mediated asymmetrical PCR, a method that is based on the concept of AFLP, but employing an asymmetrical ratio of a P1CTC-specific primer and one AFLP primer (see below) to obtain either the 5' or the 3' flanking region. The second attempt was made using inverse PCR, a method that allows both flanking regions in one reaction to be obtained.

(i) Use of Ligation-mediated Asymmetrical PCR to Obtain Flanking Regions of AFLP fragment

Using ligation-mediated asymmetrical PCR, it was attempted to obtain the 3' and 5' flanking region as displayed in Figure 4-13. *Eco*RI digested genomic *C.albicans* DNA (the same DNA samples that were previously tested in Southern blot hybridisations for complete digestions using part of the single copy gene actin as probe, see Table 2-7) linked with *Eco* linkers (see Table 2-7).

1) *Eco*R1 digest of genomic DNA

2) Ligation to *Eco* linker

3) Amplification by PCR using two primer combinations alternatively :

- **P1CTCF**/*Eco* primer
- **P1CTCR**/*Eco* primer

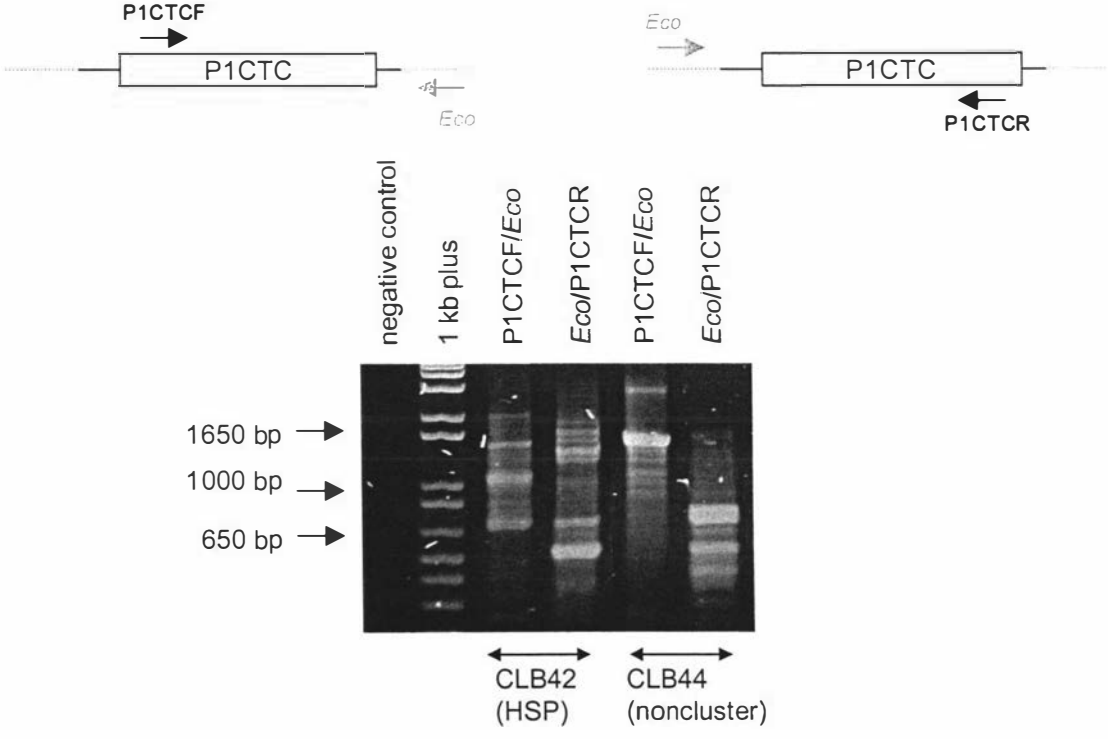


Figure 4-13: Process and amplification results of ligation-mediated asymmetrical PCR. *Eco*RI digested genomic DNA from one HSP strain (CLB42) and one noncluster strain (CLB44) were linked to *Eco* linkers and asymmetrical PCR was perfomed using either P1CTCF/*Eco* or P1CTCR/*Eco* primers in a ratio of 1:5.

To compensate for the excess of *Eco* priming sites in the genomic DNA, *Eco* primers (see Table 2-7) were added to the PCR reaction (see Table 2-8 for PCR protocol) in asymmetric ratios (10:1, 5:1). The concept for this experiment and the amplification results obtained using a primer ratio of 5:1 (*Eco* primer: AFLP primer) are displayed in Figure 4-13. Amplification products were isolated (see section 2.5.3.4), cloned (see section 2.5.8), and sequenced (see section 2.5.8.3). Most cloned fragments derived from only the *Eco* primer (*Eco* - *Eco* fragments). However, one PCR fragment of approximately 750 bp contained the P1CTC fragment and its 3' flanking region.

(ii) The Use of Inverse PCR to Obtain Flanking Regions

The process of inverse PCR (see section 2.5.7) is illustrated in Figure 4-14. This technique allows isolation of both flanking regions in one reaction. Genomic DNA was digested with *Eco*RI (as described above and see section 2.5.2). This choice of enzyme was made because no *Eco*RI restriction site was present within fragment P1CTC. Restriction fragments were religated to produce monomeric circles. PCR amplifications were performed using primers that were designed to the ends of P1CTC (primers P1 and P3, see Table 2-7 and Figure 4-16).

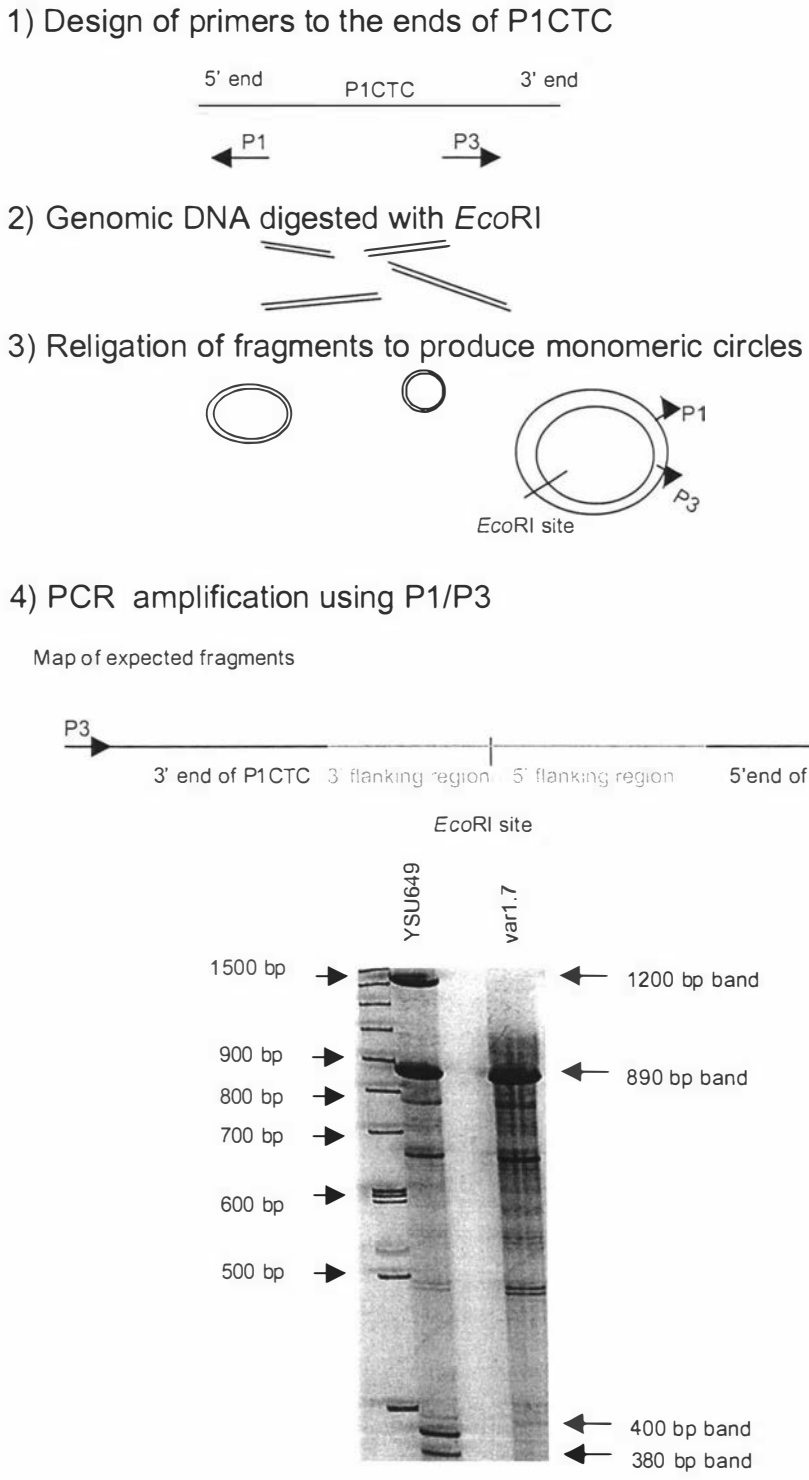


Figure 4-14: Inverse PCR process and amplification products. Amplification with primers P1 and P3 in one HSP (YSU649) and one noncluster (var1.7) isolate yielded in 1200 bp, 890 bp, 680 bp , 500 bp, 400 bp, and 380 bp products. Most bands were successfully isolated, reamplified, and sequenced.

The amplification with primers P1 and P3 (see Table-2-7) produced up to eight different products (1200 bp, 890 bp, 780 bp, 670 bp, 530 bp, 500 bp, 400 bp, and 380 bp) when resolved on 5% polyacrylamide gels (Figure 4-14). Four PCR products (1200 bp, 890 bp, 400 bp, and 380 bp) were successfully excised from polyacrylamide gel, cleaned (see section 2.5.3.4), reamplified (see section 2.5.7), and sequenced (see section 2.5.8.3). Sequencing revealed that products 400 and 380 bp resulted from amplification with only primer P3 (P3-P3 fragments). The sequence of the 890 bp fragment appeared unrelated to P1CTC. However, a 1200 bp large product (P3-P1 fragment) contained the flanking regions (see expected fragments in Figure 4-14) of a P1CTC region. The 5' end of this fragment showed high sequence identity to sequence "265153H01.x1.seq" mentioned above, whereas the 3'end showed high sequence identity to a 729 bp fragment previously obtained by ligation-mediated asymmetrical PCR (see above). Based on the sequence of the 3' flanking region, a new primer 3'endR (see Figure 4-16) was designed and used to amplifiy a product containing P1CTC and its flanking regions.

4.3.2.5 Amplification of P1CTC and its Flanking Regions

Two new primers, 188F (see Figure 4-11b; Figure 4-16, and Table 2-7) and 3'endR (see Figure 4-16 and Table 2-7), were designed in the flanking regions of P1CTC to amplify both HSP and noncluster strains. PCR amplification resulted in a strong product of approximately 800 bp in HSP strains as shown in Figure 4-15a. However, only very weak amplification levels of a similar sized product were present in some noncluster strains in combination with other products (no similar sized product was amplified in strain var1.7). Changes to the reaction conditions (e.g., concentration of primer and nucleotides, and different annealing temperatures) failed to increase specific amplification of this fragment in noncluster strains (data not shown). Various organic compounds (1.25%, 5% DMSO, Q- and enhancer solution) were also added to PCR reactions.

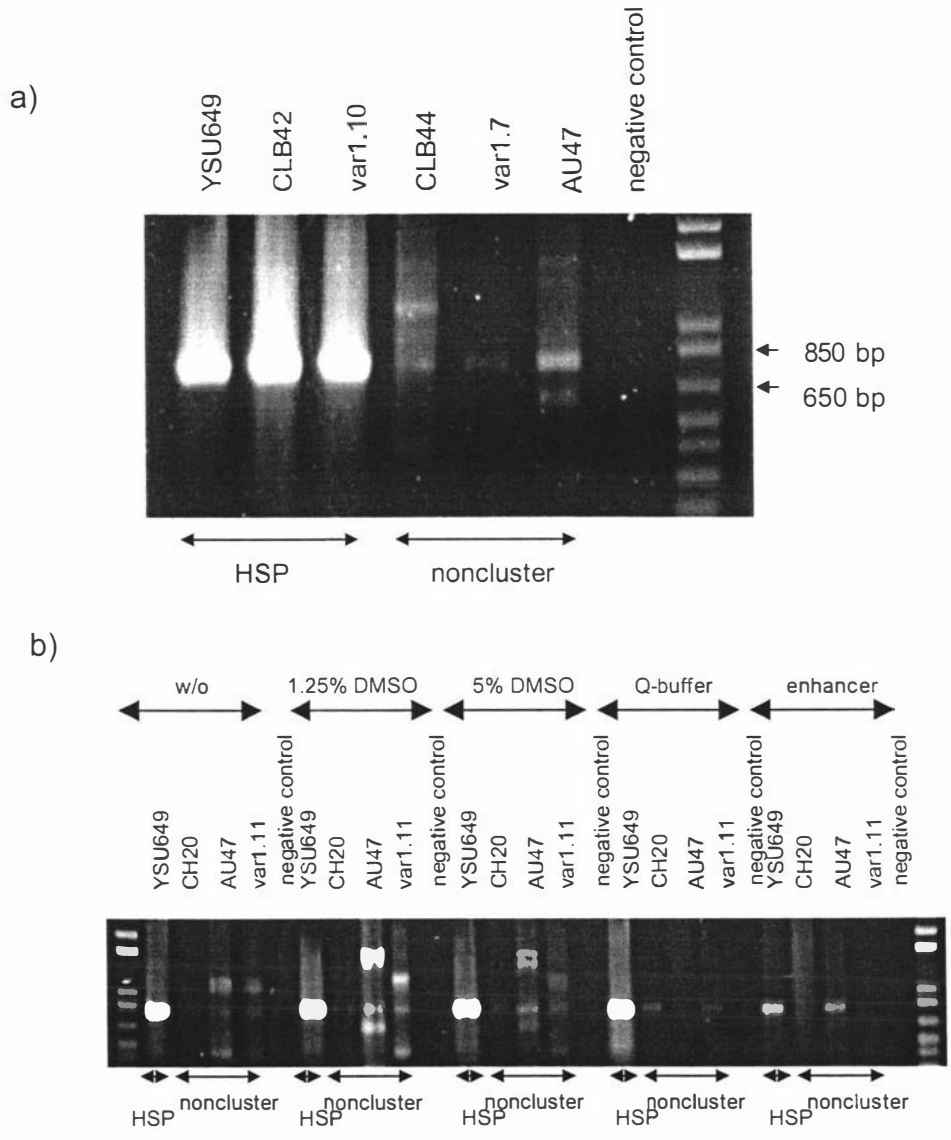


Figure 4-15a: PCR amplification of P1CTC-plus in three HSP strains (YSU649, CLB42, var1.10) and three noncluster isolates (CLB44, var1.7, AU47).

Figure 4-15b: PCR amplification of P1CTC-plus in one HSP strain (YSU649) and three noncluster isolates (CLB44, var1.7, AU47) with various additives (1.25% and 5% DMSO, Q-solution, and enhancer).

Figure 4-15b shows that the addition of DMSO (1.25% and 5%) resulted in more intense nonspecific products, while Q-buffer and enhancer solution decreased nonspecific amplification in noncluster strains.

To test for sequence variation within this larger amplification product, the fragments from different HSP *C.albicans* strains have been cloned (see section 2.5.8) and sequenced (see section 2.5.8.3). All HSP strains showed high sequence identity (data not shown).

The sequence of the 792 bp PCR product (comprising P1CTC and flanking regions) amplified in strains associated with the HSP phenotype is shown in Figure 4-16. A BLAST search using the default parameter in the Stanford database of unpublished *C.albicans* sequences revealed high sequence percentage identity to the same sequences as the smaller (541 bp) P1CTC fragment (see Figure 4-11). Identity between the 5' flanking region of P1CTC and the above mentioned unpublished sequence termed "265153H01.x1.seq" was 94% (491/517).

→ 188F
TGTTGTCTTC TCCCATTCTA CCTTGTGGGG GGTACGACAA AGACTAGCAA

CCCGGAGGCG CAGCCAATGA ACACAAGTCG GAGGCTTCAC CTGATCATTC
→ P1CTCF ← P1
GTTGTTAACT CGCTGGAGGC CATTTGCCGT TGCAACTACA TGCAGTACCA

ACGTCACCAT CGTAACAAAA TTACAAAAAA TATTAGTTTG TACTACAAAG

GGCACTACAA TTATGAATTG GCAGAATATA CATAAACAAC CAACTTAGCC

AATCAGCTAA TCCATTTACT GGATGAAAAC AGGAATATAG TTGTATTTAG

GTGGATACTC TTTTAGACAT ACAAACAAAG CGTGGTTTGA TAAATTGTGT

TTCGATATTT TTCATTTTGT TTCAATTGTA ATCAATTTTT CTCTTATTTT

TGCAACCGCT CTGAAATCCA AATTTTATTC CTTTTGAGGA ACTTTTATAT

TCAATAGTTG CAATACAACC ACTTTCATTA TGAATTACAT TACTTCGTTA
→ P3
TTGTTGCTCT CAAGTAATAC TTTCTTACAT CCAACAACAA CCTACTTACA

ATCTATGAAT GATTCAATTG TTCTTGTGAC ATCTTCAGTC TCCACAGAAT
← P1CTCR
TGACAGCACT TGCTTATGAT CCC ATATCAA CAATTTCTGG AGTTAACGGC

ACAAACAATA TCGACTATAT AAAGTTGTTA AACGATACTA ATTCCACGTT

TGTTCAATTG GATAATTCCG ATACAGATAT CGACGATAGC TCATCCAATT

CTGAAGATGT CTCATTCCAA TGATGAGCAA ATAGCACTAA TG
← 3'endR

Figure 4-16: Sequence of 792 bp large PCR fragment comprising P1CTC (bold) and its flanking regions. This sequence was obtained from an HSP associated strain YSU649. Sequences of primers used in this section (188F, P1CTCF, P1, P3, P1CTCR, and 3end'R; see Table 2-7) and their directions (←/→) with respect to present sequence are marked (underlined).

4.3.2.6 P1CTC is Not Related to the Ca3 Repetitive DNA

To test homology between the Ca3 probe and AFLP Fragment P1CTC, different amounts of the amplified P1CTC fragment (approximately 300 ng - 3 ng) were hybridised under low stringency conditions (see section 2.5.6.6) with Ca3 probe DNA. This Southern blot analysis showed no hybridisation signal (data not shown).

Two different controls were performed for these experiments. First, the Ca3 probe, consisting of Ca3 DNA and lambda phage DNA (see section 2.5.6.2), showed a positive hybridisation signal to the from lambda phage derived marker (1kb Plus ladder, Gibco BRL). Secondly, the same Ca3 probe, was successfully reused for Southern blot hybridisation of genomic *C.albicans* DNA (Ca3 fingerprints).

Additionally, the AFLP fragment P1CTC did not show sequence similarity to sequenced subfragments of the Ca3 probe in other studies (*Eco*RI fragments C1 (Lockhart *et al.*, 1995) and B (Anderson *et al.*, 1993)). These observations indicated that the locus P1CTC is not part of the moderately repetitive sequence Ca3.

4.3.2.7 Distribution of P1CTC as a Molecular Marker for HSP Strains

To examine the usefulness of P1CTC as a molecular marker for HSP cluster strains, the P1CTC marker was superimposed onto a tree constructed with a data set that used all available information that was independent of the P1CTC profile. (P1CTC was excluded from the data matrix). This reduced AFLP data set was combined with Ca3 fingerprinting data (without intensities), and also with the ITS2 restriction data (see section 4.3.1.3). Using this joint data set, quartet puzzle

trees were reconstructed using parsimony (see Figure 4-17) and distance criteria (data not shown).

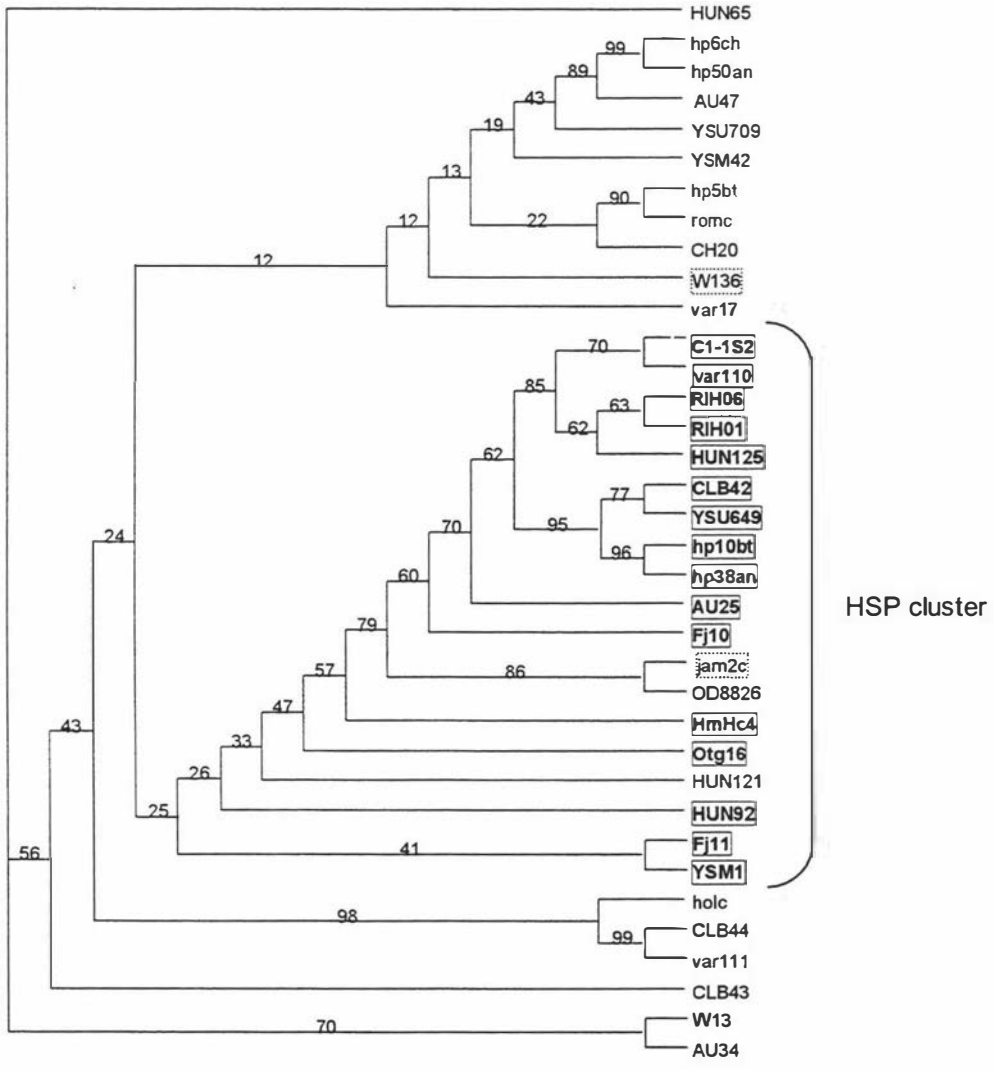


Figure 4-17: Quartet puzzle tree reconstructed from a joint data set combining all information available independent of the P1CTC profile. This included four AFLP profiles (without P1CTC bands), Ca3 fingerprinting data (without intensities), and the ITS2 restriction data. The puzzle tree was built using parsimony criteria. Onto this tree the presence of the P1CTC band was superimposed. ⬚ = faint P1CTC band, □= strong P1CTC band in AFLP profile *Mse*I-CTC.

The quartet puzzle tree using a parsimony criteria is displayed in Figure 4-17. Comparison of the quartet puzzle trees built using parsimony and distance criteria showed a tree to tree partition metric difference of 4. This result indicated that the trees were highly similar (only 4 internal branches differed out of a possible 66 i.e. 2 (n-3)). When superimposed onto these trees, P1CTC was found present in 17 out of 19 (89%) HSP strains. In 16 of these strains a strong amplification product was visible. In the putative HSP strain jam2c only a faint P1CTC band was amplified.

4.3.2.8 Development of a Multiplex PCR Assay for the Rapid Identification of HSP Strains Using AFLP Band P1CTC

Two internal PCR primers P1CTCF and P1CTCR (see Table 2-7) were designed based on the sequence of the P1CTC fragment. These primers were used to amplify a P1CTC fragment in 36 *C.albicans* strains. In 89% (17/19) of strains associated with the HSP phenotype (see section 4.2.1.2), a 514 bp P1CTC product P1CTC was amplified (only a faint band was amplified in HSP strain jam2c). This product did not amplify in noncluster strains. To test the quality of all DNA templates, an 838 bp large fragment of the single copy gene actin was amplified in the same reaction using ana and ans primers (see Table 2-7). Results of this multiplex PCR are shown in Figure 4-18.

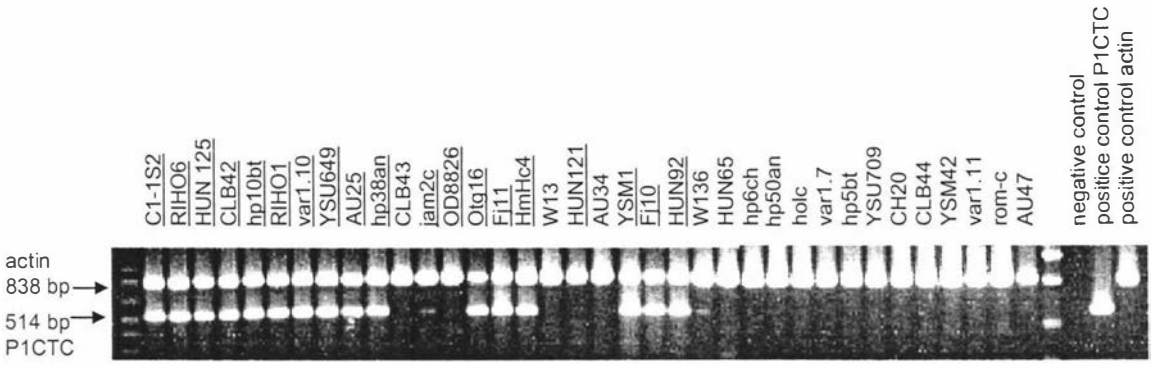


Figure 4-18: Multiplex PCR performed with two primer pairs simultaneously in 36 *C.albicans* isolates. The primer pair for P1CTC amplified a 514 bp fragment in HSP associated strains (C1-1S2 - hp38an, Otg16 - HmHc4, YSM1 - HUN92; underlined; see section 4.2.1.2). A faint 514 bp product was present in isolates jam2c and W136, while it was not amplified in the other isolates (CLB43, OD8826, W13 - AU34, and HUN65 – AU47). The positive control amplification of a single copy actin gene fragment (838 bp) was successful with all isolates.

The actin fragment was amplified in all strains investigated. When amplification results of the 514 bp fragment P1CTC were superimposed onto a tree reconstructed with all available data independent from P1CTC, the result was similar to the tree shown in Figure 4-17. These results suggest that a simple PCR test is possible for detecting HSP strains.

4.3.2.9 Discussion

4.3.2.9.1 Characterisation of Locus P1CTC

Sequencing of the AFLP fragment P1CTC revealed a GC content of 36%, which compares to that of other studies (32-40%, Kamiyama *et al.*, 1989, Iwagushi *et al.*, 1992). P1CTC has shown putative sequence homology to (i) the unpublished sequence "265253HO1.x1.seq." (Stanford database of unpublished microbial sequences), of which the function is not known to date, and (ii) the upstream region of a high affinity corticosteroid binding protein gene (CBP1).

The latter observation initially raised the possibility that at least part of the P1CTC fragment may be involved in the regulation of the CBP1 gene and thus, may represent an upstream regulatory sequence (URS element), which can be found up to 1.5 kb upstream from the coding region of a gene. Deletion or even point mutation in URS elements can greatly affect transcription levels (Mellor, 1993). Further, the affinity of corticosterone and progesterone, which are influenced by the gene CBP1, have also been suggested as being important in pathogenicity of *C.albicans* (Malloy *et al.*, 1993). However, sequence characterisation of the upstream region of CBP1 and also of AFLP fragment P1CTC (and its flanking regions) showed that the AFLP polymorphism distinguishing HSP and noncluster strains does not occur at the CBP1 locus. The exact location of this polymorphism still remains unclear as does whether it has any potential role in gene regulation.

4.3.2.9.2 Multilocus P1CTC-like Sequences in HSP and Noncluster Strains

Because the AFLP fragment P1CTC contains neither an *Mse*I nor an *Eco*RI restriction site, the number of hybridisation signals suggests multiple copies of P1CTC or related fragments in the *C.albicans* genome. Southern blot analysis

(see Fig 4-12) suggests 1-4 copies of P1CTC or P1CTC-like fragments in *Mse*I digests and 4-6 copies in *Eco*RI digests in the *C.albicans* genome. All P1CTC positive isolates showed a strong signal for an approximately 500 bp *Mse*I restriction fragment. This fragment was about the same size as the sequenced P1CTC fragment.

Two possible explanations may account for the higher number of hybridisation signals on *Eco*RI blots in comparison to *Mse*I blots. First, the additional signals may come from different-sized *Eco*RI fragments containing identical *Mse*I fragments, which are detected as only one band on *Mse*I blots. Second, *Eco*RI fragments may contain numerous *Mse*I restriction sites, which, in *Mse*I digestion, may generate fragments that are too small for detection on nylon membrane (<50 bp; Sambrook *et al.*, 1989).

The exact copy number of P1CTC or of related DNA fragments could not be resolved and the number may differ in different strains. The high degree of RFLPs could also indicate that P1CTC represents a mobile genetic element (Adams and Oeller, 1986), e.g., a retrotransposon. It is interesting that McLenachan *et al.* 2000 in their study on plants also report identification of multicopy markers derived using AFLP and that one of these found also had a high % match to a truncated retrotransposon.

Several retrotransposon-related elements have been identified in *C.albicans* (Goodwin and Poulter, 1998). They usually consist of an internal domain which is flanked by two long terminal repeats (LTRs). Previously characterised LTRs are 280 bp (Matthews *et al.*, 1997), 388 bp (Chen and Fonzi, 1992) and 395 bp (Perreau *et al.*, 1997). Solo LTRs, formed following recombination between two LTRs of a retrotransposon, can be quite numerous. The transposition cycle involves reverse transcription of a genomic mRNA into a full-length double-stranded DNA copy, followed by the insertion of this DNA into a

new site within the host genome. Transposition characteristically results in 5 bp duplications of the host DNA flanking the inserted element (Goodwin and Poulter, 1998). In the sequence of P1CTC, the same 5 bp duplication is present as described in Goodwin *et. al* (Goodwin and Poulter, 1998) for the LTR element kappa. However, the highly conserved beginning or end of an LTR (TG and CA) was not found in combination with this or any other 5 bp duplication, and none of the other described conserved regions of a retrotransposon were found. Thus, it is suggested that the P1CTC locus comprises only a portion of a mobile element (e.g., retrotransposon).

4.3.2.9.3 P1CTC and Its Flanking Regions

To further investigate the genomic location of P1CTC, the flanking regions have been obtained. Based on the 5' and 3' region, new primers were designed and a larger 792 bp fragment was amplified. Interestingly, the phenomenon of strong amplification in HSP-associated strains and no or less intense amplification in noncluster strains, was observed for this fragment (see Figure 4-15). Although amplification results of this larger fragment further support the diversity between HSP and noncluster strains, the assumed restriction and or priming site variations that may provide an explanation for differential amplification of P1CTC in HSP and noncluster strains, have not been resolved. Further characterisation of additional regions flanking P1CTC may provide this information.

4.3.2.9.4 P1CTC as Molecular Marker Specific for HSP Strains

Southern blot hybridisations of P1CTC DNA using as a probe, the moderately repetitive sequence Ca3, have shown that P1CTC is not homologous to the Ca3 probe. Thus, it can be assumed that Ca3 DNA and the AFLP fragment P1CTC occur at independent loci. This suggstion makes the marker potentially

useful for future comparisons between Ca3 typed and P1CTC identified HSP strains.

On the basis of the observed polymorphism associated with the P1CTC band (presence and absence of P1CTC band in AFLP profile *Mse*I-CTC, see section 4.2.1), a multiplex PCR assay (see section 4.3.2.8) was developed to distinguish between HSP and noncluster strains. A strong 514 bp PCR product was exclusively amplified in HSP associated strains, whilst a positive control (part of the single copy gene actin) was amplified in all strains. The 514 bp band corresponding to P1CTC was absent in two strains associated with the HSP cluster as defined by tree building with combined data sets (see section 4.2.1.2). This means that approximately 90% of HSP strains can be identified by the presence of this amplification product.

CHAPTER 5 - CONCLUSIONS

The present study sought to investigate whether certain strains, termed HSP strains, could replace other strains and subsequently cause infection. For this purpose the *C.albicans* flora from patients was monitored before and after the impact of a factor that was thought to predispose the individual to acquire a Candidosis.

During the course of study none of the sampled individuals developed a Candidosis. Risk factors that were used for categorizing differently predisposed patients were based on statistical data obtained from individuals who developed an infection. Without the development of any symptoms for Candidosis it was difficult to determine whether the individuals investigated actually became predisposed, which may be an essential condition for strain replacement. To increase the chance for observation of the transition from healthy to pathogenic state, it is suggested that in future studies, monitoring would be best made on large numbers of patients at high risk to develop a Candidosis. Such individuals would include organ transplant patients, burn/trauma patients, cancer patients, and neonates.

An indication for strain replacement was observed in one patient at low risk to acquire Candidosis. However, we were cautious in interpreting this observation, since we also found that healthy individuals could be colonised with different commensal *C.albicans* strains and substrains within one body location. Therefore it may be possible that the potential replacing strain was already present on the patient. Results by Soll *et al.* (1994) also have suggested that different body locations can be colonised by distinct strains. These observations highlight the difficulty in evaluating evidence of replacement from limited sampling data.

Future studies on replacement need to sample larger numbers of isolates from one body location and to sample as many body locations as possible. Ca3 fingerprinting, although a highly discriminatory method, is likely to be too labor

intensive and time consuming to investigate the number of patients and isolates required in such a study. For this reason, more rapid methods of screening were investigated in the present study. In the second part of the project a multiplex-PCR test was developed and this may now provide a tool that can be employed in future studies for the rapid screening for strain replacement caused by HSP strains. If potential replacement is detected with rapid initial screening then a more detailed investigation of putative replacing strains could then be performed using the Ca3 technique. This would allow detailed characterization of potentially interesting strains to be made, without the need to type all isolates.

At the time of present study, genetic fingerprint data were not available for cancer patients, nor was there other indirect evidence for strain replacement in this patient group. Further, contrary to expectations, Ca3 fingerprint results obtained in the present study indicated that the cancer patients studied were colonized with more diverse strains than healthy individuals. Interpretation of this finding was that strain replacement had not occurred in individuals from the cancer group, since replacement is expected to reduce strain diversity (Schmid, 1993; Schmid *et al.,* 1995; Schmid *et al.,* 1999). Thus while cancer patients, may be a useful study group, the finding may also suggest other groups need to be tested for evaluation of strain replacement. For example, previous studies have provided indirect evidence for replacement in AIDS patients (Schmidt *et al.,* 1992), and these may be a suitable group for future studies on replacement.

In investigating potential sources of *C.albicans* infections, we have successfully isolated *C.albicans* from the direct environment of high risk patients, demonstrating the potential of the hospital environment as potential source for infection causing strains. In characterizing strains, a nonradioactive fingerprinting protocol was developed for a more convenient use of Ca3 fingerprinting.

Ca3 fingerprinting has been evaluated for its potential and limitations for testing the replacement hypothesis. By analysis under split decomposition it was shown that Ca3 fingerprinting does not always provide unambiguous phylogenetic

information. This is expected to be the case when comparisons are made between more divergent strains but less of a problem for more closely related strains.

Until now the existence of the HSP group was based entirely on Ca3 fingerprinting data. To test for the existence of this group we have analysed amplified fragment length polymorphisms (AFLP) of 36 *C.albicans* isolates from different geographical regions. Phylogenetic reconstruction from both data forms (Ca3 and AFLP data) were highly congruent and suggest a worldwide distribution of HSP strains (a closely related group of strains with certain physiological properties (Schmid *et al.,* 1995). Study of the tree building properties of AFLP and Ca3 data using Quartet Puzzling and a tree comparison metric showed that AFLP data were more treelike than the Ca3 data.

The results obtained in the present study indicate the usefulness of AFLP (i) to investigate intraspecific phylogenetic relationships within *C.albicans* and (ii) to identify strain specific makers. Future studies are required to investigate the epidemiology of HSP strains and to assess their significance in causing disease. For such studies a rapid method for the identification of HSP strains is needed. In the present study one potential AFLP derived marker was developed.

A combined molecular data set, using all information that was independent of AFLP fragment P1CTC, showed the usefulness of this marker "P1CTC" to identify HSP strains. Based on this marker, a PCR based assay was developed. In this multiplex assay, amplification of the actin gene was used as a control for determining the suitability of the DNA template for PCR. In this assay, while actin amplified in all strains, a 514 bp P1CTC fragment only successfully amplified in HSP strains. Phylogenetic analysis of such amplification results showed that this marker identified approximately 90% of HSP strains studied. It will be important to further evaluate P1CTC as a diagnostic marker for future epidemiology studies of HSP strains. Of great interest will be determining whether this marker shows the request sensitivity to detect strain replacement in individuals predisposed to Candidosis.

Attempts were made to characterise the amplification specific nature of P1CTC in HSP strains. Southern blot experiments revealed multiple copies of P1CTC-like fragments in HSP and noncluster strains. It was shown that a 500 bp *Mse*I-restriction fragment and an 850 bp *Eco*RI-fragment were exclusively present in HSP strains. Using inverse PCR and ligation-mediated asymmetrical PCR, the 3` and 5` flanking regions of P1CTC were obtained. However, it still remains unclear why HSP strains and noncluster strains show amplification differences at this locus. Thus this question, and that of the genomic location of P1CTC, remain for future studies.

APPENDICES

Appendix 1

a) Scoring data from Ca3 fingerprinting patterns; 9 *C.albicans* isolates from one criteria 1 (healthy) individual (A.26)

```
A.26.1  0100000001000000000003011003200003020030030 33
A.26.2  0000010001000000000200200102000002010020130 33
A.26.4  0000000001000000000002011002110002010020020 33
A.26.6  0000010001000000000200200102000020020120030 33
A.26.7  0100010201000000000100111002210002010020020 33
A.26.8  0000010001000000000200200102000002010120030 33
A.26.9  0000010001000000000200200102000020010120030 33
A.26.10 0200000001000000000002011002210002010020030 33
A.26.13 0100000001000000000002011002200002010020030 33
```

b) Matrix containing S_{AB} values for all pairwise comparisons between nine *C.albicans* isolates from one criteria 1 (healthy) individual (A.26) generated using the Dendron software package. S_{AB} values ranged between 0.642 and 1.0.

```
A.26.1
A.26.2      0.654
A.26.4      0.824 0.711
A.26.6      0.667 0.958 0.681
A.26.7      0.750 0.760 0.816 0.731
A.26.8      0.642 0.979 0.696 0.980 0.745
A.26.9      0.642 0.979 0.696 0.980 0.745 1.000
A.26.10     0.873 0.694 0.917 0.667 0.830 0.680 0.680
A.26.13     0.906 0.723 0.913 0.694 0.824 0.708 0.708 0.960
```

Appendix 2

a) Scoring data from Ca3 fingerprinting patterns; *C.albicans* isolates obtained from saliva (S) samples and rectal (R) swabs taken from a highly predisposed individual (C1). *C.albicans* clones were isolated from samples taken before (1) and after the first course of neutropenia-inducing chemotherapy treatment. At the same sampling time *C.albicans* clones were also taken from saliva samples from the patient's partner (FM1). Scoring data used to analyse the extent of genetic diversity between strains harboured at a single site (Figure 3-7) within an highly predisposed patient (C1) are underlined.

C1-1S1	2001020202000020101000300203000003020130030000
C1-1S2	2002000202000020100100300203000003020030030000
C1-1S3	0000001102000020100100300203000003020030030000
C1-1R1	0001010202000020000000300203000003020030030000
C1-1R2	0000020202000020100100300203000003020030030000
C1-1R3	0000020202000020100100300203000003020030030000
C1-4R1	0000020202000020101000300203000003020030030000
C1-4R2	0000020202000020100100300203000003020030030000
C1-4S1	2002020002000020101000300203000003020030030000
C1-4S2	2002020002000020101000300203000002020030030000
C1-4S3	0000020002000020100100300203000003020030030000
FM1-1S1	0020020002000020101000300203000003020030030000
FM1-1S2	0011000002000020100100300203000003020030030000
FM1-1S3	0000000002000020100100300203000003020030030000

b) Matrix containing S_{AB} values for all pairwise comparisons between *C.albicans* isolates from patient C1 and patient's partner FM1. S_{AB} values ranged between 0.69 and 1.0

```
FM1-1S1
FM1-1S4 0.893
FM1-1S3 0.889 0.962
FM1-4S1 0.862 0.857 0.889
FM1-4S2 0.840 0.792 0.783 0.760
C1-1S1  0.871 0.833 0.828 0.806 0.741
C1-1S2  0.800 0.897 0.893 0.800 0.692 0.906
C1-1S3  0.857 0.926 0.962 0.857 0.750 0.833 0.897
C1-1R1  0.857 0.889 0.885 0.821 0.750 0.900 0.897 0.889
C1-1R2  0.897 0.893 0.926 0.828 0.760 0.903 0.900 0.929 0.929
C1-1R3  0.897 0.893 0.926 0.828 0.760 0.903 0.900 0.929 0.929 1.000
C1-4R1  0.900 0.828 0.857 0.833 0.769 0.969 0.903 0.862 0.897 0.933 0.933
C1-4R2  0.897 0.893 0.926 0.828 0.760 0.903 0.900 0.929 0.929 1.000 1.000 0.933
C1-4S1  0.900 0.862 0.857 0.833 0.769 0.938 0.903 0.828 0.862 0.867 0.867 0.935 0.867
C1-4S2  0.881 0.842 0.836 0.814 0.784 0.921 0.885 0.807 0.842 0.847 0.847 0.918 0.847 0.984
C1-4S3  0.929 0.926 0.962 0.857 0.792 0.867 0.862 0.926 0.889 0.964 0.964 0.897 0.964 0.897 0.877
```

Appendix 3

a) Scoring data from Ca3 fingerprinting patterns; *C.albicans* isolates from criteria 1 (healthy) individuals

```
A.26.1  0100000001000000000003011003200003020003003033
CH2F1   0202001001000001002002200202000002020020003203
CH4F1   0201002222000120020002200202000001120030033332
CH6F1   0000101122002020010000200022001002020020003222
CH7F1   0101001002000001002002200202010002120030033332
CH13F1  0000001002000010010201220102000002020020003023
CH14F2  0000001002000002020003320213020102020022021 32
CH18F1  0020000001010100002002200102100002020020003203
CH19F1  0000021100200000200003300003000002020230030 23
CH20F1  0000000001010101012002200102200002020020003003
CH22F1  0200000002000011002022200102000002020120020023
CH23F1  0000001100200020020000200203000002020230030003
CH24F1  0100101002000020010222002022000002020230030 13
```

b) Matrix containing S_{AB} values for all pairwise comparisons between *C.albicans* isolates from criteria 1 (healthy) individuals generated using the Dendron software package. S_{AB} values ranged between 0.46 and 0.89.

```
A26.8
CH2F1       0.618
CH4F1       0.545 0.685
CH6F1       0.586 0.585 0.711
CH7F1       0.586 0.892 0.737 0.588
CH13F1      0.792 0.667 0.648 0.698 0.667
CH14F2      0.563 0.620 0.634 0.595 0.649 0.696
CH16F1      0.622 0.769 0.540 0.473 0.691 0.600 0.492
CH18F1      0.642 0.800 0.592 0.571 0.762 0.655 0.551 0.760
CH19F       0.655 0.585 0.553 0.559 0.588 0.635 0.595 0.509 0.571
CH20F5      0.654 0.780 0.571 0.548 0.742 0.702 0.588 0.735 0.877 0.581
CH22F       0.667 0.754 0.639 0.594 0.719 0.712 0.629 0.706 0.678 0.594 0.724
CH23F       0.630 0.623 0.639 0.625 0.625 0.678 0.571 0.549 0.576 0.750 0.621 0.567
CH24F       0.576 0.576 0.571 0.638 0.609 0.688 0.507 0.464 0.531 0.580 0.571 0.646 0.646
```

Appendix 4

Scoring data from Ca3 fingerprinting patterns; *C.albicans* isolates obtained from criteria 3-5 individuals (see section 2.4.1.3) before (F) and after predisposition to acquire Candidosis.

CH2F1	02020010010000010020022002020000202002003203
CH2L1	00220010010000010020022002020000202002003203
CH4F1	02010022220001200200022002020000112003003332
CH4L1	02010022220001200200022002020000112003003332
CH7F1	01010010020000010020022002020100021200300332
CH7L4	01010010020000010020022002020100021200300332
CH13F1	00000010020000100102012201020000202002003023
CH13L2	00001010020000100102012201020000202002003023
CH19F1	00000211002000002000033000030000202023003023
CH19L1	00000102001000001000022000020000202012003023
CH20F1	00000000010101010120022001022000020200200300
CH20L1	02000110000100000120022002030000202012003003
CH22F1	02000000020000110020222001020000202012002023
CH22L1	01001010020000200002002002020000202023003003
CH23F1	00000011002000200200002002030000202023003003
CH23L1	00001010020000200002002002020000202023003003
C3F1	00002011200200010020003000032000302003003033
C3L1	00002011200200010020003000032000302003003033
C7F1	00222221000030000300220000230000230200310303
C7L1	00222220000030000300220000230000230200310303
C4F1	02020200020020001001021011020000202002203033
C4L1	00220001020020001001021011020000202002203032
C8F1	00000010002000002020033002030000021222310301
C8L1	03001020002000002020033002030000021222310300
C15F1	00001000100100000002012001020000202102002001
C15L1	00000000100100000002012001020000202102002001
C16F1	00000010100000100010012001022000020201200221
C16L1	00000010100000100010012001022000020201200221

b) Matrix containing S_{AB} values for all pairwise comparisons between *C.albicans* isolates obtained from criteria 3 individuals CH22 and CH23 prior (F) and after (L) treatment.

```
CH22F
CH22L       0.667
CH23F       0.567 0.800
CH23L       0.667 0.967 0.800
```

Appendix 5

Matrix containing S_{AB} values for all pairwise comparisons between *C.albicans* isolates obtained from criteria 2, 4 and 5 individuals (cancer patients).

C4F1														
C1-1S1	0.551													
C3F1	0.479	0.559												
C7F1	0.634	0.456	0.568											
C8F	0.568	0.592	0.603	0.500										
C9F	0.635	0.600	0.484	0.466	0.554									
C10F	0.657	0.625	0.576	0.519	0.638	0.793								
C13F2	0.486	0.448	0.493	0.525	0.472	0.295	0.400							
C14F1	0.649	0.595	0.658	0.552	0.608	0.618	0.750	0.507						
C15F1	0.526	0.481	0.571	0.388	0.542	0.583	0.577	0.436	0.645					
C16F1	0.540	0.533	0.645	0.438	0.646	0.667	0.621	0.426	0.588	0.667				
C21F1	0.542	0.571	0.552	0.493	0.590	0.600	0.593	0.351	0.563	0.591	0.720			
C22F1	0.588	0.615	0.627	0.538	0.743	0.576	0.603	0.455	0.603	0.528	0.712	0.655		
C23F1	0.588	0.585	0.627	0.487	0.571	0.542	0.571	0.515	0.548	0.491	0.712	0.618	0.750	
C24F2	0.554	0.581	0.594	0.507	0.537	0.536	0.600	0.540	0.629	0.440	0.536	0.615	0.590	0.656

Appendix 6

Scoring data from preliminary AFLP analysis with 12 *C.albicans* isolates

RIH010
0000111111111111111001111111111111001111111101011111100001100100001011111111110111010111111111100
1101111111011011101110111111001101111111111101111
RIH06
0001111110111011110011110110011000111111101000111100001100100001011111101101110101111111100
1101111111011011010111011111001101---------101111
gee2c
0011111111111111110011111111001110011111111010001111100001100100001011101111101110101111111100
1101011111111011011110111111001101111111011101111
hp42bt
0000111111111111111001111011001100011111111010001111100001100100001011111111101110101111111100
1101111111111011111110111111001101111111111101111
hp55bt
0010111111111111110001111111001110011111111010001111100001100100001011101111100110111111111111
1101111111111001111111011011001101111111111101111
HmHc4
1011111110001111100011111111111100111111101000111100001100100000011111111101111101111111100
1101111111011001111111011011000100111111100101111
wo1c
0111110011011011110011110111111110011110000010000000011100001010001001101111110101111111100
1111111100100011111110111100110110--111111001111
hp15bt
1010111111010011110011111111110110111110111000011000000111101010011110111111111010111111100
1111110001111111101111111111010010011111011100111
HmHc12
0011110010000011101111111110011111110010000001111111111100000001111100111111110001111111100
1111111111111111101110111111010010011111011000111
svoboc
1010110010000001101111111110011000000000000100111111111100000001111100111111110001111111100
1111111111111111101110111011001101111111100010111
C1-1R2
1010111111111011110011111110011100111101101110111110000110010000101111100110111010111111100
1101111111011011111111011011001101111111111101111
C1-1S2
0010111111110011110011111110011100111111101001111100001100100001011111001101110101111111100
1101111111011011111111011011101101111111111101111

Appendix 7

Scoring data from Ca3 fingerprinting analysis (including intensities) of 36 *C.albicans* isolates

```
HUN65   300230003002030000030100300100101000010000000
hp6ch   002031003002030000022000100001101000021100002O
hp50an  002031002002030000022003000000110100021001000O
holc    302030003002030000030200120102000100030002000O
var17   302030003002020002220002001001002000100000010
hp5bt   300230102001030002013000300100002002000000011
YSU709  303031013021100100032001020001000010000100100O
CH20    303230013002000001000200020010000000011000000O
CLB44   302220003001010000020100101001100010010010000O
YSM42   300030003001010020223000220000101000020210000O
var111  302230003002100000030200121001100200020002000O
romc    300030101001030000122000200100010002000010000l
AU47    200230003002030000031000000001000000200000001
C1-1S2  000030003002030000030200300100102000020202000O
RIH06   102030003102030000020100200100002100020000000O
HUN125  000030003102020000020100300000001000010000000O
CLB42   202030003101020000020200200200002000020000000O
hp10bt  102030003130031000020100200200002001000001000O
RIH01   102030003102020000030100300100002000020020000O
var110  101030003102020000020100210100102000020202000O
YSU649  102030003102020000020200210201001000010000000O
AU25    202030003102000200030200210000102000022000000O
hp38an  101030003102030000020100200200002000020000000O
CLB43   101030003101030000020100200200010100100000000
jam2c   202030003102030000020200200200102000020000000O
OD8826  102030003002030000020100200200000000020000000O
Otg16   332030003012020000120200200201012000020010011O
Fj11    301030003002000000020200210101000000020010000O
HmHc4   300030003102020100020000200300001000020200000l
W13     300230003002020000020100020000010000000000000
HUN121  300030003101010000020100100000001000010000000O
AU34    300030003002030000020100220100100000000020000O
YSM1    202231000202020020022100202001000010001000000O
Fj10    302030003002000000022010200000200010100000000O
HUN92   202030003002200000000001100300001000012000000O
W136    202030203002000000220100220001000100020100200O
```

Appendix 8

Scoring data from Ca3 fingerprinting analysis (without intensities) of 36 *C.albicans* isolates

```
HUN65   10011000100101000001010010010010100001000000000
hp6ch   00101100100101000001100010000110100001110000010
hp50an  00101100100101000001100100000011010001100100000
holc    10101000100101000001010011010100010001000100000
var17   10101000100101000101100010010010010001000000010
hp5bt   10011010100101000101100010010000100100000000011
YSU709  10101101101110010001100101000100001000010010000
CH20    10111001100100000100010001001000000000011000000
CLB44   10111000100101000001010010100110001001001000000
YSM42   10001000100101001011100011000010100000101100000
var111  10111000100110000001010011100110010001000100000
romc    10001010100101000011100010010001000100010000100001
AU47    10011000100101000001100000000010000001000000001
C1-1S2  00001000100101000001010010010010100001010100000
RIH06   10101000110101000001010010010000110001000000000
HUN125  00001000110101000001010010000000100001000000000
CLB42   10101000110101000001010010010000100001000000000
hp10bt  10101000111001100001010010010000100100000100000
RIH01   10101000110101000001010010010000100001001000000
var110  10101000110101000001010011010010100001010100000
YSU649  10101000110101000001010011010100100001000000000
AU25    10101000110100010001010011000010100001100000000
hp38an  10101000110101000001010010010000100001000000000
CLB43   10101000110101000001010010010000101001000000000
jam2c   10101000110101000001010010010010100001000000000
OD8826  10101000100101000001010010010000000001000000000
Otg16   11101000101101000011010010010101100001001001l0
Fj11    10101000100100000001010011010100000001001000000
HmHc4   10001000110101010001000010010000100001010000001
W13     10011000100101000001010001000001000000000000000
HUN121  10001000110101000001010010000000100001000000000
AU34    10001000100101000001010011010010000000001000000
YSM1    10111100010101001001110010100100001000100000000
Fj10    10101000100100000001101010000010001010000000000
HUN92   10101000100110000000000110010000100001100000000
W136    10101010100100000011010011000100010001010010000
```

Appendix 9

Scoring data from four AFLP profiles (*Mse*I-CTC, -CAG, -CTG, -CAG) of 36 *C.albicans* isolates

```
HUN65   0110110000000001000110111111111011100000010000000110010111111C0110001011011
hp6ch   0110100000000010001100000000000010000011000010111001011110100001110000001
hp50an  0000100000000001000110000001110001000001100110011100001111010000111000000 1
holc    0110101011100010001100000011001100000011000000011110011111000000001011011
var17   1110100001100110001100111100111110000101000000111000010000100001110101011
hp5bt   0110100111100110001111001000101010000001000100111001010110100000000001011
YSU709  0111100110000110001100000100111110000001000000111000010000100001111100001
CH20    0010100110000010000101001000000010000001000100111001010000000110101001111
CLB44   0110101010000010001111000000001100000001000000111111011111000000010011011
YSM42   0110100100000010001101110000001110000001000000011000010000000000100100011
var111  0110100010000010001111000011101100000001000000111111001110010000011011011
romc    0110100110000010001100111000001110000001000100111001010110110000010001011
AU47    0001001000000000000110011000011110000011000001101000010010000000000000000
C1-1S2  0110100111100111110111000000111111000000100100001100111000000100111110111
RIH06   1110100111100111111111000000111111000000100100001100111000000100111110111
HUN125  1110100111100111111111000000111111000000100100001100111000000100111110111
CLB42   1110100111100111111111100111111100000001001000111001110000001001111000111
hp10bt  1110100101100111111111110011111100000001001000011001110000001001111100111
RIH01   1110100111100111111111000000111111000000100100011100111000000100111110111
var110  1110100111100111111111000000111111000000100100011100111000000100111110111
YSU649  1110100111100111111111110011111110000001001000111001110000001001111100111
AU25    1110100001100111111111000000111111000000100010011100111000000100111110111
hp38an  1110100001100111111111110011111110000001001000011001110000001001111100111
CLB43   0110110001100111011111110110111110000101100000011001001111010111011011011
jam2c   1110100111100111111111110011111110000001100110011100111110011001111100111
OD8826  1110100111111011111111110011111110010001001110111001111110010001111100011
Otg16   1110100011100010101111111011111111111001011101111001110000001001011000011
Fj11    0000100001100010111111001000111110000011000000011111110000001001110000011
HmHc4   1110100111100010011101110000101110000000100100001100111000000100000110011
W13     0110110101100010001110110100101110000001000000111001001111011111010011011
HUN121  1110100101100111011111110000001100000001000000010000011010010000101100011
AU34    0110110001100111111000110111101110000001000000111001001111010111011011011
YSM1    0001100111100111111111011100011111000000100000001100111000000100101100 1011
Fj10    1110100111100111111111000011101110000011001000011001110000001001111100111
HUN92   0110100001110010001101110011101110000001000100111001110000001000111100111
W136    0111100100000010011111111110001111111101101010111111111000000000010000011
```

Appendix 10

Combined scoring data from four AFLP profiles (*Mse*I-CTC, -CAG, -CTG, -CAG) and Ca3 fingerprinting data (including intensities) of 36 *C.albicans* isolates

HUN65
300230003002030000030100300100101000010000000001101100000000100011011111111011100000010000000110010111110001100011011011
hp6ch
002031003002030000022000100001101000021100002001101000000000100011000000000000100000110000101110010111101000011100000001
hp50an
002031002002030000022003000000110100021001000000001000000000100011000000111000100000110011001110000011110100001110000001
holc
302030003002030000030200120102000100030002000001101010111000100011000000110011000000110000000111100111110000000001011011
var17
302030003002020002022000200100100200010000001011101000011001100011001111001111100001010000001110000100010000111010101011
hp5bt
300230102001030002013000300100002002000000001101101001111001100011110010001010100000010001001102001020010300020130003001
11100101011010000000000001011
YSU709
303031013021100100032001020001000010000100100001111001100001100011000001001111100000010000001110000100001000011111100001
CH20
303230013002000001000200020010000000001100000000101001100000100001010010000000100000010000100111001010000000110101001111
CLB44
302220003001010000020100101001100010010010000001101010100000100011110000000011000000010000001111110111110000000010011011
YSM42
300030003001010020223000220000101000020210000001101001000000100011011100000011100000010000000110000100000000000100100011
var111
302230003002100000030200121001100200020002000001101000100000100011110000111011000000010000001111110011100100000011011011
romc
300030101001030000122000200100010002000010000101101001100000100011001110000011100000010000100110010101101100000010001011

AU47

2002300030020300000310000000001000000200000000100010010000000000001100110000111100000110000011

01000010010000000000000000

C1-1S2

0000300030020300000302003001001020000202020000011010011110011111011100000011111000000010010000

01100111000000100111110011

RIH06

1020300031020300000201002001000021000200000000111010011110011111111100000011111000000010010000

01100111000000100111110011

HUN125

0000300031020200000201003000000010000100000000111010011110011111111100000011111000000010010000

01100111000000100111110011

CLB42

2020300031010200000202002002000020000200000000111010011110011111111111001111110000000010010000

11100111000000100111100011

hp10bt

1020300031300310000201002002000020010000010000111010010110011111111111001111110000000010010000

01100111000000100111110011

RIH01

1020300031020200000301003001000020000200200000111010011110011111111100000011111000000010010000

11100111000000100111110011

var110

1010300031020200000201002101001020000202020000111010011110011111111100000011111000000010010000

11100111000000100111110011

YSU649

1020300031020200000202002102010010000100000000111010011110011111111111001111110000000010010000

11100111000000100111110011

AU25

2020300031020002000302002100001020000220000000111010000110011111111100000011111000000010001000

11100111000000100111110011

hp38an

1010300031020300000201002002000020000200000000111010000110011111111111001111110000000010010000

01100111000000100111110011

CLB43

1010300031010300000201002002000010100100000000011011000110011101111111011011111000010110000000

01100100111101011101101101

jam2c

2020300031020300000202002002001020000200000000111010011110011111111111001111110000000011001100

01110011111001100111110011

OD8826

1020300030020300000201002002000000000200000000111010011111101111111111001111110010001001110

11100111111001000111110001

Otg16

3320300030120200001202002002010120000200100110111010001110001010111111101111111111001011101

11100111000000100101100001

Fj11

3010300030020000000202002101010000000200100000000010000110001011111100100011111000001100000

01111111000000100111000001

```
300030003102020100020000200300001000020200000211101001111000100111011100001011100000010010000
011001110000001000001100011

W13
300230003002020000020100020000010000000000000001101101011000100011101101C010111000000010000000
111001001111011111010011011
HUN121
300030003101010000020100100000001000010000000011101001011001110111111100000011000000010000000
010000011010010000101100011
AU34
300030003002030000020100220100100000000002000000110110001100111111000110111101110000000100000
111001001111010111011011011
YSM1
202231000202020020022100202001000010001000000000011001111001111111110111000111110000001000000
011001110000001001011001011
Fj10
302030003002000000022010200000200010100000000011101001111001111111110000111011100000110010000
011001110000001001111100111
HUN92
202030003002200000000001100300001000012000000001101000011100100011011100111011100000010001000
111001110000001000111100111
W136
202030203002000000220100220001000100020100200001111001000000100111111111000111111111011010101
111111110000000000010000011
```

Appendix 12

Combined scoring data from four AFLP profiles (*Mse*I-CTC, -CAG, -CTG, -CAG) and Ca3 fingerprinting data (without intensities) of 36 *C.albicans* isolates

HUN65
10011000100101000001010010010010100001000000000110110000000010001101111111101110000001000000
01100101111100011000101101 1
hp6ch
00101100100101000001100010000110100001110000100110100000000010001100000000000010000011000010
11100101111010000111000000 1
hp50an
00101100100101000001100100000011010001100100000000100000000010001100000011100010000011001100
11100001111010000111000000 1
holc
10101000100101000001010011010100010001000100000110101011100010001100000011001100000011000000
01111001111100000000101101 1
var17
10101000100101000101100010010010010001000000101110100001100110001100111100111110000101000000
11100001000010000111010101 1
hp5bt
10011010100101000101100010010000100100000000110110100111100110001111100100010101000000100010 0
11100101011010000000000101 1
YSU709
10101101101110010001100101000100001000010010000111100110000110001100000100111110000001000000
11100001000010000111110000 1
CH20
10111001100100000100010001001000000000110000000010100110000010000101001000000010000001000100
11100101000000011010100111 1
CLB44
10111000100101000001010010100110001001001000000110101010000010001111100000000110000000100000 0
11111101111100000001001101 1
YSM42
10001000100101001011100011000010100001011000000110100100000010001101110000001110000001000000
01100001000000000100100011 1
var111
10111000100110000001010011100110010001000100000110100010000010001111000011101100000001000000
11111100111001000001101101 1
romc
10001010100101000011100010010001000100001000010111010011000001000110011100000111000000100010 0
11100101011011000001000101 1

AU47

10011000100101000001100000000010000001000000010001001000000000001100110000111100000110000011

0100000100100000000000000000

C1-1S2

0000100010010100000101001001001010000101010000011010011110011111011100000111110000001001000

0110011100000001001111100111

RIH06

101010001101010000010100100100001100010000000001110100111100111111111000000111110000001001000

0110011100000001001111100111

HUN125

000010001101010000010100100000001000010000000001110100111100111111111000000111110000001001000

0110011100000001001111100111

CLB42

101010001101010000010100100100001000010000000001110100111100111111111110011111100000001001000

1110011100000001001111000111

hp10bt

101010001110011000010100100100001001000000100001110100101100111111111110011111100000001001000

0110011100000001001111100111

RIH01

101010001101010000010100100100001000010010000001110100111100111111111000000111110000001001000

1110011100000001001111100111

var110

101010001101010000010100110100101000010101000001110100111100111111111000000111110000001001000

1110011100000001001111100111

YSU649

101010001101010000010100110101001000010000000001110100111100111111111110011111100000001001000

1110011100000001001111100111

AU25

101010001101000100010100110000101000011000000011101000011001111111110000001111100000001000100

1110011100000001001111100111

hp38an

101010001101010000010100100100001000010000000001110100001100111111111110011111100000001001000

0110011100000001001111100111

CLB43

101010001101010000010100100100001010010000000000110110001100111011111110110111110000101100000

011001001111010111011011011

jam2c

101010001101010000010100100100101000010000000001110100111100111111111110011111100000001100110

011100111110011001111100111

OD8826

101010001001010000010100100100000000001000000001110100111111011111111110011111110010001001110

111001111110010001111100011

Otg16

11101000101101000011010010010101100001001001101110100011100010101111110111111111111001011101

1110011100000001001011000011

Fj11

101010001001000000010100110101000000001001000000000100001100010111111001000111110000011000000

0111111100000001001110000011

HmHc4

100010001101010100010000100100001000010100000111101001111000100111011100001011100000010010000110011100000010000011100111

W13

100110001001010000010100010000010000000000000001101101011000100011101101001011100000010000001110010011110111110100110111

HUN121

100010001101010000010100100000001000010000000001110100101100111011111110000001100000001000000010000001101001000010110011

AU34

100010001001010000010100110100100000000010000001101100011001111110001101111011100000010000001110010011110101110110111

YSM1

101111000101010010011100101001000010001000000000011001111001111111110111000111110000001000000011001110000001001011001011

Fj10

101010001001000000011010100000100010100000000001110100111100111111111000011101110000011001000011001110000001001111100111

HUN92

101010001001100000000000110010000100001100000000110100001110010001101110011101110000001000100111001110000001000111100111

W136

101010101001000000110100110001000100010100100001111001000000100111111111000111111111011010101111111110000000000010000011

Appendix 13

Combined available information independent of P1CTC. Scoring data from four AFLP profiles (*Mse*I-CTC,- CAG, -CTG, -CAG) of which the band P1CTC was excluded, Ca3 fingerprinting data (without intensities), and ITS2 restriction data of 36 *C.albicans* isolates

HUN65
00110110000000001000110111111111011100000010000000110011111110001100010110111001100010010100000
10100100100101000100000000
hp6ch
00110100000000010001100000000000010000011000010111001111101000011100000010010110010010100000
11000100001101000011100001O
hp50an
00000100000000001000110000001110001000001100110011100011110100001110000001001011001001010000O
11001000000110100011001O000
holc
10110101011100010001100000011001100000011000000011111011111000000000101101110101000100101000O0
10100110101000100010001O000
var17
11110100001100110001100111100111111000010100000011100010000100001110101011101010001001010001O
11000100100100100010000001O
hp5bt
10110100111100110001111001000101010000001000100111001101101000000000001011100110101001010001O
11000100100001001000000011
YSU709
10111100110000110001100000100111110000001000000111000100001000011111000011010110110111001000
11001010001000010000100100O
CH20
00010100110000010000101001000000010000001000100111001100000001101010011111101110011001000001O
00100010010000000011000000
CLB44
10110101010000010001111000000001100000001000000111111111110000000100110111011100010010100000
10100101001100010010010000O
YSM42
10110100100000010001101110000001110000001000000011000100000000001001000111000100010010100101
11000110000101000010110000O
var111
10110100010000010001111000011101100000001000000111111011100100000110110111011100010011000000
10100111001100100010001O000

romc
1011010011000001000110011100000111000000100010011100110110110000010001011100010101001010000111000100100010000100001
AU47
0000100100000000000110011000011110000011000001101000000100000000000000000100110001001010000011000000000100000100000001
C1-1S2
0011010011110011111101110000000111110000001001000011001100000010011111001110000100010010100000101001001001010000101010000
RIH06
0111010011110011111111110000001111100000010010000110011000000100111110011110101000110101000001010010001100001000000000
HUN125
0111010011110011111111110000001111100000010010000110011000000100111110011110000100011010100000101001000000010000100000000
CLB42
0111010011110011111111111001111110000000100100011100110000001001111000111101010001101010000010100100010000100000000
hp10bt
0111010010110011111111111001111110000000100100011100110000001001111100111101010001111001100001010010000010010000010000
RIH01
0111010011110011111111110000001111100000010010000111001100000010011111001111010100011010100000101001000001000010010000
var110
0111010011110011111111110000001111100000010010000111001100000010011111001111010100011010100000101001101001010000101010000
YSU649
0111010011110011111111111001111111100000010010000111001100000010011111001111010100011010100000101001101001000010000000000
AU25
0111010000110011111111110000001111100000010001001110011000000100111110011110101000110100010000101001100001010000110000000
hp38an
0111010000110011111111111001111111100000010010000110011000000100111110011110101000110101000001010010000010000010000000000
CLB43
0011011000110011101111111011011111000010110000000110010111101011101101110101000110101000001010010000010100100000000000
jam2c
0111010011110011111111111001111111100000011001100111001111001100111110011110101000110101000001010010001010000100000000000
OD8826
0111010011111110111111111110011111111001000100111011100111111001000111110001110101000100101000001010010000000000100000000
Otg16
1111010000111000101011111110111111111111001011101111001100000010010110000111111010001011010000110100101011100001001100110
Fj11

```
1000010001100010111111001000111110000011000000011111100000010011100000111010100010010000000101001101010000000100100000
HmHc4
0111010011110001001110111000010111000000100100001100110000001000001100011100010001101010100010000100100001000010100000001
W13
1011011010110001000111011010010111000000100000011100101111011111010011011100110001001010000010100001000000100000000000000
HUN121
1111010010110011101111111000000110000000100000001000011010010000101100011100010001101010000010100001000000100000000000000
AU34
0011011000110011111100011011110111000000100000011100101111010111011011011100010001001010000010100111010010000000000100000
YSM1
0000110011110011111111101110001111100000010000000110011000000100101100101110111110001010100100111001010010000100010000000
Fj10
_1110100111100111111111000011101110000011001000011001100000010011111001111010100010010000000110101000001000101000000000
HUN92
0011010000111001000110111001110111000000100010011100110000001000111100111101010001001100000000001100100001000011000000000
W136
1011110010000001001111111110001111111110110101011111111100000000000100000111010101010010000001010011000100010001010010000
```

Appendix 14

Bipartitions and frequency of occurrence (puzzling support values) found in one quartet puzzle trees built using a combined data set (AFLP and Ca3 without intensities):

HSP strains were no.s 14-23, 25-29, 31, 33-35;

14 (C1-1S2), 20 (var1.10), 15 (RIHO6), 19 (RIHO1), 16 (HUN125), 17 (CLB42), 21 (YSU649), 18 (hp10bt), 23 (hp38an), 22 (AU25), 34 (Fj10), 25 (jam2c), 26 (OD8826), 27 (Otg16), 29 (HmHc4), 35 (HUN92), 31 (HUN121), 28 (Fj11), and 33 (YSM1).

ambiguous noncluster strains (strains that tended to group with the HSP cluster) were no.s: 5 (var1.7), 24 (CLB43), 36 (W136)

```
         1         2         3
12345678901234567890123456789012345678      Freq      %
-----------------------------------------------------------
.**.....................................       995   99.5%
.........*.*............................       984   98.4%
...*....*.*.............................       981   98.1%
..................**..*.*...............       949   94.9%
..................*....*...............       948   94.8%
.....*.....*............................       927   92.7%
..............***..**...................       868   86.8%
.............................*.*.....          838   83.8%
..................*...*.................       770   77.0%
.**.........*...........................       732   73.2%
..............*.....*...................       724   72.4%
..............**********................       669   66.9%
..............**..*.....................       647   64.7%
.**********************.*****.*.****           635   63.5%
..............**********.***.*....*..          623   62.3%
..............**********.**.......*..          616   61.6%
..............**********.*****.*.***.          605   60.5%  ← HSP split
........................**..............       601   60.1%
.****************************.*.****           592   59.2%
..............*...*.....................       587   58.7%
..............********.*................       581   58.1%
..............**********..........*..          515   51.5%
..............**********.***......*..          460   46.0%
.**...*.....*...........................       429   42.9%
..........................*....*...            412   41.2%
.**.****.*.************.*****.*.****           392   39.2%
```

```
..............**.....................      388  38.8%
.............**********.***.*.*..**.      373  37.3%
.....*.*...*.........................      372  37.2%
............**********.*........*..      364  36.4%
.............***..**.*..............      328  32.8%
.............**********.**..*....*..      327  32.7%
.............**********.***.*.*..*..      311  31.1%
.............**********.***.*....**.      306  30.6%
........................*.....*.*....      261  26.1%
........................**.........      246  24.6%
......**.............................      235  23.5%
.............**********.*...........      228  22.8%
.**...*..............................      226  22.6%
.**...*..*..*........................      222  22.2%
.**..***.*.**........................      215  21.5%
.............**********.***.*.*.***.      213  21.3%
.................**....*..............      212  21.2%
....*........**********.*****.*.***.      201  20.1%
.............**********.*****...***.      194  19.4%
..............*.*....................      168  16.8%
.............****************.*.***.      161  16.1%
.**.****.*.******************.*.****      160  16.0%
.**...**....*........................      159  15.9%
.**......*..*........................      149  14.9%
.............**********.*****.*.**..      147  14.7%
........................*.......*....      143  14.3%
.............*.*...*.................      139  13.9%
.............**********.**...........      139  13.9%
.**..***.*.**.......................*      135  13.5%
.............**********.*****...**..      129  12.9%
.....*...*.*.........................      128  12.8%
.**...**.*..*........................      128  12.8%
.........*..*........................      120  12.0%
....*........****************.*.***.      115  11.5%
.**.****.*.**.......................*      112  11.2%
.............**********.*****.*.****      109  10.9%
.****************************.******      106  10.6%
.............****************.*****.      104  10.4%
.**.****.*.**........................      103  10.3%
.................**..*.*.*...........      103  10.3%
............................*.*.....       99   9.9%
..............*...**.................       96   9.6%
.**..*...*.**........................       96   9.6%
.............**********.***.*.*.**..       94   9.4%
.**.****.*.******************.******       93   9.3%
.........*...........................*       91   9.1%
............................*........*       90   9.0%
.***.********.......................*       89   8.9%
.............**********.**..*.*..*..       87   8.7%
```

```
....*........***************.*****.        86   8.6%
.............********.*.*...........        84   8.4%
.............**********.***.*...**..        84   8.4%
.............**********.*...*....*..        83   8.3%
.***********.......................*        81   8.1%
..............**..**...............        77   7.7%
.............***..**.*...........*..        72   7.2%
......**........................*            71   7.1%
....*........**********.*****.*.****        68   6.8%
....*.*.........................            67   6.7%
....*........***************.*.****        59   5.9%
.............***..*.................        59   5.9%
.............**********.*****.*..**.        59   5.9%
.**...**.....:......................        58   5.8%
.............**************.*.*.***.        54   5.4%
.***....*.*.*.......................        54   5.4%
......*.........................*            53   5.3%
.............**********.***.*...***.        53   5.3%
.**.*.**.*..***********.*****.*.****        50   5.0%
.......................*......*.....        50   5.0% ← HSP incompatible split
```

LITERATURE

Adams, J. and P. W. Oeller (1986). Structure of evolving populations of *Saccharomyces cerevisiae*: adaptive changes are frequently associated with sequence alterations involving mobile genetic elements. *Proceedings of the National Academy of Sciences (USA)* 83: 7124-7127.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215(3): 403-10.

Anaissie, E. (1992). Opportunistic mycoses in the immunocompromised host: experience at a cancer center and review. *Clinical Infectious Diseases* 14(S1): S43-53.

Anderson, J., T. Srikantha, B. Morrow, S. H. Miyasaki, T. C. White, N. Agabian, J. Schmid and D. R. Soll (1993). Characterization and partial nucleotide sequence of the DNA fingerprinting probe Ca3 of *Candida albicans*. *Journal of Clinical Microbiology* 31: 1472-1480.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman and J. A. Smith (1989). Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley-Interscience.

Bandelt, H.-J. and A. W. M. Dress (1992). Split Decomposition: A new and useful approach to phylogenetic analysis of distance data, molecular phylogenetics and evolution. *Molecular Phylogenetics and Evolution* 1(3): 242-252.

Barnetson, R. and L. J. R. Milne (1973). Skin sampling for *Candida* with adhesive tape. *British Journal of Dermatology* 88: 487.

Bauer, T. M., E. Ofner, H. M. Just and F. D. Daschner (1990). An epidemiological study assessing the relative importance of airborne and direct contact transmission of microorganisms in a medical intensive care unit. *Journal of Hospital Infections* 15: 301-309.

Beck-Sague, C. M. and W. R. Jarvis (1993). Secular trends in the epidemiology of nosocomial infections in the United States, 1980-1990. *Journal of Infectious Diseases* 167: 1247-1251.

Berger, C., R. Frei and A. Gratwohl (1988). Bottled lemon juice- a cryptic source of invasive *Candida* infections in the immunocompromised host. *Journal of Infectious Diseases* 158: 654-655.

Bodey, G. P. (1984). Candidiasis in cancer patients. *American Journal of Medicine* 77 (suppl. 4D): 13-19.

Bodey, G. P. (1993). Candidiasis Pathogenesis, Diagnosis, and Treatment. New York, Raven Press Ltd.

Botelho, A. R. and R. J. Planta (1994). Specific identification of *Candida albicans* by hybridisation with oligonucleotides derived from ribosomal DNA internal spacer. *Yeast* 10: 709-717.

Brachman, P. S. (1986). Epidemiology of nosocomial infections. Boston, Little, Brown and Co.

Bross, J., G. Talbot, G. Maislin, S. Hurwitz and B. Strom (1989). Risk factors for nosocomial candidemia: a case-control study in adults without leukemia. *American Journal of Medicine* 87: 617-20.

Brown, M. W., C. Evans, J. L. Ford and M. Pilling (1986). A note on the recovery of micro-organisms from an oil-in-water cream. *Journal of Clinical and Hospital Pharmacy* 11: 117-123.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford and P. J. Wadell (1993). Partitioning and combining data in phylogenetic analysis. *Systematic Biology* 42(3): 384-397.

Butler, K. M. and C. J. Baker (1988). *Candida*: an increasingly important pathogen in nursery. *Pediatr.Clin.N.Am.* 35: 543-563.

Cannon, R. D., A. R. Holmes, A. B. Mason and B. C. Monk (1995). Oral *Candida*: clearance, colonization or candidiasis. *Journal of Dental Research* 74(5): 1152-1161.

Casewell, M. and I. Phillips (1977). Hands as a route of transmission for Kebsiella species. *British Medical Journal* 2: 1315-1317.

Chen, J.-Y. and W. A. Fonzi (1992). A temperature-regulated, retrotransposon-like element from *Candida albicans*. *Journal of Bacteriology* 174: 5624-5632.

Clarke, A. (1995). Understanding cancer. Summerville House.

Cutler, J. E., P. M. Glee and H. L. Horn (1988). *Candida albicans*- and *Candida stellatoidea*-specific DNA fragments. *Journal of Clinical Microbiology* 26: 1720-1724.

Davidson, F. and R. F. Mould (1978). Recurrent genital Candidosis in women and the effect of intermittent prophylactic treatment. *British Journal of Veneral Diseases* 54: 176-183.

de Barros Lopez, M., S. Rainieri, P. A. Henschke and P. Langridge (1999). AFLP fingerprinting for analysis of yeast genetic variation. *International Journal of Systematic Bacteriology* 49(915-924).

DeGregorio, M. W., W. M. F. Lee, C. A. Linker, R. A. Jacobs and C. A. Ries (1982). Fungal infections in patients with acute leukemia. *American Journal of Medicine* 73: 543-548.

Dembry, L. M., J. A. Vazquez and M. J. Zervos (1994). DNA analysis in the study of the epidemiology of nosocomial candidiasis. *Infection Control and Hospital Epidemiology* 15(1): 48-53.

Doebbeling, B. N., R. J. Hollis, H. D. Isenberg, R. P. Wenzel and M. A. Pfaller (1991). Restriction fragment analysis of a *Candida tropicalis* outbreak of sternal wound infections. *Journal of Clinical Microbiology* 29(6): 1268-1270.

Dollinger, M. and E. H. Rosenbaum (1998). Everyone´s guide to cancer therapy, Summerville House.

Dopazo, J., A. Dress and A. von Haeseler (1993). Split decomposition: a technique to analyze viral evolution. *Proceedings of the National Academy of Sciences (U. S. A.)* 90(21): 10320-10324.

Eigen, M. (1993). The origin of genetic information: viruses as models. *Gene* 135(1-2): 37-47.

Ellsworth, D. L., K. D. Rittenhouse and R. L. Honeycutt (1993). Artefactual variation in randomly amplified polymorphic DNA banding patterns. *Biotechniques* 14: 214-217.

Fenton Wiliams, J. (1989). Optimization strategies for the polymerase chain reaction. *BioTechniques* 7(7): 762-768.

Finkelstein, R., G. Reinhertz, N. Hashman and D. Merzbach (1993). Outbreak of *Candida tropicalis* fungemia in a neonatal intensive care unit. *Infection Control and Hospital Epidemiology* 14: 587-590.

Fisher-Hoch, S. P. and L. Hutwanger (1995). Opportunistic candidiasis: an epidemic of the 1980s. *Clinical Infectious Diseases* 21: 897-904.

France, D. (1968). Survival of *Candida albicans* in hand creams. *New Zealand Medical Journal* 67: 552-557.

Fridkin, S. K. and W. A. Jarvis (1996). Epidemiology of nosocomial fungal infections. *Clinical Microbiology Reviews* 9: 499-511.

Goodwin, T. J. D. and R. T. M. Poulter (1998). The CARE-2 and Rel-2 repetitive elements of *Candida albicans* contain LTR fragments of a new transposon. *Gene* 218: 85-93.

Hazen, K. C. and P. M. Glee (1994). Hydrophobic cell wall protein glycosylation by the pathogenic fungus *Candida albicans*. *Canadian Journal of Microbiology* 40(4): 266-272.

Hazen, K. C. and B. W. Hazen (1992). Hydrophobic surface protein masking by the opportunistic fungal pathogen *Candida albicans*. *Infection and Immunity*. 60(4): 1499-1508.

Hellstein, J., H. Vawter-Hugart, P. Fotos, J. Schmid and D. R. Soll (1993). Genetic similarity and phenotypic diversity of commensal and pathogenic strains of *Candida albicans* isolated from the oral cavity. *Journal of Clinical Microbiology* 31(12): 3190-3199.

Higgins, D. G. and P. M. Sharp (1989). Fast and sensitive multiple sequence alignments on a microcomputer. *CABIOS Commun* 5: 151-153.

Holmes, D. S. and M. Quigley (1981). A rapid boiling method for the preparation of bacterial plasmids. *Analytical Biochemistry* 114(1): 193-7.

Holmes, E. C., M. Worobey and A. Rambaut (1999). Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution* 16(3): 405-409.

Horn, R., B. Wong, T. Kiehn and D. Armstrong (1985). Fungemia in a Cancer Hospital: Changing Frequency, Earlier Onset, and Results of Therapy. *Reviews of Infectious Diseases* 7(5): 646-655.

Huelsenbeck, J. P., J. J. Bull and C. W. Cunningham (1996). Combining data in phylogenetic analysis. *Tree* 11(4): 152-157.

Hunter, P. R. (1991). A critical review of typing methods for *Candida albicans* and their applications. *Critical Reviews in Microbiology* 17(6): 417-434.

Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68-73.

Isenberg, H. D. (1989). Single-source outbreak of Candida tropicalis. *Journal of Clinical Microbiology* 27: 2426-2428.

Iwaguchi, S.-I., M. Homma and K. Tanaka (1990). Variation in the electrophoretic karyotype analysed by the assignment of DNA probes in *Candida albicans*. *Journal of General Microbiology* 136: 2433-2442.

Iwaguchi, S., M. Homma, H. Chibana and K. Tanaka (1992). Isolation and characterization of a repeated sequence (RPS1) of *Candida albicans*. *Journal of General Microbiology* 138(9): 1893-1900.

Jarvis, W. R. (1996). The epidemiology of colonisation. *Infection Control and Hospital Epidemiology* 17(1): 47-52.

Kamiyama, A., M., M. Niimi, M. Tokunaga and H. Nakayama (1989). DNA homology between *Candida albicans* strains: evidence to justify the synonymous status of *C. stellatoidea* *Mycopathologia* 107(1): 3-7.

Kashbur, I. M., G. A. J. Ayliffe and R. H. George (1980). The survival of *Candida albicans* in moist and dry environments. *Journal of Hospital Infections* 1: 349-356.

Klein, J. J. and C. Watanakunakorn (1979). Hospital-acquired fungemia. It's natural course and clinical significance. *American Journal of Medicine* 67: 51-8.

Kleinegger, C. L., S. R. Lockhart, K. Vargas and D. R. Soll (1996). Frequency, intensity, species, and strains of oral *Candida* vary as a function of host age. *Journal of Clinical Microbiology* 34(9): 2246-2254.

Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* 18: 1-32.

Lasker, B. A., L. S. Page, T. J. Lott and G. S. Kobayashi (1992). Isolation, characterization, and sequencing of *Candida albicans* repetitive element 2. *Gene* 116: 51-57.

Levitz, S. M. (1992). Overview of host defenses in fungal infections. *Clinical Infectious diseases* 14(S1): S37-42.

Lichtmann, M. A. (1980). Hematology and Oncology. New York, Grune & Stratton, Inc.

Lockhart, P., P. A. McLenachan, D. Havell, D. Glenny, D. Huson and U. Jensen (1999). Phylogeny, dispersal and radiation of New Zealand alpine buttercups: molecular evidence under split decomposition. *submitted*.

Lockhart, P. and T. McLenachan (1997). Isolating polymorphic plant DNA fragments identified using AFLP™ technology without acrylamide gels:markers for evolutionary studies. *Focus* 19(3): 70-71.

Lockhart, S. R., J. J. Fritch, A. Sturdevant Meyer, K. Schröppel, T. Srikantha, R. Galask and D. R. Soll (1995). Colonizing populations of *Candida albicans* are clonal in origin but undergo microevolution through C1 fragment reorganization as demonstrated by DNA fingerprinting and C1 sequencing. *Journal of Clinical Microbiology* 33(6): 1501-1509.

Lockhart, S. R., B. D. Reed, C. L. Pierson and D. R. Soll (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. *Journal of Clinical Microbiology* 34(4): 767-777.

Lott, T. J., B. M. Bums, R. Zancope-Oliveira and E. Reiss (1998). Sequence Analysis of the Internal Spacer 2 (ITS2) from Yeast Species Within the Genus *Candida*. *Current Microbiology* 36: 63-69.

Magee, B. B., T. M. D´Souza and P. T. Magee (1987). Strain and species identification by restriction fragment polymorphisms in the ribosomal DNA repeat of *Candida* species. *Journal of Bacteriology* 169: 1639-1643.

Magee, B. B. and P. T. Magee (1987). Electrophoretic karyotypes and chromosome numbers in *Candida* species. *Journal of Clinical Microbiology* 133: 425-430.

Majer, D., R. Mithen, B. G. Lewis, P. Vos and R. P. Oliver (1996). The use of AFLP fingerprinting for the detection of genetic variation in fungi. *Mycological Research* 100(9): 1107-1111.

Maki (1982). Relation of inanimate hospital environment to endemic nosocomial infection. *The New England Journal of Medicine* 307: 1562-1566.

Malloy, P. J., X. Zhao, N. D. Madani and D. Feldman (1993). Cloning and expression of the gene from *Candida albicans* that encodes a high affinity corticosteroid-binding protein. *Proceedings of the National Academy of Sciences (U. S. A.)* 90: 1902-1906.

Matthews, G. D., T. J. D. Goodwin, M. I. Butler, T. A. Berryman and R. T. M. Poulter (1997). pCal, a highly unusual Ty1/copia retrotransposon from the pathogenic yeast *Candida albicans*. *Journal of Bacteriology* 179: 7118-7128.

Maxam, A. M. and W. Gilbert (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods in Enzymology* 65: 499-560.

Mellor, J. (1993). Multiple interactions control the expression of yeast genes. The eukaryotic genome organisation and regulation. P. M. A. Broda, S. G. Oliver and P. F. G. Sims. Cambridge, Press Syndicate of the University of Cambridge.

Mercure, S., N. Rougeau, S. Montplaisir and G. Lemay (1993). Complete nucleotide sequence of *Candida albicans* 5.8S rRNA coding gene and flanking internal transcribed spacers. *Nucleic Acids Research* 21(19): 4640.

Meunier, F., M. Aoun and N. Bitar (1992). Candidemia in immunocompromised patients. *Clinical Infectious Diseases* 14(S1): S120-125.

Moro, M. L., C. Maffei, E. Manso, G. Morace, L. Polonelli and F. Biavasco (1990). Nosocomial outbreak of systemic Candidosis associated with parenteral nutrition. *Infection Control and Hospital Epidemiology* 11(1): 27-35.

Mueller, U. G. and L. LaReesa Wolfenbarger (1999). AFLP genotyping and fingerprinting. *Tree* 14(10): 389-394.

Mueller, U. G., S. E. Lipari and M. G. Milgroom (1996). Amplified fragment length polymorphism (AFLP) fingerprinting of symbiotic fungi cultured by the fungus-growing ant *Cyphomyrmex minutus*. *Molecular Ecology* 5: 119-122.

Neuhaus-Url, G. (1993). The use of the nonradioactive digoxigenin chemiluminescent technology for plant genomic Southern blot hybridization: a comparison with radioactivity. *Transgenetic Research*. 2: 115-120.

Odds, F. C. (1988). *Candida* and Candidosis. London, Bailliere Tindall.

Odds, F. C. (1994). *Candida* species and virulence. *ASM News* 60(6): 313-318.

Odds, F. C., C. C. Kibbler, E. Walker, A. Bhamra, H. G. Prentice and P. Noone (1989). Carriage of *Candida* species and *C. albicans* biotypes in patients undergoing chemotherapy or bone marrow transplantation for haematological disease. *Journal of Clinical Pathology* 42: 1259-1266.

Oldham, R. K. (1991). Principles of cancer biotherapy. New York, Marcel Dekker,Inc.

Ortiz, O. C. and A. Stoliar (1988). Immunological changes in human breast cancer. *European Journal of Gynaecology and Oncology* 4(6): 502-514.

Page, R. D. M. and E. C. Holmes (1998). Molecular evolution: a phylogenetic approach. Oxford, Blackwell Science Ltd.

Pechere, J. C. (1993). Microbiology of nosocomial infections. *Bull.Acad.Natl.Med.* 177: 705-717.

Penny, D. and M. D. Hendy (1985). The use of tree comparison metrics. *Systematic Zoology* 34(1): 75-82.

Perreau, V. M., M. A. S. Santos and M. F. Tuite (1997). Beta, a novel repetitive element associated with tRNA genes in the pathogenic yeast *Candida albicans*. *Molecular Microbiology* 25: 229-236.

Pfaller, M., I. Cabezudo, F. Koontz, M. Bale and R. Gingrich (1987). Predictive value of surveillance cultures for systemic infection due to *Candida* species. *Journal of Clinical Microbiology* 1987: 628-633.

Pfaller, M. A. (1994). Epidemiology and control of fungal infection. *Clinical Infectious Disease* 19(suppl. 1): S8-S13.

Pfaller, M. A. (1995). Epidemiology of candidiasis. *Journal of Hospital Infections* 30(suppl.): 329-338.

Pfaller, M. A. (1996). Nosocomial candidiasis: ermerging species, reservoirs, and modes of transmission. *Clinical Infectious Diseases* 22(suppl.): S89-S94.

Pfaller, M. A., S. R. Lockhart, C. Pujol, J. A. Swails-Wenger, S. A. Messer, M. B. Edmont, R. N. Jones, R. Wenzel and D. R. Soll (1998). Hospital specificity, region specificity, and fluconazole resistance of *Candida albicans* bloodstream isolates. *Journal of Clinical Microbiology* 36: 1518-1529.

Phleps, M., G. A. J. Ayliffe and J. R. Babb (1986). An outbreak of candidiasis in a special care baby unit: the use of a resistogram typing method. *Journal of Hospital Infection* 7: 13-20.

Prasad, J. K., I. Feller and P. D. Thomson (1987). A ten-year review of *Candida* sepsis and mortality in burn patients. *Surgery* 101: 213-216.

Pujol, C., S. Joly, S. R. Lockhart, S. Noel, M. Tibayrenc and D. R. Soll (1997). Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis, and Southern blot hybridisation with moderately repetitive DNA probe Ca3 for fingerprinting *Candida albicans*. *Journal of Clinical Microbiology* 35(9): 2348-2358.

Pujol, C., S. Jo[illegible] B. Nolan, T. Srikantha and D. R. Soll (1999). Microevolutionary changes in *Candida albicans* identified by the complex Ca3 fingerprinting probe involve insertions and deletions of the full-length repetitive sequence RPS at specific genomic sites. *Microbiology* 145: 2635-2646.

Rangel-Frausto, M. S., A. K. Houston, M. J. Bale, C. Fu and R. P. Wenzel (1994). An experimental model for study of *Candida* survival and transmission in human volunteers. *European Journal of Clinical Microbiology and Infectious Disease* 13: 590-595.

Rashid, S., M. Collins, J. Corner and R. S. Morton (1984). Survival of *Candida albicans* on fabric after laundering. *British Journal of Venereal Diseases* 60: 277.

Reagan, D. R., M. A. Pfaller, R. J. Hollis and R. P. Wenzel (1990). Characterisation of the sequence of colonization and nosocomial candidemia using DNA fingerprinting and a DNA Probe. *Journal of Clinical Microbiology* 28: 2733-2738.

Reingold, A. L., X. D. Lu, B. D. Plikaytis and L. Ajello (1986). Systemic mycoses in the United States, 1980-1982. *J.Med.Vet.Mycol.* 24: 433-436.

Robert, R., C. Mahaza, A. Marot-Leblond, G. Tronchin and J.-M. Senet (1991). Binding of mouse fibrinogen to *Candida albicans* in vivo. *FEMS Microbiology Letters* 78: 301-304.

Rosendahl, S. and J. W. Taylor (1997). Development of multiple genetic markers for studies of genetic variation in arbuscular mycorrhizal fungi using AFLP^{TM}. *Molecular Ecology* 6: 821-829.

Ross, I. K., F. De Bernardis, G. W. Emerson, A. Cassone and P. A. Sullivan (1990). The secreted aspartate proteinase of *Candida albicans*: physiology of secretion and virulence of a proteinase-deficient mutant. *Journal of General Microbiology* 136(Pt 4): 687-94.

Russel, C. and K. M. Lay (1973). Natural history of Candida species and yeasts in the oral cavities of infants. *Archieves of Oral Biology* 18: 957-962.

Rutschenko-Bulgac, E. P., F. Sherman and J. B. Hicks (1990). Chromosomal rearrangements associated with morphological mutants provide a means for genetic variation of Candida albicans. *Journal of Bacteriology* 172(3): 1276-1283.

Sachs, L. (1991). Angewandte Statistik. Heidelberg, Spriner-Verlag.

Saitou, N. and M. Nei (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: a laboratory manual. New York, Cold Spring Harbor.

Sanchez, V., J. A. Vazquez, D. Barth-Jones, L. Dembry, J. D. Sobel and M. J. Zervos (1993). Nosocomial acquisition of *Candida parapsilosis*: an epidemiologic study. *The American Journal of Medicine* 94: 577-582.

Sanger, F., J. E. Donelson, A. R. Coulson, H. Kossel and D. Fischer (1974). Determination of a nucleotide sequence in bacteriophage f1 DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 90(2): 315-333.

Scherer, S. and D. A. Stevens (1987). Application of DNA typing methods to epidemiology and taxonomy of *Candida* species. *Journal of Clinical Microbiology* 25: 675-679.

Scherer, S. and D. A. Stevens (1988). A *Candida albicans* dispersed, repeated gene family and its epidemiologic applications. *Proceedings of the National Academy of Sciences (USA)* 85: 1452-1456.

Schmid, J. unpublished data.

Schmid, J. (1993). Candidiasis: conclusions from DNA fingerprinting. *Clinical Advances in the Treatment of Fungal Infections* 4(6): 12-16.

Schmid, J., *et al.* (1999). Evidence for a general purpose genotype in *Candida albicans*, highly prevalent in multiple geographic regions, patient types and types of infection. *Microbiology* 145: 2405-2413.

Schmid, J., P. R. Hunter, G. C. White, A. K. Nand and R. D. Cannon (1995). Physiological traits associated with success of *Candida albicans* strains as commensal colonisers and pathogens. *Journal of Clinical Microbiology* 33(11): 2920-2926.

Schmid, J., F. C. Odds, M. J. Wiselka, K. G. Nicholson and D. R. Soll (1992). Genetic similarity and maintenance of *Candida albicans* strains from a group of AIDS patients, demonstrated by DNA fingerprinting. *Journal of Clinical Microbiology* 30(4): 935-941.

Schmid, J., M. Rotman, B. Reed, C. L. Pierson and D. R. Soll (1993). Genetic similarity of *Candida albicans* strains from vaginitis patients and their partners. *Journal of Clinical Microbiology* 31(1): 39-46.

Schmid, J., Y. P. Tay, L. Wan, M. Carr, D. Parr and W. McKinney (1995). Evidence for nosocomial transmission of *Candida albicans* obtained by Ca3 fingerprinting. *Journal of Clinical Microbiology* 33(5): 1223-1230.

Schmid, J., E. Voss and D. R. Soll (1990). Computer- assisted methods for assessing strain relatedness in *Candida albicans* by fingerprinting with the moderately repetitive sequence Ca3. *Journal of Clinical Microbiology* 28: 1236-1243.

Schröppel, K., M. Rotman, R. Galask, K. Mac and D. R. Soll (1994). Evolution and replacement of *Candida albicans* strains during recurrent vaginitis demonstrated by DNA fingerprinting. *Journal of Clinical Microbiology* 32(11): 2646-2654.

Seeliger, H. P. R. and C. Patzelt (1991). Host-Parasite Interaction-Mechanisms of pathogenesis. *Candida* and Candidamycosis. E. Tumbay, H. P. R. Seeliger and O. Ang. New York, Plenum Press: 55-59.

Sherertz, R. J., K. S. Gledhill, K. D. Hampton, M. A. Pfaller, L. B. Givner, J. S. Abramson and R. G. Dillard (1992). Outbreak of *Candida* bloodstream infections associated with retrograde medication administration in a neonatal intensive care unit. *The Journal of Pediatrics* 120: 455-461.

Sherwood, J., N. A. R. Gow, G. W. Gooday, D. W. Gregory and D. Marshall (1992). Contact sensing in *Candida albicans*- a possible aid to epithelial penetration. *Journal of Medical and Veterinary Mycology* 30: 461-469.

Sneath, P. H. A. and R. R. Sokal (1973). Numerical taxonomy. The principles and practice of numerical classification. San Francisco, W.H.Freeman & Co.: 230-234.

Sobel, J. D., G. Muller and H. R. Buckely (1984). Critical role of germ tube formation in the pathogenesis of candidal vaginitis. *Infection and Immunity* 44: 576-580.

Soll, D. R. (1991). Current Status of the Molecular Basis of *Candida* Pathogenicity. The Fungal Spore and Disease Initiation in Plants and Animals. G. T. Cole and H. C. Hoch, Plenum Publishing Corporation: 503-540.

Soll, D. R. (1992). Switching and its possible role in Candida pathogenesis. Edinburgh, Churchill Livingstone.

Soll, D. R. (1993). DNA fingerprinting of *Candida albicans*. *Journal of Mycological Medicine* 3: 37-44.

Soll, D. R., R. Galask, J. Schmid, C. Hanna, K. Mac and B. Morrow (1991). Genetic dissimilarity of commensal strains of *Candida* subsp. carried in different anatomical locations of the same healthy women. *Journal of Clinical Microbiology* 29(8): 1702-1710.

Soll, D. R., M. Staebell, C. Langtimm, M. Pfaller, J. Hicks and T. V. G. Rao (1988). Multiple *Candida* strains in the course of a single systemic infection. *Journal of Clinical Microbiology* 26(8): 1448-1459.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98(3): 503-507.

Srikantha, T., R. R. Gutell, B. Morrow and D. R. Soll (1994). Partial nucleotide sequence of a single ribosomal RNA coding region and secondary structure of the large subunit 25 s rRNA of *Candida albicans*. *Current Genetics* 26(4): 321-328.

Steel, M. A and D. Penny (1993).Distribution of tree comparison metrics-some new results. *Systematic Biology* 42: 126-141.

Steel, M. A., M. A. Cooper and D. Penny (1996). Confidence intervals for the divergence between two clades. *Systematic Biology* 45: 127-134.

Sternberg, S. (1995). The emerging fungal threat. *Science* 266: 1632-1634.

Strausbaugh, L. J., D. L. Sewell, T. T. Ward, M. A. Pfaller, T. Heityman and R. Tjoelker (1994). High frequency of yeast carriage on hands of hospital personnel. *Journal of Clinical Microbiology* 32: 2299-2300.

Strimmer, K. and A. von Haeseler (1996). Quartet Puzzling: a maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* 13: 964-969.

Strimmer, K.; Goldman, N and A. von Haeseler (1997).Bayesian Probabilities and Quartet Puzzling. *Molecular Biology and Evolution* 14: 210-211.

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.

Swofford, D. L. (1998). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sunderland, MA, Sinaur Associates.

Swofford, D. L., G. J. Olsen, P. J. Waddell and D. M. Hillis (1996). Phylogenetic interference. Sunderland, Massachusetts, Sinauer Associates.

Taschdjian, C. L., J. J. Burchall and P. J. Kozinn (1960). Rapid Identification of *Candida albicans* by filamentation on serum and serum substitudes. *A.M.A. Journal of Diseases of Children* 99: 212-215.

Taylor, J. W., D. M. Geiser, A. Burt and V. Koufopanou (1999). The evolutionary biology and population genetics underlying fungal strain typing. *Journal of Clinical Microbiology* 12(1): 126-146.

Tomada, T. and Y. Nakano (1984). Variations in intestinal *Candida* populations in patients receiving antileukemic therapy. *Bulletin of the Osaka Medical School* 30-1: 14-18.

Thrash-Bingham, C. and J. A. Gorman (*1992*). DNA translocations contribute to chromosome length polymorphisms in *Candida albicans*. *Current Genetics* 22: 93-100.

van Belkum, A., W. Melchers, B. E. de Pauw, S. Scherer, W. Quint and J. F. Meis (1994). Genotypic characterization of sequential *Candida albicans* isolates from fluconazole-treated neutropenic patients. *Journal of Infectious Diseases* 169: 1062-1070.

van Nues, R. W., J. M. J. Rientjes, S. A. Morre, E. Mollee, R. J. Planta, J. Venema and H. A. Raue (1995). Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precurser ribosomal RNA. *Journal of Molecular Biology* 250: 24-36.

Vazquez, J., A., A. Beckley, S. Donabedian, J. D. Sobel and M. J. Zervos (1991). Comparison of restriction enzyme analysis and pulsed-field gradient gel electrophoresis as typing systems for *Candida albicans*. *Journal of Clinical Microbiology* 29: 962-967.

Vazquez, J. A., V. Sanchez, C. Dmuchowski, L. M. Dembry, J. D. Sobel and M. J. Zervos (1993). Nosocomial acquisition of *Candida albicans*: an epidemiologic study. *Journal of Infectious Diseases* 168: 195-201.

Veldman, G. M., J. Klootwijk, H. van Heerikhuizen and R. J. Planta (1981). The nucleotide sequence of the intergenic region between the 5.8S and the 26S rRNA genes of the yeast ribosomal operon. Possible implications for the interaction between 5.8S and 26S rRNA and processing of the primary transcript. *Nucleic Acids Research* 9: 4847-4862.

Vos, P., *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23(21): 4407-4414.

Voss, A., M. A. Pfaller and B. N. Doebbeling (1995). Investigation of *Candida albicans* transmission in a surgical intensive care unit cluster by using genomic DNA typing methods. *Journal of Clinical Microbiology* 33(3): 576-580.

Weems, J. J., M. E. Chamberland, J. Ward, M. Willy, A. A. Padhye and S. L. Solomon (1987). *Candida parapsilosis* fungemia associated with parental nutrition and contaminated blood pressure transducers. *Journal of Clinical Microbiology* 25: 1029-1032.

Wey, S. B., M. Mori, M. A. Pfaller, R. F. Woolson and R. P. Wenzel (1989). Risk Factor for hospital-acquired candidemia. *Archives of Internal Medicine* 149: 2349-2353.

Wills, J. W., B. A. Lasker, K. Sirotkin and W. S. Riggsby (1984). Repetitive DNA of *Candida albicans*: nuclear and mitochondrial compounds. *Journal of Bacteriology* 157: 918-924.