Universidade de Aveiro Departamento de Biologia
2007

Inês Cátia Dinis de Sousa

Descrição e caracterização de um locus de microssatélite no gene *IFF8* de *Candida albicans*

Description and characterization of a microsatellite locus in *Candida albicans IFF8* gene

Universidade de Aveiro Departamento de Biologia
2007

Inês Cátia Dinis de Sousa

Descrição e caracterização de um locus de microssatélite no gene *IFF8* de *Candida albicans*

Description and characterization of a microsatellite locus in *Candida albicans IFF8* gene

dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Célia Pais, Professora do Departamento de Biologia da Universidade do Minho e co-orientação da Doutora Ângela Cunha, Professora do Departamento de Biologia da Universidade de Aveiro.

dedico este trabalho à minha família pelo apoio incondicional.

o júri

presidente

Prof. Dr. António Correia
professor associado com agregação do Departamento de Biologia da Universidade de Aveiro

Prof. Dra. Célia Pais
professora associada do Departamento de Biologia da Universidade do Minho

Prof. Dra. Ângela Cunha
professora auxiliar do Departamento de Biologia da Universidade e Aveiro

Prof. Dra. Paula Sampaio
professora auxiliar do Departamento de Biologia da Universidade do Minho

agradecimentos

agradeço ao Departamento de Biologia da Escola de Ciências da Universidade do Minho pela oportunidade de realizar este mestrado e, especialmente, à orientadora, a Professora Doutora Célia Pais. Agradeço ainda a todos os que tornaram possível a realização deste trabalho e àqueles que me apoiaram durante todo o percurso.

Resumo

As leveduras do género *Candida* constituem um grupo muito diverso de fungos que inclui espécies comensais da microflora humana normal, podendo tornar-se patogénicas em determinadas condições de imunodepressão do hospedeiro. Apesar de várias espécies do género *Candida* estarem associadas a infecções em humanos, o principal agente é *C. albicans*, responsável por diversas candidíases superficiais das mucosas mas também por infecções sistémicas denominadas candidémias. Uma vez que a patogenicidade e susceptibilidade a antifúngicos varia entre estirpes da mesma espécie, é crucial uma identificação rápida e fiável da estirpe causadora da infecção para que possa ser aplicado o tratamento mais adequado.

O principal objectivo do presente trabalho consiste na descrição e caracterização de um novo microssatélite de *C. albicans* localizado no gene *IFF8* e que codifica uma proteína semelhante a *Hyr1*, um componente da parede celular da forma de hifa de *C. albicans*. Para tal, utilizaram-se 78 isolados clínicos independentes provenientes de dois produtos biológicos, hemocultura e exsudados vaginais, e oriundos de pacientes internados em diversas instituições hospitalares. Foram identificados 16 alelos, com comprimentos entre 200 e 316 pares de bases, e 27 genótipos, o que resulta num poder de discriminação de cerca de 0.90. Foi ainda possível identificar genótipos específicos de isolados de hemoculturas, assim como de exsudados vaginais. Para determinar a origem da variabilidade alélica deste microssatélite a análise da estrutura dos diferentes alelos foi efectuada por sequenciação revelando a sua estrutura composta. Dada a localização deste microssatélite no gene *IFF8* de *C. albicans* e uma vez que as proteínas GPI podem estar envolvidas na biossíntese e remodelação da parede celular, na adesão às células do hospedeiro e na virulência, a sua variabilidade poderá estar relacionada com o grau de virulência das estirpes em estudo.

Abstract

Yeasts of the genus *Candida* include a variety of human commensal species that may become pathogenic in immunocompromised hosts. The predominant causal agent of candidiasis is *C. albicans,* responsible for superficial infections and systemic life threatening conditions. Since pathogenicity and antifungal susceptibility often vary between strains, it is crucial to rapidly and accurately identify the strain causing infection so that tha adequate treatment could be selected.

The aim of this work was to characterize a new microsatellite locus located inside *IFF8* gene of *C. albicans* and coding for a GPI anchored protein similar to *Hyr1* which is a cell wall component of the *C. albicans* hypha. Seventy eight independent *C. albicans* isolates from two biological products, blood cultures and vaginal exudates, and different health institutions were analysed. A total of 16 alleles, with lenghts varying between 200 and 316 bp were found and twenty seven genotypes were identified resulting in a discriminating power of 0.90. It was also possible to identify specific genotypes of blood culture isolates as well as specific of vaginal exudates isolates. Allele length variability was studied by sequence analysis showing that this is a compound microsatellite with a complex internal structure. Given the location of this microsatellite in *C. albicans IFF8* gene and since GPI proteins may be involved in cell wall biosynthesis and remodelling, adhesion and virulence, its variability may be related to strain virulence.

Abbreviations

Als	Agglutinin-like sequence
AFLP	Amplified fragment length polymorphism
bp	base pair
CGD	Candida Genome Database
CSC	Centro de Saúde Carandá Braga
DP	Discriminatory power
DNA	Deoxyribonucleic acid
ECMM	European Confederation of Medical Mycology
EDTA	Ethylenediamine tetraacetic acid
GPI	Glycosylphosphatidylinositol
GPIp	GPI anchored protein
HIV	Human Immunodeficiency Virus
HYR1	HYphally Regulated gene 1
IFF	Individual Protein File Family F
INSA	Instituto Nacional Dr. Ricardo Jorge Lisboa
HSJ	Hospital São João Porto
ICU	Intensive care unit
IPOL	Instituto Português de Oncologia Lisboa
IPOP	Instituto Português de Oncologia Porto
LOH	Loss of heterozygocity
MLEE	Multilocus Enzyme Electrophoresis
MLST	Multilocus sequence typing
PL	Phospholipases
PCR	Polymerase chain reaction
PGFE	Pulsed field gel electrophoresis
RAPD	Randomly amplified polymorphic DNA
REA	Restriction endonuclease analysis
RFLP	Restriction fragment length polymorphism
rpm	Rotations per minute
SDA	Sabouraud Dextrose Agar
SDS	Sodiumdodecyl sulphate
SAP	Secreted aspartyl proteases
SSLP	Simple Sequence Length Polymorphism
SSR	Simple Sequence Repeats
STR	Simple Tandem Repeats
TE	Tris EDTA
VNTR	Variable Number of Tandem Repeats
YPDA	Yeast Peptone Dextrose Agar
ZNF1	Zinc finger transcription factor

Introduction

1. Characteristics and medical relevance of the yeast *Candida albicans*

Over the past several decades, the incidence of fungal infections has increased, especially those acquired in health care associated settings, generally called nosocomial infections. Factors responsible for this rise include aging populations in countries with advanced medical technologies. As these trends continue, it can be predicted that the incidence of invasive nosocomial fungal infections will continue to increase.

The predominant nosocomial fungal pathogens include *Aspergillus* spp., *Mucorales*, *Fusarium* spp. and *Candida* spp. (Perlroth *et al*., 2007). Yeasts belonging to the genus *Candida* have emerged as major fungal opportunistic pathogens in humans whose immune system is debilitated. As commensals *Candida* species are harmless, they belong to the normal microbial flora of skin and mucosal surfaces. However, if the balance of the normal flora is disrupted or the immune defences are compromised, these organisms can outgrow the mucosal flora and cause disease: frequently superficial infections of mucosa. Furthermore, in hospital settings, *Candida* species may cause life-threatening invasive infections in a growing population of vulnerable patients (Mavor *et al*., 2005).

Following reduction in the host immune status, two main types of infections can be observed: superficial and invasive candidiasis. Superficial infections of mucosal epithelial tissues are frequent in immunocompromised patients. In more severe cases, *Candida* species may enter the bloodstream (candidaemia) and can reach almost all organs of the body. Theoretically, any organ can be infected, however, the organs most commonly colonised are the kidney, the brain and the heart (Fridkin & Jarvis,1996).

Approximately two hundred species of *Candida* have been described but only a few are of medical importance (Ng *et al*., 1999). *Candida albicans* is the most frequently isolated species, but other species, such as *C. tropicalis*, *C. krusei*, *C. parapsilosis*, and *C. glabrata*, have increasingly been recognized as pathogens with wide distribution (Correia *et al*., 2004; Bassetti *et al*., 2006; Colombo *et al*., 2006; Snydman, 2006).

The risk of infections caused by yeasts of the genus Candida, generally denominated candidiasis, is influenced by host resistance as well as by external factors. Concerning host resistance, there is an association between diseases causing reduction in the host immune status and candidiasis. Patients are particularly predisposed to these infections when suffering from neutropenia, heart disease, cancer, and diabetes. Patients with late stage Human Immunodeficiency Virus (HIV) infection have also an extremely high incidence of candidiasis. However, HIV infection is not an independent risk factor because the increased incidence of disseminated candidiasis in patients infected with HIV is attributable to the increased incidence of the other usual risk factors. On the other hand, treatments themselves also have immunosuppressive effects that predispose patients to candidiasis, namely chemotherapy, radiotherapy, steroid drugs, and broad-spectrum antibiotics.

Additionally, disruption of normal skin barriers (burn injury, percutaneous catheter placement, abdominal surgery or parenteral nutrition) is an important risk factor for invasive *Candida* infections because it provides a direct route of entry for pathogens. Exogenous acquisition of candidaemia has been reported, with a significantly increased risk of *Candida* bloodstream infections associated with intravascular devices and parenteral nutrition (Tortorano *et al*., 2006).

Prior colonisation with *Candida* species is considered a prerequisite for subsequent deep-seated infection and has been demonstrated to be the leading risk factor for candidaemia (Tortorano *et al*., 2006). It has been shown that patients with higher colonization burdens have a proportionately higher risk of developing disseminated disease (Perlroth *et al*., 2007).

It should be highlighted that not all these predisposing factors equally favour superficial and invasive infections. For example, HIV infected individuals suffer extremely frequently from oral *Candida* infections, but rarely develop disseminated infections (Vargas & Joly, 2002).

It is not clear whether the strains that colonize healthy hosts are responsible for causing subsequent invasive disease when those hosts acquire the appropriate risk factors, or whether infections are caused by acquisition of more virulent strains from environmental sources in the nosocomial setting. Overall data suggest that in

most cases, the source of an infecting strain of *C. albicans* is endogenous flora, but that in certain circumstances transmission of more virulent strains may occur in the nosocomial setting. However, several reports of outbreaks of infection with *Candida* species support the hypothesis that exogenous acquisition (transmission of a strain from one patient to another) of the infecting yeast strain can occur (Ruiz-Diez *et al*., 1997; Shin *et al*., 2005).The increase of vulnerable patients has made *Candida* infections increasingly important, particularly in hospital settings. The frequency of candidaemia among hospitalised patients has doubled during the 1980s and 1990s (Blot & Vandewoude, 2004; Almirante *et al*., 2005). Through the late 1980s, the predominant species causing invasive *Candida* infections was *C. albicans*. In Europe, more than half of all cases of candidemia were caused by *C. albicans* as monitored in three independent investigations (Mavor *et al*., 2005). However, since the 1990s, there has been a steady increase in the relative frequencies of non-albicans species of *Candida* causing disseminated candidiasis (Perlroth *et al*., 2007). The relative contribution of different *Candida* species varies between countries (Tortorano *et al*., 2006). These circumstances have made *Candida* infections an increasingly serious threat in European hospitals. Studies reported a leveling off of the frequency of invasive *Candida* infections during the late 1990s due to the use of prophylaxis and more effective diagnostic methods and antifungal treatment (Blot & Vandewoude, 2004). Furthermore, most recently, several studies have been published contradicting the notion that the incidence of disseminated candidiasis is leveling off, and demonstrating a continued rise in its incidence since the turn of the 21st century (Perlroth *et al*., 2007).

The severity of candidaemia is confirmed by the high crude mortality rate found in the European Confederation of Medical Mycology (ECMM) survey (38%) (Tortorano *et al*., 2006).

The three main routes of entry of *Candida* cells into the circulatory system are shown in Figure 1.

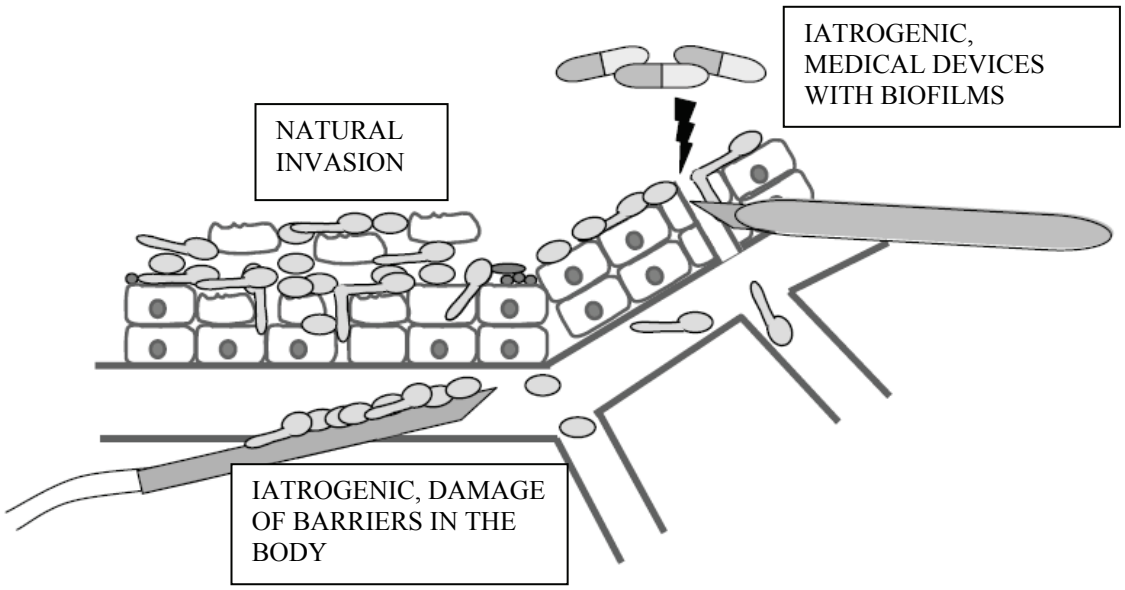

Figure 1. *Candida albicans* may enter the bloodstream by direct penetration from epithelial tissues (natural invasion), due to damage of barriers in the body or may spread from biofilms produced on medical devices (iatrogenic invasion). Adapted from Mavor *et al*., (2005).

The first is the “natural” way through penetration of epithelial cells from mucosal surfaces into deeper tissues and blood capillaries or vessels. The second and third are “artificial” (iatrogenic) routes, which are available due to the use of medical devices or caused by damage of barriers (Mavor *et al*., 2005). Natural invasion is processed in four possible stages as shown in Figure 2.

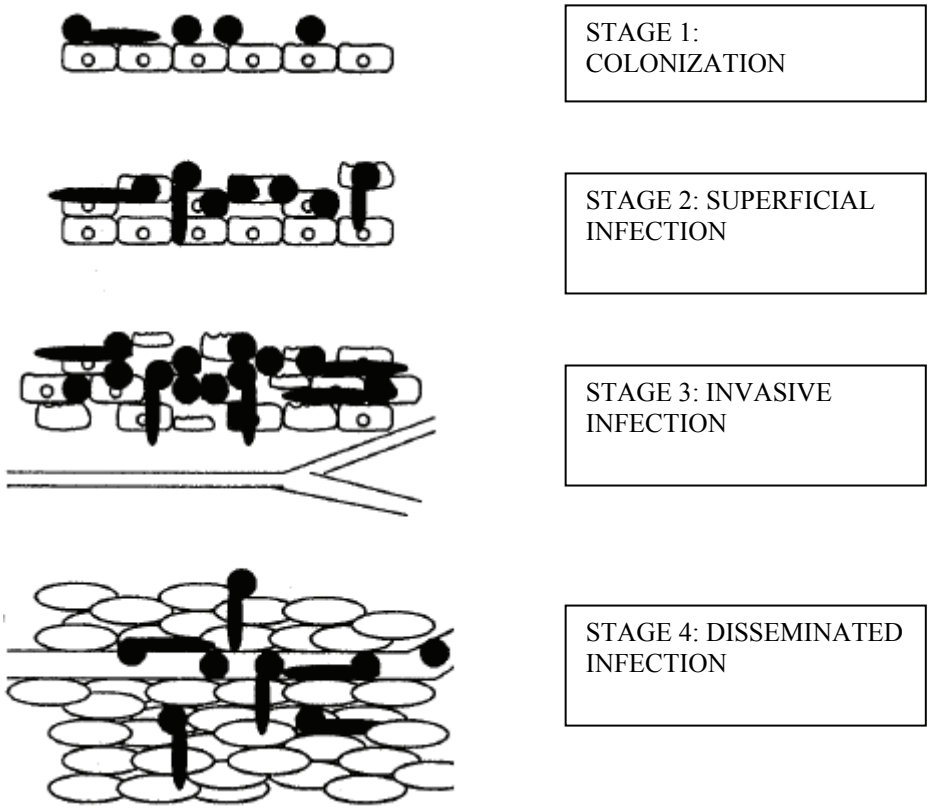

Figure 2 – Stages of infection by natural invasion. Adapted from Naglik *et al.*, (2003).

Not only are invasive *Candida* infections extremely common, they are also difficult to diagnose and treat (Blot & Vandewoude, 2004). Because of the difficulties in confirming the diagnosis with laboratory studies, empiric administration of therapy is often based on a clinical diagnosis of disseminated candidiasis typically made in a patient with symptoms and laboratory features consistent with infection, and who has risk factors for disseminated candidiasis.

Until the 1980s therapy for invasive candidiasis was limited to amphotericin B, but with the advent of new antifungal agents, such as azoles and echinocandins, less toxic therapeutic options are possible. In general, due to its favorable toxicity profile, high oral bioavailability, low cost, and impressive efficacy in randomized clinical trials, fluconazole therapy is preferred (Perlroth *et al.*, 2007).

The antifungal agents available to treat candidiasis are limited in number and effectiveness, due basically to the limited number of identified targets that are clearly distinguished between eukaryotic fungal cells and mammalian cells. In addition to their efficacy, the clinical usefulness of these drugs is hampered by

their safety, as undesirable side effects on patients are often associated with antifungal drugs, and by the emergence of resistances that parallels their clinical use (Gozalbo *et al*., 2004).

2. Virulence factors of *Candida albicans*

In order to establish infection, opportunistic pathogens have to evade the host immune system, survive and divide in the host environment, and spread to new tissues. The characteristics that allow colonization and persistence in the host, conferring the ability to adapt to a variety of microhabitats and cause infections are generally termed virulence factors (Calderone & Fonzi, 2001; Yang, 2003).

During transition from commensal growth to superficial infections, and during penetration into deeper tissue and the bloodstream, *C. albicans* has to adapt to radically changing environments.

Most of *C. albicans* virulence factors are related to the cell wall which is a fungal specific, complex and dynamic structure essential to almost every aspect of its biology including pathogenicity (Gozalbo *et al*., 2004). Since it is the outermost cellular structure, it plays a major role in the interactions between the microorganism and the environment, including the host-fungal interactions. Several studies showed that yeast cell wall proteins are associated with a number of functions in cell physiology including cell surface integrity, flexibility and remodelling, cell morphology, growth, and budding (Calderone & Fonzi, 2001; Bates *et al*., 2006). Thus, this structure is inevitably related with virulence.

Candida albicans virulence factors include (i) attributes which allow cells to adhere to the host, namely adhesins, (ii) hydrolytic enzymes that enable colonisation, invasion and tissue damage (such as secreted aspartyl proteinases (SAP), lipases (LIP), and phospholipases (PL)), (iii) morphogenesis (the reversible transition between unicellular yeast cells and, filamentous, growth forms) that permits shape and gene expression variation according to the pressures of the varying environments during infections, and (iv) phenotypic switching, which is accompanied by changes in antigen expression, colony morphology and tissue affinities (Calderone & Fonzi, 2001; Mavor *et al*., 2005).

2.1. Adherence

Adherence to host cells is seen as an essential early step in the establishment of disease. In addition, *C. albicans* can also adhere on the surfaces of medical devices and form biofilms increasing the possibility of contamination and systemic infection. It was observed that strains with the greater ability to form biofilms and to adhere to the host cells are the most virulent ones (Yang, 2003).

C. albicans displays several adhesins which are biomolecules that promote the adherence to host cells or host-cell ligands. Most of them are glycoproteins, including proteins belonging to Als (agglutinin-like sequence) family which is one of the best characterized groups of cell surface proteins in *C. albicans* (Hoyer, 2001). These proteins are members of the glycosylphosphatidylinositol (GPI) cell wall proteins and share a central domain containing tandem repeats. They are homologous to □-agglutinin, required for cell to cell recognition during mating in *Saccharomyces cereviseae*, and have been shown to present adhesin functions to mammalian cells (Hoyer, 2001; Yang, 2003). All family members have similar functions but seem to be specialized for different conditions encountered in various host environments. Als1, for example, is required in the early stages of morphogenesis, for full adherence and for full virulence in a hematogenously disseminated murine model, and Als5 (Ala1) was found to confer on *S. cerevisiae* cells the ability to adhere to extracellular matrix proteins (Alberti-Segui *et al*., 2004).

2.2. Hydrolytic enzymes

Another factor that contributes to the process of virulence is hydrolytic enzyme production, which is known to play a central role in the pathogenicity of protozoa, bacteria, and pathogenic yeasts. The secretion of hydrolases during infections may be required as a virulence attribute for the degradation of host surfaces (to aid adhesion, invasion, and destruction of host immune factors) in addition to nutrient acquisition. *C. albicans* possesses a large spectrum of hydrolytic enzymes with relatively broad substrate specificities, including secreted aspartyl proteases,

lipases, and phospholipases, which might be the reason for the outstanding position of this human pathogen.

2.2.1. Secreted Aspartyl Proteases

Although many microorganisms possess a variety of hydrolytic enzymes, proteinases particularly secreted aspartyl proteinases (SAP) are by far the most commonly associated with virulence. All proteinases catalyze the hydrolysis of peptide bonds (CO—NH) in proteins but can differ markedly in specificity. SAP are secreted primarily to provide nutrients for the cells; however, pathogenic fungi appear to have adapted this biochemical property to fulfil a number of specialized functions during the infective process. These more direct virulence functions may include digesting host cell membranes or damaging molecules of the host immune system to resist antimicrobial attack. They were also shown to contribute to several other virulence attributes of *C. albicans* during the infection process including hyphal formation, adhesion, and phenotypic switching (Naglik *et al*, 2003).

The *SAP* gene family has ten members and, as with the *Als* proteins, family members have similar functions but seem to be specialised for different conditions encountered in various host environments (Koelsch *et al*., 2000). For example, *Sap1–3* are required for mucosal infections and *Sap4–6* for systemic infections (Albrecht *et al*., 2006). Albrecht *et al*. (2006) showed that *Sap9* and *Sap10*, in contrast to the other family members, are GPI-anchored aspartic proteinases that target proteins of fungal origin necessary for cell surface integrity, cell separation, and adhesion.

2.2.2. Lipases

Lipases are enzymes that are able to catalyse both the hydrolysis and synthesis of ester bonds of lipids. The most prominent role of extracellular lipases for a microorganism is also the digestion of lipids for nutrient acquisition. These enzymes might help both bacteria and fungi to grow in a carbohydrate-restricted

area or environments where lipids are the sole carbon source. Another putative role of these proteins is the enhancement of adhesion, cell growth and hyphal formation to support colonization.

In *C. albicans*, the lipases are also encoded by a gene family with ten members: *LIP1* to *LIP10*. In a similar manner to the *SAP* gene family, the *LIP* genes have been shown to be differentially expressed at different stages and sites of infection (Stehr *et al*., 2003).

2.2.3. Phospholipases

Phospholipases (PL) hydrolyze one or more ester linkages of glycerophospholipids and are classified according to the specific ester bond cleaved. They belong to a group of enzymes capable to derange or destroy cell surface membranes after the adherence is realized ensuring their penetration into the host cells (Ivanovska, 2003). Results from several studies reveal that cells producing less PL are less virulent than cells producing more PL indicating that these enzymes can be classified as virulence factors (Calderone & Fonzi, 2001; Ghannoum, 2000). In addition, *C. albicans* strains isolated from blood produce higher levels of PL than commensal ones (Ibrahim *et al*., 1995).

In *C. albicans*, all four types (A, B, C and D) of PL have been detected but only proteins encoded by the *PL B* family seem to be extracellular. *PLB1* appears to be responsible for most of the extracellular phospholipase activity and it has been shown to be required for virulence in an animal model of disseminated candidiasis (Calderone & Fonzi, 2001).

2.3. Cell morphogenesis

Cell morphogenesis is thought to be an important virulence factor of *C. albicans*. This species exhibits considerable morphogenetic plasticity, growing in either a yeast or hyphal form, or as pseudohyphae, previously thought to represent an intermediate stage between yeast cells and true hyphae but proved to be a distinct developmental form (Whiteway & Oberholzer, 2004). This morphogenetic

variability is linked to pathogenicity. It was shown that cells that are trapped in either the yeast or filamentous forms are significantly reduced in virulence in several infection models (Saville *et al*., 2006).

Both types of morphology may be present in infected tissue, and it is therefore possible that both may play important roles in the pathogenesis. Nevertheless, hyphal growth may be more critical since hyphae adhere more strongly to mammalian cells, promote tissue penetration, and provide a mechanism to escape the attack by macrophages (Ernst, 2000).

Although morphogenesis in *C. albicans* may be an important regulator of the specific response in the host, a variety of extracellular conditions can in turn induce yeast cells to form hyphae and pseudohyphae. The external signals that promote filamentous growth also include pH in neutral-basic range, elevated temperature (37ºC), nutrient starvation (e.g. iron), and exposure to an inductor (serum, *N*-acetylglucosamine, or proline) (Romani *et al*., 2003).

It is generally accepted that dimorphism is a virulence trait per se, but since changes in the cell surface accompany the morphological transition, it is intimately linked with other virulence factors.

2.4. Phenotypic switching

Many pathogenic protozoa, bacteria, and fungi have evolved strategies for alternative expression of surface-related phenotypes enabling their adaptation to changing environments and evasion of the immune response (Lan *et al*., 2002).

In addition to the yeast-hypha transition, *C. albicans* is capable of undergoing a different type of morphological change, termed phenotypic switching, which involves the spontaneous and reversible generation of different morphological and physiological states (Soll, 1997). Besides changes in the morphology of the colonies, phenotypic switching also affects levels of resistance to antifungals, levels of proteinase secretion, adhesion, hyphal formation and sensitivity to neutrophils. Due to this influence over other virulence traits, phenotypic switching is referred to as “higher order” virulence attribute (Mavor *et al*., 2005).

Both processes, dimorphic transition and switching, confer on *C. albicans* the ability to generate variants which allow a better selective adaptation to changing environmental conditions, and particularly to evade the host immune system (Gozalbo *et al*., 2004).

2.5. Other attributes

Other traits have been implicated as important for the pathogenicity of *C. albicans* (Mavor *et al*., 2005). Increased antimicrobial drug resistance, for example, helps a pathogen evade clearance in the presence of antifungals. In *C. albicans*, resistance mechanisms include the mutation of the antifungal targets themselves and the upregulation of some multidrug transporters. Besides, several *Candida* species, including *C. albicans*, have an antioxidative stress response consisting of the production of antioxidants that defend them against reactive oxygen species during phagocytosis. In addition, the ability to assimilate iron is necessary for survival of all pathogens in the host and could hence be called a persistence factor, being frequently considered a virulence determinant.

The virulence factors required by *C. albicans* to cause infections may well vary and show distinct complex interactions depending on the type of infection, the site and stage of infection, and the nature of the host response.

3. Genotyping of *Candida albicans*

The typing of *C. albicans* at the strain level has become crucial for medical mycology. For this species, as for all microbial pathogens, a reproducible and discriminatory strain typing system is of benefit for clinical and epidemiological studies to provide information on sources, carriage, and transmission of infection and on relations between strain types and properties, such as virulence and antimicrobial resistance (Odds *et al*., 2006, Tay *et al*., 2005). Thus, studying the relatedness of clinical strains is relevant in clinical management (Chong *et al*., 2007, Chen *et al*., 2005).

A number of different approaches for typing *C. albicans* isolates have been developed since the early 1980s. Molecular typing methods should be reproducible, discriminatory, high throughput, easy-to-use, digitally portable and amenable to standardization and library typing (Soll, 2000). Most molecular typing methods are based on techniques such as pulsed field gel electrophoresis (PGFE), restriction fragment length polymorphism (RFLP), polymerase chain reaction (PCR), and sequencing.

Each of the developed methods described next has its own set of assets as well as limitations, ensuring the continuous search in this field aiming to overcome them.

Multilocus Enzyme Electrophoresis

Multilocus Enzyme Electrophoresis (MLEE) is a technique based on electrophoresis of total proteins under native conditions and subsequent visualization of the enzymes present in the gel by specific enzyme staining procedures. It has been extensively used to fingerprint *C. albicans*, among other *Candida* species, and was shown to be useful to assess the mode of reproduction of microorganisms (Pujol *et al*., 1997; Soll, 2000). The main drawback to this method is that, being an indirect analysis of DNA performed through analysis of proteins, only coding regions are monitored and it is not possible to differentiate DNA sequences that code for the same protein. Also, minor aminoacid variation may not affect electrophoretic mobility of the protein. On the other hand, MLEE is relatively laborious and time consuming, because it is necessary to combine data from at least ten enzymes that provide variability among isolates (Soll, 2000). For instance, in a study of clinical *C. albicans* strain discrimination using MLEE by Pujol *et al*. (1997), 21 enzymes were tested but only 13 exhibited variability and were therefore used in the analysis.

Pulsed field gel electrophoresis

For more than two decades, several PFGE-based typing methods have been widely used, including karyotyping (Chen *et al*., 2005). This analysis is based in the differential migration of entire chromosomes with different sizes in an agarose matrix under the influence of an electric field. After DNA fragments are separated according to size they are visualized by ethidium bromide staining. The high intraspecific variability of karyotypes of fungi makes this technique discriminatory and useful for population studies in pathogenic yeasts, particularly *C. albicans*. Because of the high chromosome instability of this species there are some questions regarding reproducibility of the method. Additionally, this technique is time-consuming so it is no longer frequently used for *C. albicans* strain typing (Soll, 2000).

Another related approach is PFGE of restriction fragments, also called restriction endonuclease analysis (REA) of genomic DNA. It consists in electrophoresis of fragments generated by rare-cutting restriction endonucleases, such as *BssH*II and *Sfi*I useful for investigating microevolution of *C. albicans* clinical strains (Shin *et al*., 2005). REA may be a more sensitive method than karyotyping in the investigation of *C. albicans* infections but the use of more than one restriction enzyme may be necessary for optimal strain discrimination (Voss *et al*., 1995).

Restriction fragment length polymorphism

Restriction fragment length polymorphism (RFLP) techniques assay variations of the genome using restriction endonucleases that recognize and cut specific DNA sequences. There are basically two variants of this method; the first is called direct RFLP because it goes without probe hybridization, and the second is indirect RFLP because it is followed by specific probe hybridization.

One of the first DNA fingerprinting methods used to assess strain relatedness in a variety of infectious fungi, including *C. albicans*, was RFLP without any probe hybridization (Vazquez *et al*., 1991, Voss *et al*., 1995). In this technique, DNA extracted from spheroplasts is digested with one or more endonucleases, and

separated by electrophoresis in an agarose gel. The banding pattern of digested DNA is then visualized, usually by staining with ethidium bromide. This pattern is based on different fragment lengths determined by the restriction sites identified by the particular endonucleases employed. Low-intensity bands in most RFLP patterns are poorly resolved and, therefore, comparisons usually rely only on differences between intense bands, representing mainly ribossomal and mitochondrial DNA sequences. These fragments do not provide enough information to assess the relatedness of moderately close isolates. Even so, there is abundant evidence from the many RFLP studies of *C. albicans* that the method has been successful in identifying the same strain in independent isolates and in distinguishing among unrelated isolates.

The RFLP pattern of eukaryotic cell DNA as described above is poorly resolved primarily because all restriction fragments are stained. To allow visualization of particular fragments in the pattern, one can probe a Southern blot of the RFLP gel with a labelled DNA sequence that recognizes one or more fragments as a result of sequence homology. Any region of DNA can be used, including repetitive regions, if variations can be visualized with probes. Ribossomal and mitochondrial DNA probes have not been generally used in broad epidemiological studies of the infectious fungi mainly because of the presence of homologous genes and low discriminating power (DP). To date, the most successful and popular hybridization probes for the fungi have been cloned fragments containing repetitive genomic sequences. In a Southern blot of endonuclease digested genomic DNA, such a probe will hybridize to repetitive sequences dispersed throughout the genome, thus identifying variability among isolates at a variety of dispersed loci. It will also hybridize to additional sequences that are less variable, including sequences that vary as a result of allelic polymorphisms. Finally, it will hybridize to some hypervariable sequences, revealing microevolutionary changes within a strain. The virtue of the complex probe is that all this information is provided by a single Southern blot hybridization pattern.

Two complex probes for *C. albicans*, 27A and Ca3, were cloned approximately at the same time in the late 1980s and subsequently found to be related, although not identical. These have been used in a number of studies with *C. albicans*

(Shmid *et al*., 1993; Schroppel *et al*., 1994; Ruiz-Diez *et al*., 1997; Marco *et al*., 1999; Boccia *et al*., 2002; Taylor *et al*., 2003). Ca3 probe generates a pattern that is, on average, more complex than the one generated by 27A, and has proven to be reproducible and highly amenable to computer-assisted analysis. Databases have been established for comparisons of the Ca3-generated patterns of strains from different studies and for retrospective analyses. However, this methodology is technically demanding, not amenable to high throughput sample processing. In general, RFLP with hybridization with *C. albicans*-specific probes is very informative but time-consuming since Southern blots are needed (Soll, 2000).

Polymerase chain reaction

The techniques referred above are time consuming and not always compatible with the imperatives of the clinical laboratory. Due to their rapidity, PCR-based methods generally seem more appropriate.

Randomly amplified polymorphic DNA analysis

Although a variety of PCR-based strategies have been developed for DNA fingerprinting purposes, randomly amplified polymorphic DNA analysis (RAPD) has been one of the most commonly used for the infectious fungi. RAPD assay has become one of the most favourable choices for DNA fingerprinting of medically important *Candida* species and has been extensively applied in *C. albicans* strain typing (Metzgar *et al*., 1998).

In this technique, genomic DNA is amplified with a single short (approximately 10 bases) primer with an arbitrary sequence and products are separated on an agarose gel and stained with ethidium bromide. Although a single primer can generate a relatively complex pattern, in most cases, it provides only one to three intense bands differing among isolates. Therefore, it is usually necessary to select a number of primers, run independently for each test isolate, and combine the information. This strategy is illustrated by the work of Pujol *et al*. (1997), who

tested 40 random primers on a limited number of *C. albicans* test isolates and selected 8 that provided maximum variability.

RAPD analysis is technically simple and often detects variation among isolates that are invariant with RFLP analysis. Of the currently available *Candida* genotyping techniques, RAPD is relatively cost-effective and matches the resolving power of PFGE karyotyping, use of Ca3 complex probes and MLEE (Pujol *et al*., 1997; Soll, 2000). This, together with the availability of computer-assisted software systems that generate dendrograms of genetic relatedness among *C. albicans* isolates, has significantly contributed to the understanding of infective episodes and asymptomatic carriage of *C. albicans* (Samaranayake *et al*., 2003). The RAPD method has also gained favour because it is less time-consuming than Southern blot hybridization based methods. Unfortunately, the RAPD technique is poorly reproducible due to the low annealing temperatures used in the PCR. The problem of reproducibility exists not only among laboratories but within a laboratory over time, making the development of a common database difficult. Besides, the patterns obtained are often complex and sometimes difficult to interpret.

Amplified fragment length polymorphism

One promising modification of RAPD is amplified fragment length polymorphism (AFLP). This method selectively amplifies restriction fragments of genomic DNA. Amplification is achieved by using the restriction sites for the annealing of primers, and fragment selection is achieved by adding selective bases to the 3' end of the primers. By using stringent reaction conditions for primer annealing, the reliability of the method has been reported to be superior to that of the RAPD method. Careful selection of the restriction enzymes and nucleotides at the primer 3' ends results in a complex fingerprint pattern in a sequencing gel with yeast DNA as template (Soll, 2000).

Sequencing

Cloning and sequencing genes used to be a slow, technically demanding, and expensive undertaking, beyond the technical capacity of most medical mycologists interested in DNA fingerprinting of large collections of isolates. Recently, the emergence of PCR, automated DNA-sequencing technologies, and gene data banks, provided sequence based fingerprinting strategies amenable and affordable for large epidemiological studies.

Multilocus sequence typing

Multilocus sequence typing (MLST) is a sequencing-based technique that consists in the analysis of nucleotide polymorphisms of internal fragments of housekeeping genes (Maiden *et al*., 1998). DNA sequences from six or seven gene fragments are compared in order to establish the level of similarity between isolates. MLST is a highly discriminatory approach to distinguishing strains within a microbial species and has been widely used for epidemiological purposes. Because it was first developed to discriminate bacteria, analysis of diploid organisms such as *C. albicans* can be subjective. Today, MLST schemes for several fungi are available, and the approach is well developed for the typing of *C. albicans* strains (Bougnoux, *et al*., 2002; Tavanti *et al*., 2003).

Results from MLST are comparable with those obtained by DNA fingerprinting with the moderately repetitive sequence Ca3, since both typing approaches assigned the same sets of isolates to the same clusters of highly related strains in a study by Odds *et al*. (2006). Chowdhary *et al*. (2006) have shown that MLST has a higher discriminatory power, compared to Ca3 fingerprinting, in describing the genetic relatedness among clinical *C. albicans* isolates. When compared to other typing methods, MLST offers several advantages: it is rapid, reproducible and amenable to automation, provides an objective comparison of molecular typing data among laboratories, permits the exchange of that data via the internet to support local and global epidemiological studies, is easily implemented into any laboratory with DNA sequencing capacity, and supports the creation of a

comprehensive MLST database of *C. albicans* nucleotide polymorphisms. To assess the value of MLST relative to those of other DNA fingerprinting tools for discriminating among strains of *C. albicans*, Robles *et al*. (2004) applied it to a previously well-characterized set of *C. albicans* isolates evaluated by RAPD, MLEE, and Ca3 southern hybridization probe techniques. These results demonstrated that MLST is a highly effective technique that performs at least comparably to other established DNA fingerprinting techniques. Nevertheless, because it is necessary to combine information from at least six genes, this is not suitable to type a large set of isolates.

Microsatellite genotyping

Another class of molecular typing methods includes techniques that directly analyse the polymorphism of microsatellite markers. These markers are stable, easy to assay, adaptable to a large series, and discriminatory enough to be used as a typing system to investigate clinical issues. Because of their high levels of allelic variation and their co-dominant character, they deliver more information than any other marker system, thus reducing analysis time and costs. Besides, since microsatellites test the presence of different alleles at a given locus, distinguishing heterozygotes in diploid organisms such as *C. albicans* is possible in contrast to the RFLP and RAPD methods.

The most efficient type of microsatellite analysis consists in the amplification of the repetitive region by PCR with the use of primers complementary to conserved flanking regions. After, amplification products of different length are sized in order to determine the number of repetitions of the motives. Various methods to size these alleles can be used but the majority is based on electrophoresis in which the size determination of DNA molecules relies on calculation of the mobility in the gel matrix relative to an internal standard (Christensen *et al*., 1999). This can be performed using simple conventional laboratory equipment, electrophoresis with agarose or polyacrilamide gels, or using a more expensive but more sensitive and rapid method: capillary electrophoresis system using laser fluorescence detection in an automatic DNA sequencer (Deforce *et al*., 1998).

Microsatellite analysis has been extensively used for a wide range of biological questions, namely mapping and positional cloning of genes, population studies (structure, diversity and epidemiology), evolutionary studies, phylogenetics, linkage analysis, forensics, molecular pathology, paternity testing, relatedness and identification of individuals (Naidoo & Chetty, 1998; Schlötterer, 2000; Buschiazzo & Gemmell, 2006; Eloy *et al*., 2006).

4. *Candida albicans* microsatellite markers

Several studies have already reported the application of microsatellite markers for the genotyping of *Candida albicans* (as we will see below). At least 17 microsatellite loci have already been characterized for this pathogenic yeast species, providing adequate targets for the molecular typing of its strains.

In 1996, Field and co-workers described seven *C. albicans* microsatellite markers (ERK1, ZNF1, CCN2, MNT2, CPH1, EFG1 and EFG2) consisting of small repetitive regions from coding sequences and in which most of the repetitive motifs coded the same aminoacid, glutamine. Although these presented low polymorphism, this study revealed the great abundance and potential of microsatellite markers in *C. albicans*. ERK1 polymorphism was confirmed in another study aiming to assess specific site polymorphisms in *C. albicans* isolates from HIV infected patients, receiving fluconazole for fungal prophylaxis. This analysis allowed three different scenarios to be discerned: strains can remain identical, be replaced by clearly different strains, or undergo small changes (Metzgar *et al*., 1998).

In 1997, a microsatellite marker called CEF3 was described. It was found in the upstream sequence of the elongation factor 3 gene (EF3), between a putative TATA box and the transcription start site, in chromosome 5 of *C. albicans*. When applied to the study of sixty *C. albicans* isolates, it showed higher level of polymorphism than the previous ones (Bretagne *et al*., 1997). This marker was used in many posterior studies. Lott *et al*. (1999) compared independent isolates of *C. albicans* from two different geographic regions using both CEF3 and ERK1 and concluded that there was no indication of geographical partitioning. Dalle *et al*.

(2000) compared bloodstream and nonbloodstream strains of *C. albicans* using CEF3 concluding that they have a heterogeneous structure at this locus with three major and multiple minor allelic combinations. This group found an undescribed allele as well as seven new combinations. In 2003, the same team characterized *C. albicans* strains from healthy individuals using CEF3 and compared them with strains from nonhealthy individuals showing an overall similarity that suggested the ability of all commensal strains to develop as pathogens (Dalle *et al*., 2003). Lott & Effat (2001) in an attempt to characterize special groups of *C. albicans* isolates, combined information obtained from the analysis of five microsatellite markers (and other loci), three of them already described: ZNF1 (zinc finger transcription factor), ERK1 (extracellular-signal-regulated kinase), CEF3, and two new ones: A3 (Anonymous locus 3) and A4 (Anonymous locus 4) that presented very low polymorphism.

Microsatellite markers CDC3 and HIS3, present near coding regions and named after the gene they stand by, were described and applied together with CEF3 (Botterel *et al*., 2001). With the aim of obtaining a rapid genotyping method of *C. albicans*, these three markers were investigated by multiplex PCR and revealed a combined discriminatory power (DP) of 0.97. The same three marker combination was used again to evaluate the colonization of *Candida* species and the importance of *C. albicans* cross-contamination by Stephan *et al*. (2002). Clinical specimens obtained from surgical patients who had a high risk of yeast colonization were screened and it was concluded that acquisition of *C. albicans* in the surgical intensive care unit (ICU) seems to be mainly endogenous.

In a population study of *C. albicans* isolates from USA, Europe and Asia, two new markers were described: KRE6 and LOC4 (previously referred by Lott *et al*. (1999)) by Fundyga *et al*. (2002). These microsatellites were located in coding regions and exhibited low polymorphism. This study also included CEF3, ERK1, and ZNF1, assessing genotypes at five loci to address the question of genetic variation of strains isolated from different human hosts and geographic distinct regions.

In 2003, Sampaio and co-workers investigated the polymorphism of a new microsatellite locus, CAI, located in chromosome 4 of *C. albicans*. It was revealed

to be species-specific and showed a low mutation rate, since no amplification product was obtained when testing other pathogenic *Candida* species and no genotype differences were observed when testing over 300 generations. This new microsatellite was very polymorphic presenting the greater individual DP reported to date (0.97) the same obtained with the multilocus analysis described above, proving to be a valuable tool to differentiate *C. albicans* strains. Furthermore, when compared to other molecular genotyping techniques, CAI proved to be very simple, highly efficient, and reproducible, being suitable for low-quantity and very-degraded samples and for application in large-scale epidemiological studies.

Two years later, the same group described five new *C. albicans* microsatellite loci (CAIII, CAIV, CAV, CAVI, and CAVII) and developed a multiplex strategy to differentiate *C. albicans* strains using CAI, characterized earlier, and the two new microsatellites: CAIII and CAVI (Sampaio *et al.*, 2005). The DP obtained by combining the information generated achieved 0.99, the highest value ever reported. The multiplex PCR was later used to test *C. albicans* clinical isolates. The analysis of microsatellites by this multiplex PCR strategy was found to be a highly efficient tool for the rapid and accurate differentiation of *C. albicans* strains and adequate for the identification of fine microevolutionary events that could be related to strain microevolution in response to environmental stress conditions.

Besides their potential as molecular markers, it has been suggested a significant role of microsatellite sequences as relevant genomic information. Therefore, microsatellites are the focus of this work.

5. Microsatellite DNA

A great percentage of genomic DNA of all living organisms is made up of repetitive sequences, which fall into two classes: transposable elements-like sequences, which move around the genome, and internally repetitive sequences, such as satellite DNA, which belong to a group of genomic sequences known as Variable Number of Tandem Repeats (VNTR). Microsatellites are short DNA sequence stretches in which a motif of one to six bases is tandemly repeated and minisatellites are sequences composed of longer (10-100 bases) repeated motifs.

Microsatellites, also known as STR (Simple Tandem Repeats), SSR (Simple Sequence Repeats) and SSLP (Simple Sequence Length Polymorphism), are present in both Eukaryotes and Prokaryotes, and can be found either in protein-coding or non-coding regions. Microsatellite sequences are stably inherited and hence highly conserved from one generation to the next. Although they are unique to an individual and the same in all cells from the same individual, they exhibit high level of allelic variation (polymorphism) among individuals (Tautz, 1989).

Polymorphism is a crucial feature of microsatellites that makes them powerful and highly versatile genetic markers. It is predominantly manifested as changes in the number of the repeated motif caused by mutations that typically occur at rates that are orders of magnitude greater than single-nucleotide point mutations, from 10^{-7} to 10^{-2} events per locus per generation. These rates are influenced, among other factors, by environmental conditions, repeated motif, allele size, interruptions in the microsatellite, recombination rate, transcription rate, chromosome position, GC content in flanking DNA, and genotype.

Microsatellites were usually considered selectively and evolutionarily neutral sequences that are randomly distributed over the genome. However, conflicting results about their distribution are documented (Levinson & Gutman 1987, Schlotterer, 2000). Numerous lines of evidence available today suggest that microsatellite genomic distribution is not random and that it is dependent of many factors, namely location and repeat type (from mononucleotides up to hexanucleotides). Repeats are rarer in coding regions than in non-coding regions and this is probably attributable to negative selection against frame-shift mutations (Metzgar *et al.,* 2000). Tóth *et al.* (2000) conducted a detailed analysis that revealed highly taxon-specific patterns in the distribution of different repeat types in coding and noncoding sequences. Although the majority of microsatellites are composed of mono- and dinucleotide repeats (Dieringer & Schlötterer, 2003), tri-nucleotide repeats (and motif lengths that are multiples of three) are overrepresented in coding sequences. This happens because they can be accommodated more readily within coding regions as change in their length simply results in gain or loss of a single amino acid from a protein sequence. Thus, long mono- and di-tracts are almost exclusively distributed in non-translated regions

(Field & Wills 1998). Also, Katti *et al*. (2001) and Dieringer & Schlötterer (2003) found dramatic differences among repeat types within and between species. The presented examples demonstrate various non-random patterns of microsatellite variation that call for functional interpretation.

Microsatellites used to be commonly regarded as DNA sequences with no significant role as genomic information. Questions concerning the role of these sequences arose because of their frequent presence in highly economical genomes such as those of yeasts. Since then, the functional significance of a substantial part of microsatellites has been proven in various biological phenomena namely chromatin organization, and regulation of both DNA metabolic processes and gene activity.

In many cases, microsatellite repeat number appears to be a key factor for gene expression and expression level. Kashi & King (2006) stated that whatever the role and location of a microsatellite, changes in its number of repeated motifs can modulate its genetic function. When in coding regions, trinucleotide microsatellites variation in the number of repeated motives (codons) results in a variation in the length of homopolymeric stretches of amino acid affecting protein properties (Richard *et al*., 1999; Hancock & Simon, 2005). Mutations that change the number of repeats in coding non-triplet microsatellites cause frameshifts, which can effectively inactivate gene expression or encode different protein sequences. Because frameshifting based on microsatellite mutation is readily reversible by subsequent mutation, such microsatellites can function as on–off switches for their genes which is proven to happen in some bacteria (Li *et al*., 2004). Repeat variation commonly exerts a functional influence even when the microsatellites are located in noncoding sites where they do not affect protein structure directly.

Precise changes in microsatellites with high repeat numbers are long known to regulate critical virulence factors in several prokaryotes (Field & Wills, 1996). The presence of microsatellites in these organisms is rare, but most that do occur are related to pathogenic ones. In pathogenic bacteria, for example, infection processes require adaptation to several host environments and that is possibly achieved through microsatellite modulation, as has been described for a multitude of different genes (Li *et al*., 2004). Recently, Kashi & King (2006) showed the

results of a study that strongly implicates a mononucleotide-repeat polymorphism as a causal basis for differentiation in sporulation efficiency of yeast, a significant life-history trait for these organisms. Experimental manipulation of some microsatellites has demonstrated a linear correlation between repeat number and the extent of cell adhesion (Verstrepen *et al*., 2005). There is also evidence suggesting that the variation in repeat size of microsatellites coding for cell wall proteins of *S. cerevisiae* can alter its phenotype and may also transform the antigenic properties of the cell, thus allowing it to deceive the immune system of the host (Richard & Plaine, 2006).

In humans, instability of microsatellite may contribute to the development of cancer in several sites by causing either relatively small or drastic changes in the repeat structure. It has also been shown that many neurological disorders result from the expansion of unstable trinucleotide repeats located either in non-coding or in coding sequences (Katti *et al*., 2001).

6. *IFF* gene family

Microsatellite CAIV was investigated in this work. The first reference indicated that it was found in a non coding region of the *C. albicans* genome (Sampaio *et al*., 2005). Instead, new available data from Candida Genome Database (CGD) accounts for its location inside *IFF*8 gene, in chromosome 5. Analysis of the *C. albicans* genome has identified the *IFF* gene family (Individual Protein File Family F), with 12 genes, as encoding the largest family of cell wall related proteins. Ten genes (*IFF*1-9 and *HYR*1 (HYphally Regulated)) are predicted to have GPI anchors and two genes, *IFF*10 and *IFF*11, do not have any signal sequences for GPI anchor linkages (Richard & Plaine, 2006). Phylogenetic analysis of this family shows that the two non GPI anchored proteins *Iff10* and *Iff11* diverge obviously from the original group. No function has been established so far for any of the proteins, but their presence in 12 copies in the genome implies that they perform a useful function in terms of *C. albicans* biology, an idea reinforced again by the fact that *S. cerevisiae* has no clear ortholog.

Very little is known about this family and only two members, *HYR*1 and *IFF*11, have been the subject of published work (Bailey *et al*., 1996)). Nevertheless, it has been shown that this family is conserved in a range of other *Candida* species suggesting that the observed differences might be one of the reasons explaining the virulence variations between very close species (Bates *et al*., 2007).

A null mutant lacking *IFF11* was studied by Bates *et al*. (2007) and shown to be hypersensitive to cell wall-damaging agents, suggesting a role in cell wall organization. In a murine model of systemic infection this null mutant was highly attenuated in virulence, and survival-standardized infections suggest it is required to establish an infection. This work provides the first evidence of the importance of this gene family in the host-fungal interaction and virulence.

Although no clear function has been yet attributed to the referred proteins, the fact that most of them are putative GPI anchored proteins (GPIp) suggest that they are worthy of further investigation.

GPI anchored proteins are an abundant class of cell surface proteins present in both lower and higher eukaryotic organisms (De Groot *et al*., 2003). Their wide occurrence does not dictate specific functions, as many types of proteins are GPI anchored. The functions of GPIp have been extensively investigated and several roles have been assigned to it. Most are thought to be involved in cell wall biosynthesis and remodelling, hydrophobicity, antigenicity, flocculation, proteasic activity, sporulation, adhesion, mating, and virulence (De Groot *et al*., 2003). There is evidence demonstrating that a functional GPI anchor is required for normal cell wall structure and for full hyphal formation in *C. albicans*, and that perturbation of the GPI-anchor biosynthesis also results in hypersensitivity to host defences (Gozalbo *et al*., 2004). It has been shown that complete GPI anchors are required in *C. albicans* for full morphogenesis, virulence and resistance to macrophages (Richard *et al*., 2002; Richard & Plaine, 2006).

Albrecht *et al*. (2006) described two *C. albicans* cell surface-associated aspartic proteases, Sap9 and Sap10, which are GPIp. They have shown that Sap9 and Sap10 are crucial for the infection process. In contrast to the other SAP family members, they are GPIp necessary for cell surface integrity, cell separation, and adhesion.

Another family of GPIp, termed the fungal adhesins, is important for fungal pathogenesis permitting pathogens such as *C. albicans* to adhere to mammalian epithelial and endothelial cells. Expression of these proteins in *S. cerevisiae* permits this organism, which does not normally adhere to mammalian cells, to adhere to them (Guo *et al*., 2000).

7. Importance and aims of this study

Candida albicans is a serious human health concern, especially for the growing population of immunocompromised patients. Since pathogenicity and antifungal susceptibility often vary between strains, it is crucial to rapidly and accurately identify the strain causing infection so that the best treatment strategy is selected.
The aim of the present work is to study a new microsatellite locus, CAIV, located inside *IFF*8 gene of *Candida albicans.* CAIV characterization is focused on its polymorphism, allelic frequencies, genotypic frequencies and discrimination power. This molecular marker will be used to assess strain relatedness of a group of *C. albicans* clinical isolates from several Portuguese health institutions as well as to compare strains collected from two biologic products, blood and vaginal exudates. A multilocus approach using CAI and CAIV molecular markers will be evaluated according to the discrimination power and ability to detect microevolutionary events when multiple isolates from the same patient are compared.

Materials & Methods

1. Origin and culture of the clinical isolates

A total of 97 *Candida albicans* clinical isolates from five different Portuguese institutions were studied. The isolates were obtained from one health centre, Centro de Saúde Carandá Braga (CSC), two oncology hospitals: IPO Porto (IPOP) and IPO Lisboa, Instituto Nacional Dr. Ricardo Jorge Lisboa (INSA), and Hospital São João Porto (HSJ). Approximately half of the strains were collected from blood cultures and the others were obtained from vaginal exudates, in a total of 78 independent patients. Considering the number of clinical isolates studied this means that, in some cases, more than one isolate from the same patient was examined.

All the isolates had been previously identified by conventional phenotypic methods like the ability to produce germ tubes and by rapid biochemical identification kits, such as API-32C (BioMérieux, St Louis Mo.) or by YBC Card (Vitek Systems Inc., Hazelwood Mo.), according to the clinical laboratories of origin. Isolates were named after their institution of origin followed by a number in order to preserve the anonymity of the patients.

Isolates were received in plates with Yeast Peptone Dextrose Agar (YPDA) medium or Sabouraud Dextrose Agar (SDA) medium, and were subjected to macroscopic analysis to confirm their purity. Pure cultures of each strain were then preserved at 4ºC in YPDA tubes. For long term analysis, they were also frozen at -80ºC in 1.5 ml of a 30% (w/v) glycerol solution.

2. Microsatellite selection and Primer design

A search in *C. albicans* genome sequences, available in databases from Stanford's DNA Sequencing and Technology Center (http://www.sequence.stanford.edu/group/candida), was conducted for sequences containing microsatellite repeats. Previously, a search for repetitive sequences expected to have a very high degree of polymorphism was conducted based on two criteria: the number of simple repeat units (more than 20) and the location, outside a coding region. Following this search several microsatellites were

selected, characterized and used for *C. albicans* strain differentiation (Sampaio *et al*., 2003; Sampaio *et al*., 2005).

According to assembly 20 of *C. albicans* genome sequencing project, it was found that some of the selected sequences containing repetitive regions were, in fact, located inside genes. A microsatellite consisting of trinucleotide repeats, located in *IFF8* gene, which codes for a GPI anchored protein, was then chosen for further characterization in the present work and designated as locus CAIV.

To determine the chromosomal localization of the microsatellite the sequence selected was searched by BLAST against the latest release of *C. albicans* genome sequence, to give a location to a sequence contig (http://www-sequence.stanford.edu/group/candida).

Specific primers for the amplification of this locus were designed from the microsatellite flanking regions by using software Primer 3 available from www-genome.wi.mit.edu/cgi-bin/primer/primer3www.cgi. The conditions for primer design were that they should allow PCR amplification at an annealing temperature of 60ºC and that the amplified fragments should not exceed 500bp.

The forward primer was fluorescently labelled with FAM for detection in the automatic sequencer.

3. Characterization of the selected microsatellite locus

3.1. Genomic DNA isolation

The procedure followed was adapted from the method described by Kaiser *et al*., 1994. Cells were grown overnight in 15ml of YPD medium at 30ºC on a rotary shaker at 200rpm and then harvested by centrifugation at 5000rpm for 5 minutes at 4ºC. After ressuspending the cells in 1ml sorbitol buffer, 10µl of liticase were added. This mixture was incubated for an hour at 37ºC to break the cell wall. Another centrifugation at 5000rpm for 5 minutes was performed and spheroplasts in the pellet were ressuspended in 1ml of tris-EDTA. To break down the cell membranes, 30µl of 10% (v/v) SDS (sodiumdodecyl sulphate) solution was added. This suspension was heated at 65ºC for 30 minutes. After addition of 250µl of

potassium acetate (3M), the mixture was stored on ice for one hour for protein precipitation and centrifuged at 10000rpm for 10 minutes. To precipitate the nucleic acids an equal volume of cold isopropanol was added. The pellet thus obtained was subsequently washed twice with 70% ethanol (v/v) for 5 minutes. The dried pellet was finally ressuspended in 200µl of TE buffer (pH 7.4) and the DNA solution stored at 4ºC. From this solution, new ones were prepared with 25ng/µl for subsequent use in PCR reactions; these were stored at -20ºC.

3.2. PCR amplification

PCR reactions were performed in several independent isolates in order to evaluate the locus-specific amplification and its polymorphism.

This locus was amplified using the reagent mixture presented in Table 1. While preparing this mixture, 10% excess of all components were added to ensure sufficient volumes. 1µl of each DNA template solution (total 25ng) isolated as described above was added to 24 µl of the mixture.

The sequences of the primers used were:

F- 5'CCGCTTCTAATGATACTGGTGTT 3'

R- 5'TTTCCGTGGCATCAGTATCA 3'

Table 1. Reagents used to prepare the PCR amplification mixture.

Reagent	**Volume (µl)**
Pure water	17.8
Amplification buffer (10x)	2.5
$MgCl_2$ (25mM)	2.0
Primer F (50pmol/µl)	0.5 (25µM)
Primer R (50pmol/µl)	0.5 (25µM)
dNTPs mixture (10nM)	0.5 (µ)
Taq DNA polymerase (5U/µl)	0.2 (1U)

Amplification reactions were performed in the thermocycler (Icycler, BioRad) following the programme presented in Table 2.

Table 2. PCR programme for the amplification of locus CAIV

Number of cycles	Step	Step duration	Temperature (ºC)
1	Inicial denaturation	5 minutes	95
30	Denaturation	30 seconds	94
	Annealing	30 seconds	64
	Extension	1 minute	72
1	Final extension	7 minutes	72

The annealing temperature of 64ºC was chosen after analysis of a range of values, from 58ºC to 66ºC.

3.3. Fragment size determination

3.3.1. Polyacrylamide gel electrophoresis

PCR products were analysed by electrophoresis in 6% (w/v) polyacrylamide gel under denaturing conditions provided by addition of 6.5% (w/v) urea. Buffer systems used were described by Gusmão *et al*. (1997) for the Multiphor II electrophoretic system, and consisted of gel buffer Tris-HCl (0.375M, pH 8.8) and loading buffer Tris-glicina (0.125M, pH 8.8). DNA fragments obtained were visualized by silver staining. Gels were washed with 10% (v/v) ethanol solution followed by a 1% (v/v) nitric acid solution. After rinsing with water, gels were incubated with a 0.2% (w/v) AgNO solution in the dark. They were then submerged in a 0.28M NaCO solution with 0.02% (v/v) formaldeid to reveal the DNA fragments, and finally fixed with 10% (v/v) acetic acid.

3.3.2. Genescan Fragment determination

When PCR conditions were optimized the genotyping of all the clinical isolates was performed by Genescan analysis. DNA samples were prepared by adding to 1µl of PCR products, 14.7µl of formamide with 0.3µl of molecular size marker TAMRA GeneScan^{TM} 500 (Applied Biosystems). These samples were denatured by incubation at 96ºC for 5 minutes.

Samples were run in an ABI PRISM 310 genetic analyser (Applied Biosystems) and capillary electrophoresis was performed using POP4 polymer at 60ºC and

15Kv. Fragment size of PCR products was determined automatically using GeneScan 3.7 Analysis software (Applied Biosystems).

3.4. DNA sequencing

3.4.1. Allele sequence

The sequence of the alleles was determined in order to confirm the specific amplification of CAIV locus and to study the number of repetitions and the structure of the motif.

Single alleles were excised from polyacrylamide gels and eluted in 200µl of TE buffer, and subjected to three cycles of freezing (-20ºC) and towing (65ºC). These DNA fragments were reamplified in a PCR using the previous conditions and purified using Microspin S-300 HR (Amersham Biotech). 2µl of this product was added to 2µl of Mix BigDie Terminator v3.1 (Applied Biosystems) and 1µl of primer solution, either forward or reverse. The sequencing reaction took place in the thermocycler (Icycler, BioRad) following the programme described in Table 3.

Table 3. Programme used in the sequencing reaction of locus CAIV.

Number of cycles	Step	Step duration	Temperature (ºC)
1	Inicial denaturation	10 minutes	96
35	Denaturation	10 seconds	94
	Annealing	20 seconds	55
	Extension	4 minutes	60
1	Final extension	10 minutes	60

These products were purified using AutoSeqTM G-50 columns (Amersham Biotech) according with the instructions. 20 µl of formamide were added and the solution was denatured at 96ºC for 3 minutes. These fragments were separated by capillary electrophoresis in an ABI PRISM 310 genetic analyser (Applied Biosystems) using POP4 polymer at 50ºC and 15Kv. Analysis of the sequence was performed with Sequence Analysis software 3.1 (Applied Biosystems).

3.4.2. Gene sequence

The sequence of the final portion of gene *IFF8* where microsatellite CAIV can be found was also determined. PCR amplification of this region was performed as indicated above using primers previously published (Sampaio *et al*., 2005). The purification steps, sequencing reaction, and automatic analysis were performed as described above.

4. Differentiation of *Candida albicans* isolates using two microsatellite markers

The clinical isolates of *C. albicans* were also genotyped with marker CAI and other microsatellite markers described by Sampaio *et al*. (2005), to allow further comparative studies. Amplification and fragment size determination were performed as described previously.

5. Data analysis

Statistic analysis of the data obtained was performed by using software package GENEPOP, version 3.4 (Raymond & Rousset, 1995).

For each locus in each sample, several parameters were calculated: allele frequencies, observed and expected genotype proportions, observed and "expected" number of homozygotes and heterozygotes, the genotypic matrix. A table of allele frequencies for each locus and for each sample was also computed.

Population differentiation between groups of samples from two different biologic products was assessed. An unbiased estimate of the *P*-value was performed using a Markov chain method (Guo & Thompson, 1992).

Comparative analysis of the isolates was performed calculating genetic distances. Distance matrices were calculated from allele (or genotype) frequency of the locus CAIV and trees were constructed using UPGMA (unweighted pair-group analysis method using arithmetic averages) and visualized using software NTSYS-PC.

Discrimination power (DP) was calculated according to Simpson index (Hunter & Gaston, 1988), using the following expression:

$$DP = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{s} nj(nj-1),$$

where N is the strain number, s is the total number of distinct genotypes, and n*j* is number of strains with *j* genotype.

These parameters were studied both with CAIV data alone and with the combination of CAI and CAIV.

Results & Discussion

1. Specific amplification, genotyping and sequencing of locus CAIV

With the aim of identifying new highly informative microsatellite polymorphisms for molecular discrimination of *Candida albicans*, a previous search for short repetitive sequences was conducted as described in Materials & Methods. A microsatellite consisting of trinucleotide repeats, located in *IFF8* gene, which codes for a GPI anchored protein, was then chosen for characterization in the present work and designated as locus CAIV.

Microsatellite marker CAIV was selected from *IFF8* genomic sequence (ORF19.570) available at Candida Genome Database (www.candidagenome.org). This gene is located in chromosome 5 of *C. albicans* (164619, 166763) and it is composed of 2145 nucleotides. The microsatellite contains 24 trinucleotide repeated units AAC, interrupted by AAT and ACT (Figure 3).

Orf19.570, Assembly 20

…tcatcggttacttccagttcagacgcttctactgattccatcacttcagaaacagcttcatcttcatcaaccccattatctggtgattcatctcaa
gtatcttctttgacaactggaacatcaccagataccattgccagtttccaaaccgactctacttcttttggctttggctctggctctccttcttctggt
gctgtccaatcttctggtgttaccaatagtacaccaaatactggtgatgttaatactcaatctaatactgctaatatcgctacatctgataatactg
ctacatctaccgcttctaatgatactggtgttaatacagctacagctacaactactggtactggtactggtcctgacaacaacaacaacaaca
ataacaacaataacaacaataacaacaataacaacaataacaacaataacaacaataacaacactaacaacagtggtgtttctgcagc
tgattctaaagcaagtggtgatatttctactgtgacagcttcaagtactactttaatttccgtggcatcagtatcatcaacatatcctattgctaatg
aatcaagttcaccttcttcttcttcttcttcttcttcttcttctggtactcctggtgaagtgattccaaatgctaatggatcttctaaattatcaattggt
atgacattcatgattagtggatttgctacaatgtttgctttattcatg…

Figure 3- Parcial *IFF8* sequence from Candida Genome Database (ORF 19.570) showing the repetitive region corresponding to microsatellite CAIV (blue). Sequences of the primers used for the specific amplification of locus CAIV are underlined.

Using the primers designed, locus CAIV was amplified in only twenty independent strains in order to evaluate its polymorphism. Specific amplification of microsatellite locus CAIV with the primers designed revealed the presence of one or two PCR products for each strain. As *C. albicans* is a diploid species, the different PCR products were considered alleles of this locus, and the strains presenting only one fragment were homozygotic, while the ones presenting two fragments were considered heterozygotic.

Examples of the fragments obtained by PCR amplification of this locus in four *C. albicans* isolates are presented in Figure 4.

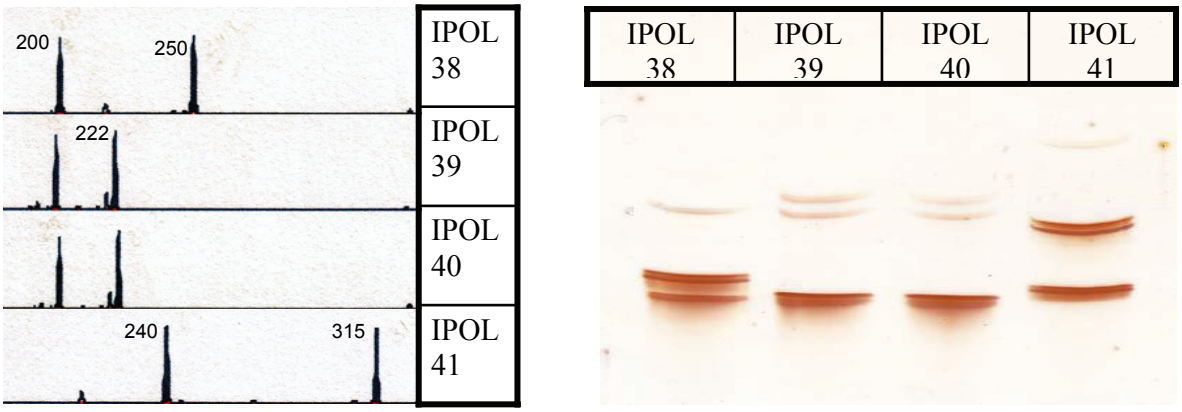

Figure 4. A. Genescan profiles depicting the results of automatic fragment sizing in four different strains, using marker CAIV. B. Denaturing gel electrophoresis of the fragments obtained by PCR amplification of the same *Candida albicans* strains.

However, genescan analysis of several clinical isolates using microsatellite CAIV revealed some amplification problems. Sometimes more than two fragments were detected and in other cases it was difficult to obtain amplification, indicating the possible presence of null alleles. These amplification problems may be related to a high polymorphism in primer annealing regions. Frequently, the sequences that immediately flank the repeats are liable to variability and sometimes this may be an obstacle to the design of adequate PCR tests (Lunel *et al.*, 1998). Therefore, both the electropherograms obtained from automated genotyping and the polyacrylamide gels were analysed and compared in order to avoid misinterpretations (Figure 4).

Some PCR artefacts can occur due to the addition of extra nucleotides by some *Taq* polymerases. This problem can be controlled by the systematic use of a reference strain in each experiment. Since the expected length is known, it is possible to detect an artefact and to correct the sizing of the alleles.

It was not possible to obtain the sequence of individual alleles excised from polyacrylamide gels. Thus, in order to determine allelic structure and establish the correspondence between the size (base pairs) of the alleles and their number of repeated units this microsatellite region was sequenced in some selected strains by using previously designed primers (Sampaio *et al.*, 2003). Due to several

problems raised with sequencing reactions it was only possible to obtain a clear sequence in a few strains. As shown in Figure 5, the results confirm the presence of the referred repetitive region and the expected structure of microsatellite CAIV composed of trinucleotide units AAC, interrupted by AAT and ACT.

It was also possible to observe some differences between the sequence obtained in the database (consensus sequence) and the sequence here represented. These can reveal that possible incorrections may exist in the database or simply that the strains studied are different, not only in the length of the repetitive sequences, but that there is also polymorphism outside the repetitive region.

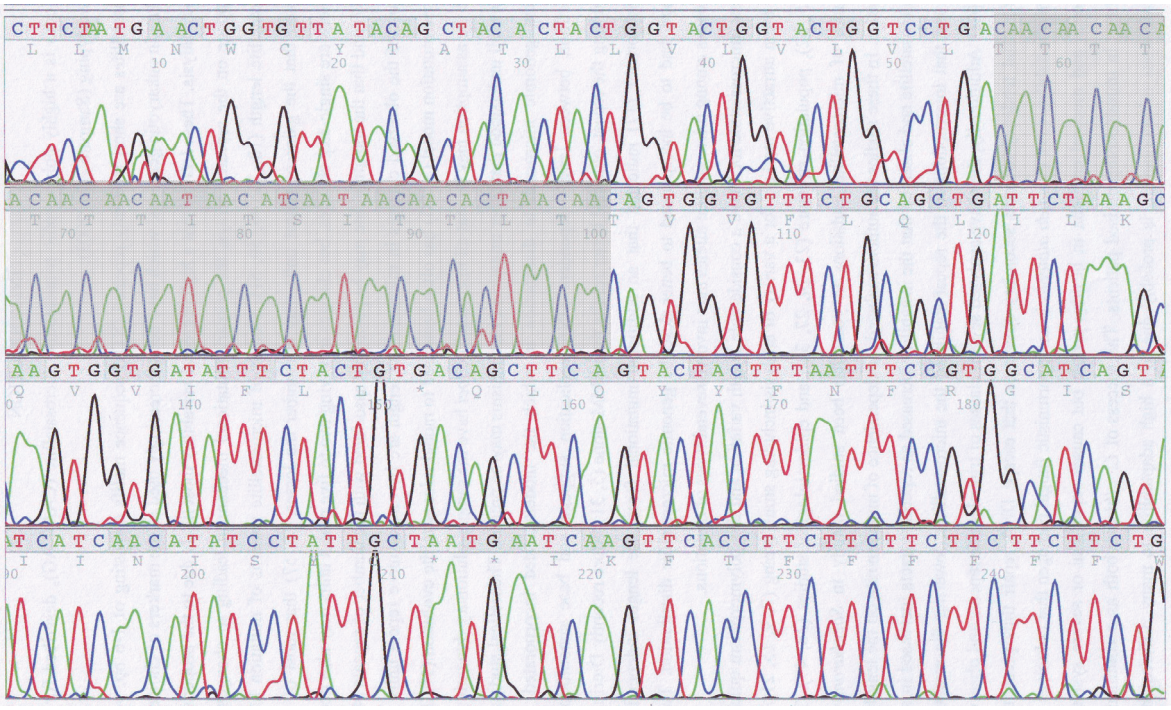

Figure 5. Partial sequence of *IFF8* gene from strain HSJ155 showing the repetitive region composed of AAC units. Analysis of the sequence was performed with Sequence Analysis software 3.1 (Applied Biosystems).

Thus, sequence analysis may reveal three levels of polymorphism: (1) the number of repeats, (2) the structure of the repeated region and, (3) point mutations outside repetitive regions. Here, we have only considered the first level but the other two may also play important roles in strain differentiation of *C. albicans*.

It was assumed that the variation in length was always due to differences in the number of trinucleotide repeat units and, therefore, the alleles were named according to the total number of repeats independently of the structure variation.

2. Analysis of locus CAIV

A total of 96 clinical isolates were genotyped for CAIV. For an easier and more accurate allele size determination, genotype analysis was performed automatically by using a fluorescent labelled primer.

CAIV revealed to be species specific, since no amplification products were obtained when using the described primers in the amplification of other clinically relevant *Candida* species. Due to time limitations it was not possible to test reproducibility for every sample and stability of the marker was not assessed either.

Allelic Frequencies

Only independent clinical isolates were considered for the genetic and genotypic analysis to avoid overrating some of the alleles due to their possible repetition in isolates from the same patient.

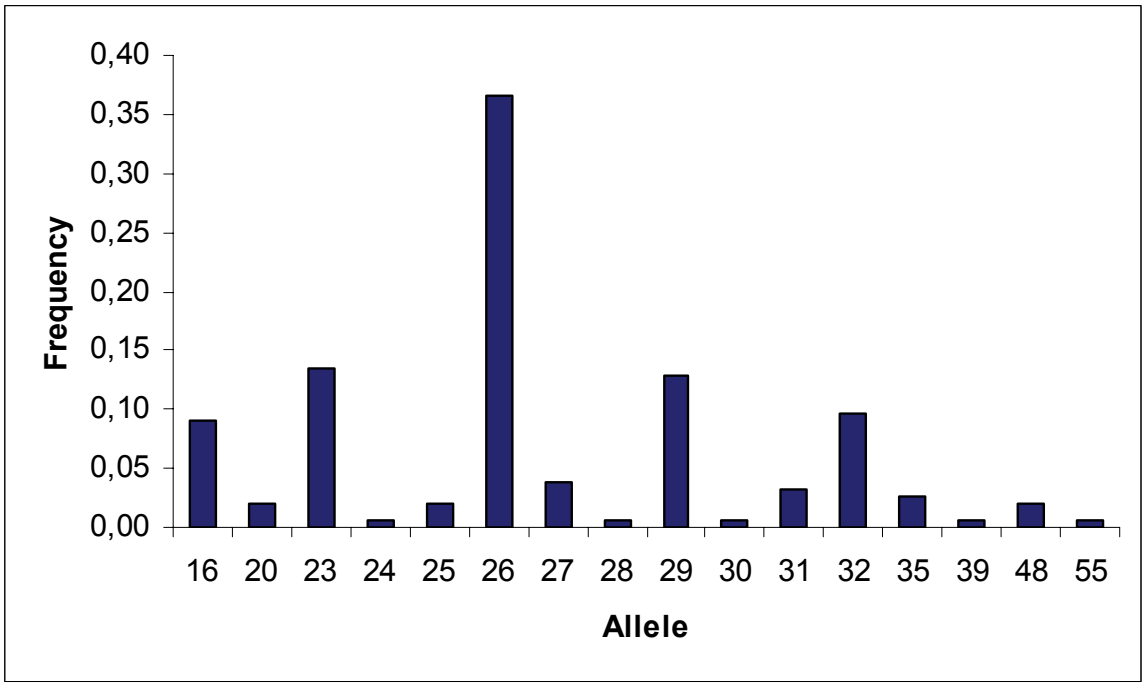

Figure 6. Frequency of CAIV alleles present in the 78 non-related *C. albicans* strains studied.

Among the 78 independent strains, 16 distinct alleles were found with molecular weights ranging from 200 and 316bp. Allelic frequencies varied from 0.006, when the allele was present only once, to 0.365 that corresponded to the allele being found in 57 isolates. The most frequent was allele 26, followed by 23, 29, 32 and 16 (Figure 6).

As referred in the Material & Methods the strains of *C. albicans* used in this study had been isolated from two distinct body products, blood cultures and vaginal exudates, representing deep-seated systemic infections and superficial infections, respectively.

In order to investigate if the allelic distribution was identical in these two groups of isolates an unbiased estimate of of the *P*-value of the probability test (or Fisher exact test) was performed, as described by Raymond & Rousset (1995), indicating a statistically significant difference in allelic distribution in the two groups ($p>0,005$).

The relative distribution of CAIV alleles present in the isolates from blood cultures and from vaginal exudates is shown in Figure 7.

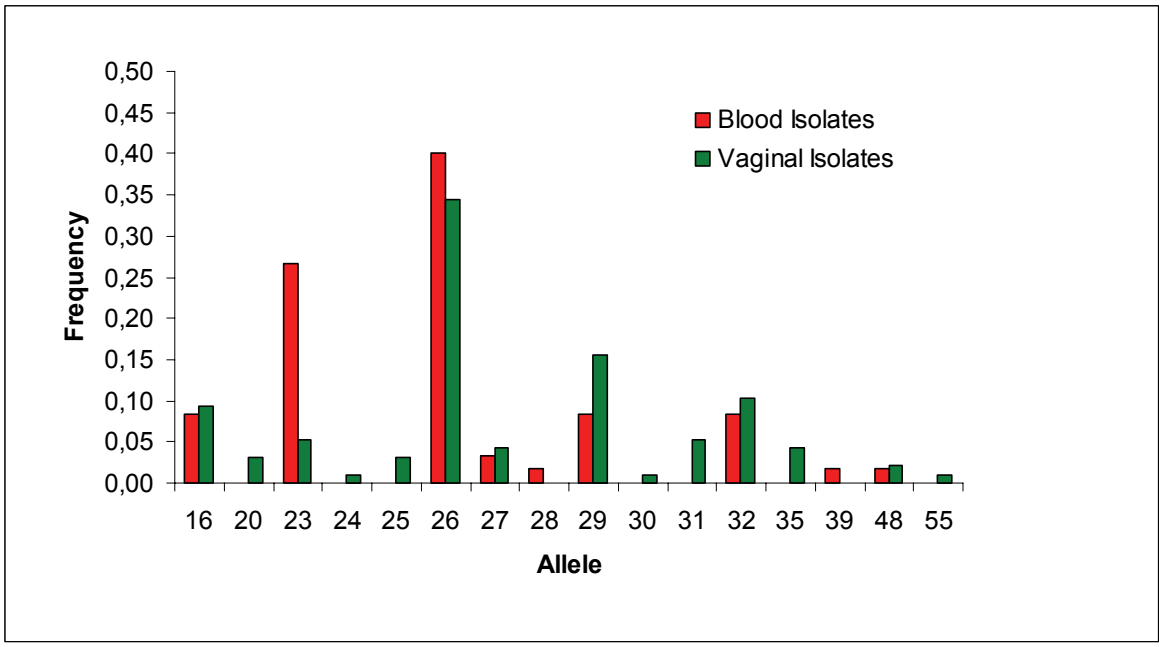

Figure 7. Relative frequencies of CAIV alleles present in isolates from different blood cultures and vaginal exudates from the group of 78 independent *Candida albicans* strains.

From the analysis of the figure it can be seen that several alleles were present only in vaginal isolates (20, 24, 25, 30, 31, 35, and 55) while alleles 28 and 39 were only found in blood isolates. The most frequent alleles are present in both groups of isolates but allele 23, the second most frequent, is overrepresented in the blood isolates.

Genotypic Frequencies

Table 4. CAIV genotypes observed in the group of 78 independent strains and respective observed and expected frequencies.

Genotypes	Number of Strains	Observed Frequency	Expected Frequency
16 - 16	1	0,013	0,587
16 - 20	1	0,013	0,271
16 - 23	5	0,064	1,897
16 - 26	2	0,026	5,148
16 - 27	1	0,013	0,542
16 - 32	3	0,038	1,355
20 - 20	1	0,013	0,019
23 - 23	7	0,090	1,355
23 - 32	1	0,013	2,032
23 - 39	1	0,013	0,135

24 - 31	1	0,013	0,032
25 - 25	1	0,013	0,019
25 - 31	1	0,013	0,097
26 - 26	20	0,256	10,297
26 - 28	1	0,013	0,368
26 - 29	12	0,154	7,355
26 - 31	1	0,013	1,839
26 - 48	1	0,013	1,103
27 - 27	2	0,026	0,097
27 - 30	1	0,013	0,039
29 - 29	2	0,026	1,226
29 - 32	1	0,013	1,935
29 - 48	2	0,026	0,387
29 - 55	1	0,013	0,129
31 - 35	2	0,026	0,129
32 - 32	5	0,064	0,677
35 - 35	1	0,013	0,039

In the group of 78 strains a total of 27 different genotypes were observed with frequencies between 0.013, for genotypes found only once, and 0.256, for genotypes found in 20 strains (Table 4). The most frequent genotype was 26-26 followed by 29-26, both combinations including the most frequent allele, 26. Other frequent genotypes were 23-23, 16-23, 32-32 and 16-32. Approximately half (40) of the observed genotypes were homozygotic which is clearly above expectations (14.4) as computed using Levene's correction.

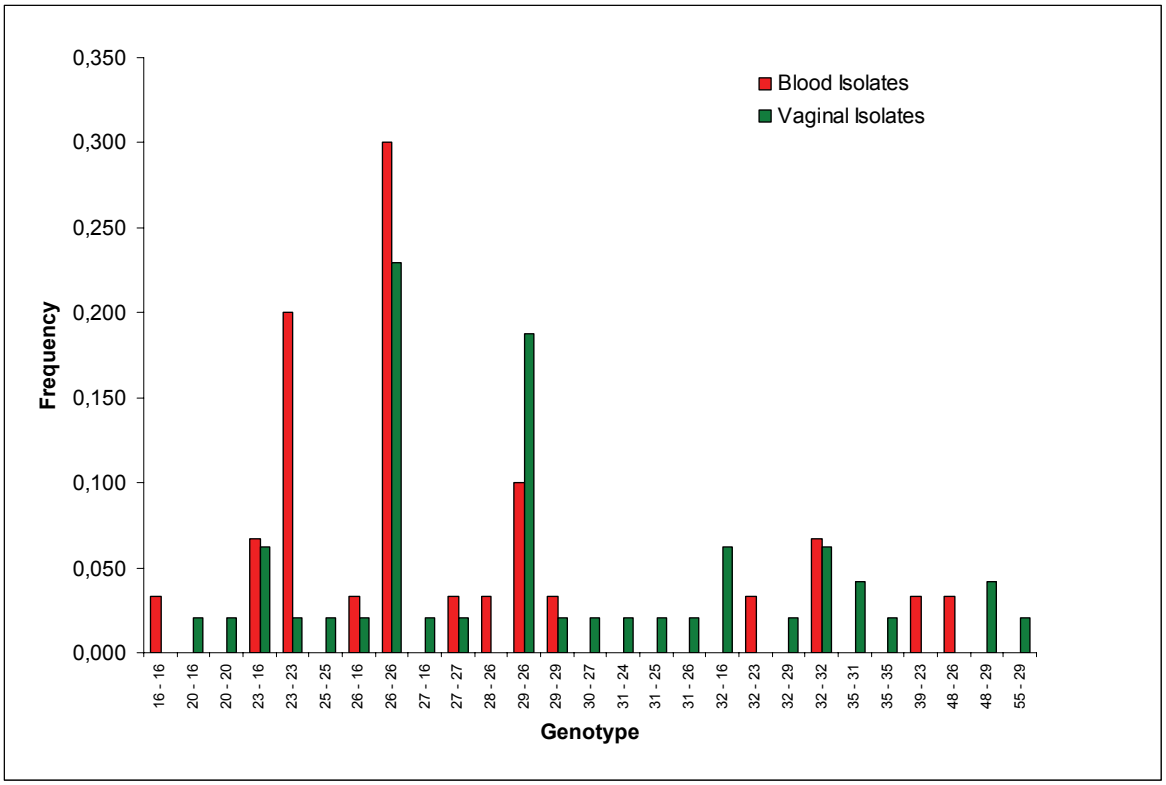

Figure 8. Relative frequencies of CAIV genotypes present in isolates from different biologic products (blood cultures and vaginal exudates) from the group of 78 independent strains.

The relative distribution of CAIV genotypes present in the isolates from blood cultures and from vaginal exudates is shown in Figure 8. From the analysis of the figure it can be seen that 14 of the genotypes were found exclusively in vaginal exudates and five were observed only in isolates from blood.

CAIV discriminating power

The discriminatory power (DP) of the microsatellite marker CAIV, calculated according to the Simpson index, was 0.90. This DP is comparable or even higher than the values reported for other microsallites located in coding regions. However, CAIV has relatively low polymorphism and DP when compared with microsatellite CAI, 0.97 (Sampaio *et al*., 2003).
In combination with other microsatellite markers this new microsatellite seem to be a valuable tool to differentiate *C. albicans* strains, and for application in large scale epidemiological studies.

Strain Grouping

The similarity among the unrelated strains was studied in order to determine if there was any association between the different isolates according to their origin. Although a significant difference was found in genotype distribution between the two populations studied, as it can be seen in Figure 9, strains isolated from vaginal exudates and from blood cultures were distributed by all the groups obtained.

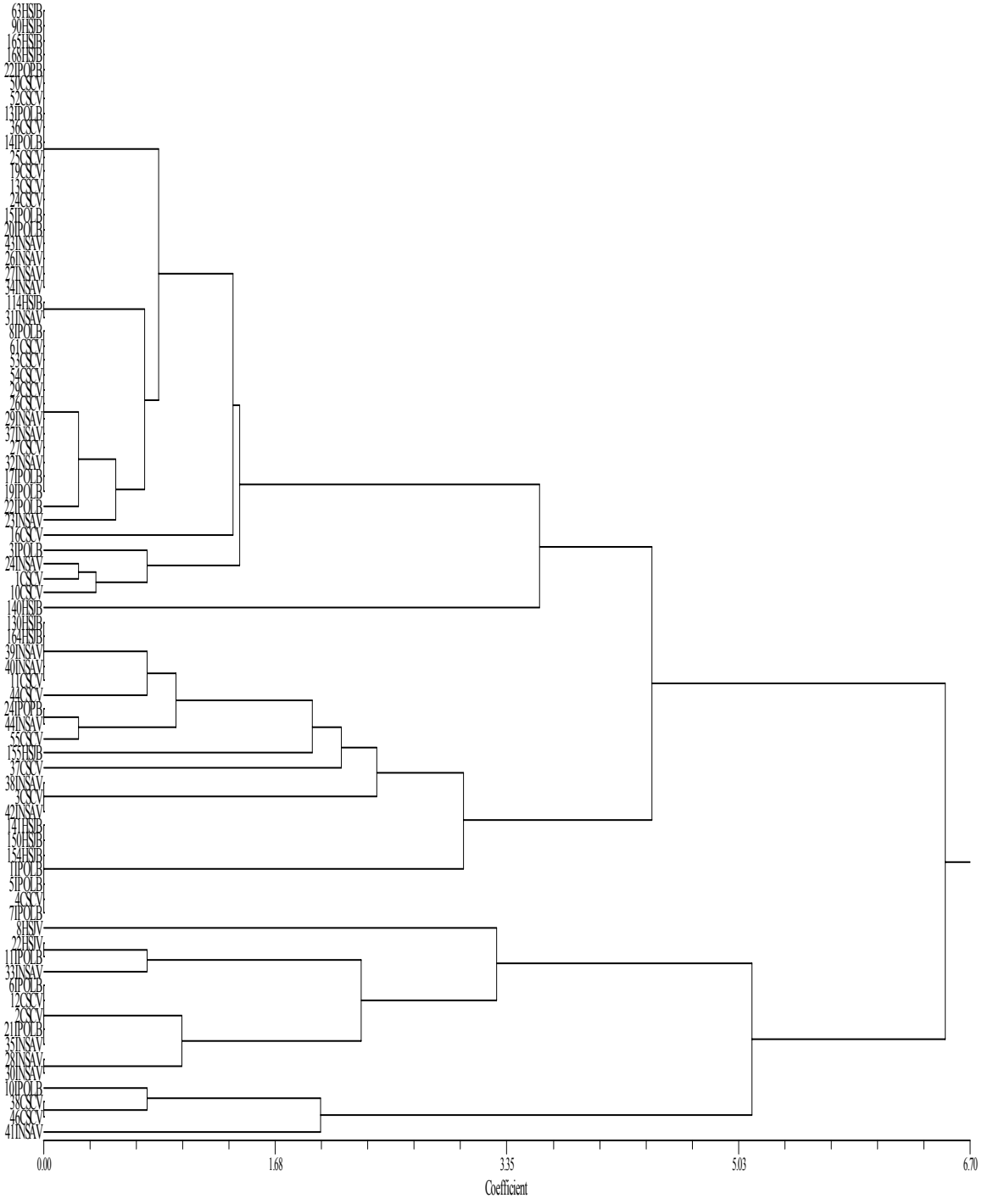

Figure 9. Dendrogram representing the grouping of clinical isolates based on CAIV genotypes. Genetic distances were calculated by using Dsw coefficient and the UPGMA aglomerative method. The isolates from blood cultures are designated by B and isolates from vaginal exudates by V.

3. Multilocus analysis

In order to complete their characterization, the *C. albicans* isolates were also genotyped with another microsatellite marker, CAI, previously described by Sampaio *et al*. (2003). This second locus was chosen because it has a very high DP (0.97) and also because it is located on a different chromosome so as to improve the chances of finding polymorphisms.

Genotypic Frequencies

Sixty-four of the 78 independent isolates in study were analysed simultaneously with markers CAIV and CAI and 58 different genotypes were observed, with frequencies between 0.0156 for genotypes found only once, and 0.0625 (4 strains). The most frequent genotype was CAI:21-25/CAIV:26-26 (Table 5). Most combinations were unique, indicating a very good discrimination provided by the simultaneous analysis with these two markers. In fact, the combined discriminatory power (DP) obtained with the two microsatellite markers was 0.996, indeed a very high value.

Table 5. Genotypes observed with microsatellites CAI and CAIV in the 64 independent *C. albicans* isolates and respective frequencies.

CAI	CAIV	Strains	Frequency	CAI	CAIV	Strains	Frequency
11-18	16-20	1	0,0156	21-26	26-28	1	0,0156
11-18	20-20	1	0,0156	21-26	26-31	1	0,0156
11-28	16-27	1	0,0156	22-34	26-26	1	0,0156
12-12	16-23	1	0,0156	**23-27**	**29-29**	**2**	**0,0312**
12-12	26-48	1	0,0156	24-26	32-32	1	0,0156
12-12	27-27	1	0,0156	25-25	16-23	1	0,0156
12-17	16-23	1	0,0156	25-26	16-23	1	0,015625
12-21	26-26	1	0,0156	**26-26**	**16-23**	**2**	**0,0312**
16-25	26-26	1	0,0156	26-26	16-32	1	0,0156
16-27	26-29	1	0,0156	26-26	23-23	1	0,0156
16-38	26-29	1	0,0156	26-28	16-23	1	0,0156
17-21	25-31	1	0,0156	26-33	32-32	1	0,0156
17-21	26-29	1	0,0156	26-35	16-32	1	0,0156
18-18	16-23	1	0,0156	27-27	26-29	1	0,0156
18-18	23-23	1	0,0156	27-27	29-48	1	0,0156
18-24	16-23	1	0,0156	27-30	27-30	1	0,0156
18-28	26-29	1	0,0156	27-34	26-29	1	0,0156
18-34	23-39	1	0,0156	27-42	26-29	1	0,0156
19-34	32-32	1	0,0156	27-47	26-29	1	0,0156
20-27	26-26	1	0,0156	27-49	26-26	1	0,0156
20-28	16-26	1	0,0156	27-50	26-29	1	0,0156
20-28	26-26	1	0,0156	28-47	27-27	1	0,0156
20-28	29-48	1	0,0156	29-29	16-16	1	0,0156
20-33	29-55	1	0,0156	30-32	35-35	1	0,0156
20-37	16-32	1	0,0156	30-48	32-32	1	0,0156
21-25	25-25	1	0,0156	32-46	29-32	1	0,0156
21-25	**26-26**	**4**	**0,0625**	35-44	26-26	1	0,0156
21-26	26-26	1	0,0156	39-46	26-26	1	0,0156
21-26	**26-26**	**2**	**0,0312**	40-40	26-26	1	0,0156

As shown in Table 5, there were no predominant genotypes and a large number presented a very low frequency. It is expected that this study has not uncovered the absolute number of alleles for the population at large because of the limited number of clinical isolates. However, increasing the total number of strains studied may not cause significant variations of the actual distribution pattern. Similar results were found by other authors using another set of microsatellite markers when comparing a group of bloodstream isolates with isolates from other provenances (Lott *et al*., 1999).

Strain grouping

The grouping of genotypes was performed as described previously but considering the two loci analysed, CAI and CAIV. The results are presented in Figure 10.

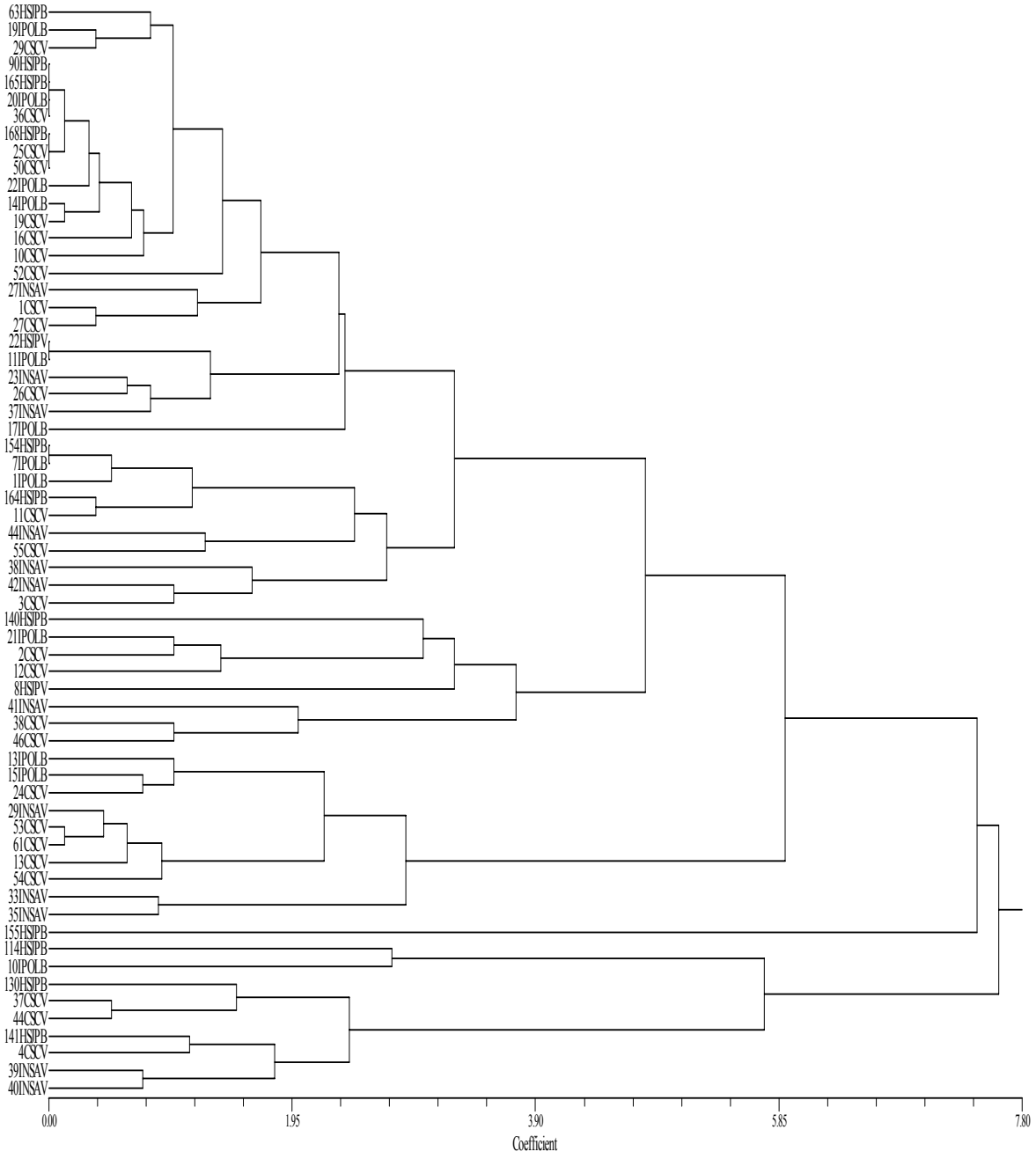

Figure 10. Dendrogram representing the grouping of clinical isolates based on CAI and CAIV genotypes. Genetic distances were calculated by using Dsw coefficient and the UPGMA aglomerative method. The isolates from blooculturesare designated by B and isolates from vaginal exudates by V.

Considering both markers analysed, it was observed that different groups were formed but there is no significant clustering of strains according to their origin. Nevertheless, some small groups with low coefficient consist only of vaginal or blood isolates, indicating that some genotypes might be better adapted to specific characteristics of these biologic products.

Use of CAIV microsatellite for *Candida albicans* strain differentiation

Genotyping with CAIV was used to differentiate the infecting *C. albicans* strains in cases of recurrent infections. In Table 6 are presented the results obtained in 12 of the studied cases using both CAI and CAIV microsatellites.

Of the thirteen patients with multiple isolates analyzed (Table 6), only two of the presented strains showed different CAI genotype (patients C and H). Regarding CAIV genotypes, the analysis revealed that multiple isolates from patients C, H, I, K and L were different. The two vaginal isolates from patient I presented the same CAI genotype but completely different CAIV genotypes, indicating that the relapse was caused by a different strain. The same is observed with patient H. It seems to have occurred strain substitution or some kind of evolutionary event. In patient L a microevolutionary event seems to have occurred since the strains presented a change at the CAIV locus, from genotype 23-32 to genotype 32-32. This occurrence may be the result of one genetic change that could lead to the LOH. Similar results were reported by other authors and it was suggested by Sampaio *et al* (2005) that this kind of microevolution may arise as an adaptative response of the strains to new environments, what may happen following fluconazole treatment. Because isolates from patient K have unknown CAI genotypes, it is difficult to study strain differences but it is possible that this is a case of loss of heterozygocity (LOH), assuming that CAI genotype are the same.

Table 6. Genotypes observed with microsatellites CAI and CAIV in different *C. albicans* isolates from the same patient.

Patient	Isolate	CAI	CAIV
A	HSJ 63	16-25	26-26
	HSJ 69		
B	HSJ 90	21-25	26-26
	HSJ 93		
C	HSJ 124a	18-34	23-39
	HSJ 140	18-34	
	HSJ 140a	**34-34**	
	HSJ 144	18-34	
D	HSJ 143	26-26	23-23
	HSJ 150		
E	HSJ 154	26-26	23-23
	HSJ 154b		
F	HSJ 164	26-28	16-23
	HSJ 164b		
G	HSJ 168	21-26	26-26
	HSJ 168a		
	HSJ 168b		
H	HSJ 7	**30-30**	35-35
	HSJ 8	30-32	35-35
	HSJ 9	30-32	**29-29**
I	HSJ 22	23-27	**29-29**
	HSJ 23		**32-32**
J	IPOL 1	25-25	23-23
	IPOL 2		
K	IPOL 3	unknown	**23-32**
	IPOL 4	unknown	**32-32**
L	IPOL16	16-38	**26-26**
	IPOL17		26-29
	IPOL18		26-29
M	CSC 36	21-25	26-26
	CSC 36e		

In summary, analysis of multiple isolates from the same patient, in recurrent infections, revealed that they are not always the same. Patients H, I, L had at least two strains with the same CAI genotype but different CAIV genotype, proving the usefulness of analysing multiple loci.

It has been suggested that multiple loci need to be monitored in order to reliably establish potential relatedness between strains of *C. albicans*. Moreover, Lunel *et*

al., (1998) showed that strains that are epidemiologically unrelated may still be identical for approximately 40% of their repetitive loci.

It is important to stress that the results here presented clearly show the need for a multilocus analysis in strain typing, even when highly discriminatory molecular markers like CAI are used.

Conclusion & Final Remarks

Among the yeasts that have emerged as major fungal pathogens in recent years, the commensal species *Candida albicans* is the most prevalent, acting as an opportunistic agent in immunocompromised patients. This yeast is responsible for a wide range of superficial infections but also for systemic life threatening conditions. As a commensal *C. albicans* is harmless, however if the balance of normal flora is disrupted or the host immune defenses are compromised the fungus can outgrow the mucosal flora causing superficial and invasive candidiasis. In severe cases *Candida* species may enter the bloodstream (candidaemia) and can penetrate almost all organs in the body. Cancer chemotherapy, neutropaenia, organ transplantation, indwelling catheters and devices, autoimmune diseases, burns, antimicrobial therapy, abdominal surgery, radiotherapy, and intensive care are among the main risk factors predisposing for infections by *C. albicans* and other *Candida* species.

It has been shown that the primary mode of genetic inheritance in *C. albicans* is clonal (Lott *et al.*, 1999; Dalle *et al.*, 2003). However, such clonal lineages may have variations in biological traits, such as pathogenicity, host adaptation, and antifungal susceptibility. Thus, it is crucial to a rapid and accurate identification of the strains causing infection. Several DNA-based methodologies have been used for strain identification, including RFLP analysis, PFGE and Southern blot hybridization with probes that hybridize with repetitive genetic elements. However, it has been demonstrated that PCR based ones have several advantages over the others. PCR is a less time consuming technique and automated procedures have a high throughput and should be suitable for screening large number of isolates with limited workload. Results can be easily reproduced and, since the lengths of the alleles are numeric data, it is possible to compare them between laboratories. Microsatellite markers are stable, easy to assay, and discriminatory enough to be used as a typing system to investigate clinical issues and yeast epidemiology. Furthermore, it can be used to complement studies based on other techniques, by which evolutionary related strains can be distinguished, particularly when searching for micro-evolutionary events.

Analysis of multiple microsatellite loci enables accurate typing, if high DP markers are selected and high-speed typing, if a multiplex system is developed. As a

multilocus analysis is more accurate in discriminating strains, in this work a new polymorphic microsatellite locus was described to be used, in combination with other markers, for *C. albicans* strain differentiation. This microsatellite, designated CAIV, revealed to be very polymorphic and 16 distinct alleles and a total of 27 different genotypes were found in 78 unrelated *C. albicans* isolates. The strains studied were obtained from several health institutions and were isolated from blood cultures or from vaginal exudates. It was possible to identify alleles and genotypes specific of each of the groups of strains and a statistically significant allele and genotype differentiation was found between the two groups of strains.
So far, no pathogenic genotype, i.e., no genotype with a propensity for bloodstream invasion, has been identified. However, we can not rule out the possibility that among commensal *C. albicans* strains some are more prone than others to undergo increased mucosal proliferation in nonhealthy individuals. A study comparing series of strains with different origins, including bloodstream isolates, was performed previously but the isolates were obtained only from nonhealthy individuals and no significant differences were observed (Dalle *et al.*, 2003). The results obtained in the present study are very promising since we found two alleles and five genotypes exclusive from the blood cultures isolates. It would be of interest to determine why a large number of strains represented in this population have common alleles, and whether these have a selective advantage in the host. Specific studies focused on the minor genotypes should also be designed to assess whether some genotypes are more pathogenic than others.

Both microsatellites and minisatellites are found to have a role of biology function in the genomes with more and more evidences (Gao & Kong, 2005). These DNA sequence and structure elements provide interesting targets in the analysis of controlled gene expression in *C. albicans* in analogy with studies in other microrganisms. Several potentially regulatory VNTR domains have been recently described for *C. albicans*, and various modes of involvement in gene expression regulation were suggested. Field & Wills (1998) noted that many of the VNTRs encode amino acid homopolymers at the protein level, like CAIV, which could be involved in modulation of transcriptional activity of genes. Moreover, the same

group of researchers has detected elaborate sequence and repeat number variability at various SSR loci in the *C. albicans* genome. These observations emphasize the relevance of these elements in this particular yeast species. However, further studies on possible structure-function relationships are required. A putative link between VNTR polymorphism and virulence (or avirulence) of eukaryotic microbial pathogens was recently established for *Toxoplasma gondii*. Based on the nature of a repetitive DNA domain, parasites which show clear virulence in a mouse infection model could be separated from the avirulent strains. Several authors link entire genes containing VNTRs to pathogenicity related features of *C. albicans*. This finding raises the question of whether VNTRs controlling fungal invasiveness in a similar fashion exist in the genome of *C. albicans* (Lunel *et al*., 1998; Al-Aidan *et al*., 2007).

The microsatellite locus CAIV, described in the present work, is located inside *C. albicans IFF8* gene, which codes for a GPI anchored protein similar to *Hyr*1, a cell wall protein of *C. albicans*, exclusive of the hypha form. The allelic variability observed at this locus may be related with the virulence of the isolates in study and to their adaptability to the host. It was observed that the shorter alleles are more frequent than the heavier ones that have long stretches of repetitive units. Therefore, it is possible that the isolates that possess shorter alleles have an adaptive advantage over the others and that can explain their prevalence.

From the putative list of *C. albicans* 115 GPI proteins most of them (66%) are of completely unknown function, including *Iff*8 protein (Richard & Plaine 2006). These may be relevant to the future discovery of novel genes involved specifically in *C. albicans* pathogenicity. Firstly because of the putative localization of these proteins at the cell surface, thus in a very good position to be involved in any interaction with the surrounding environment: the host cells. Secondly because *C. albicans* is highly adapted to its environment compared to other opportunistic fungi which suggests that it developed during evolution several mechanisms to colonize his host: these original functions might be carried out by these surface proteins of unknown function.

Given the location of microsatellite CAIV, and since GPI proteins may be involved in cell wall biosynthesis and remodelling, adhesion and virulence, the

characterization of isolates bearing short and long alleles would be relevant to determine whether allele variability may be associated with differences in morphogenesis and strain response to stress agents.

References

AL-AIDAN, A. W.; GOESSENS, W.; LEMMENS-DEN TOOM, N.; AL-AHDAL, M.; VAN BELKUM, A. - Microevolution in genomic short sequence repeats of *Candida albicans* in non-neutropenic patients. Yeast. Vol.24 (2007): p.155-160.

ALBERTI-SEGUI, C.; MORALES, A.; XING, H.; KESSLER, M.; WILLINS, D.; WEINSTOCK, K.; COTTARE, G.; FECHTEL, K.; ROGERS, B. - Identification of potential cell-surface proteins in *Candida albicans* and investigation of the role of a putative cell-surface glycosidase in adhesion and virulence. Yeast. Vol.21 (2004): p.285-302.

ALBRECHT, A ; FELK, A.; PICHOVA, I.; NAGLIK, J. R.; SCHALLER, M.; DE GROOT, P.; MACCALLUM, D.; ODDS, F. C.; SCHÄFER, W.; KLIS, F.; MONOD, M.; HUBE, B. - Glycosylphosphatidylinositol-anchored proteases of *Candida albicans* target proteins necessary for both cellular processes and host-pathogen interactions. The Journal of Biological Chemistry. Vol.281 (2006): p.688–694.

ALMIRANTE, B.; RODRIGUEZ, D.; PARK, B.; CUENCA-ESTRELLA, M.; PLANES, A.; ALMELA, M.; MENSA, J.; SANCHEZ, F.; AYATS, J.; GIMENEZ, M.; SABALLS, P.; FRIDKIN, S.; MORGAN, J.; RODRIGUEZ-TUDELA, J.; WARNOCK, D.; PAHISSA, A.; BARCELONA CANDIDEMIA PROJECT STUDY GROUP - Epidemiology and Predictors of Mortality in Cases of *Candida* Bloodstream Infection: Results from Population-Based Surveillance. Journal of Clinical Microbiology. Vol.43 Nº4 (2005): p.1829-1835.

BAILEY, D. A.; FELDMANN, P. J. F.; BOVEY, M.; GOW, N.; BROWN, A. J. P. - The *Candida albicans* HYR1 Gene, Which Is Activated in Response to Hyphal Development, Belongs to a Gene Family Encoding Yeast Cell Wall Proteins. Journal of Bacteriology. Vol.178 (1996): p.5353–5360.

BASSETTI, M.; RIGHI, E.; COSTA, A.; FASCE, R.; MOLINARI, M.; ROSSO, R.; PALLAVICINI, F.; VISCOLI, C. - Epidemiological trends in nosocomial candidemia in intensive care. BMC Infectious Diseases. Vol.6 (2006): p.21-27.

BATES, S.; DE LA ROSA, J. M.; MACCALLUM, D. M.; BROWN, A. J.; GOW, N. A.; ODDS, F. C. - *Candida albicans* Iff11, a secreted protein required for cell wall structure and virulence. Infec Immun. Vol.75 (2007): p.2922-2928.

BATES, S.; HUGHES, H.; MUNRO, C.; THOMAS, W.; MACCALLUM, D.; BERTRAM, G.; ATRIH, A.; FERGUSON, M.; BROWN, A.; ODDS, F.; GOW, N. - Outer Chain N-Glycans Are Required for Cell Wall Integrity and Virulence of *Candida albicans*. The Journal of Biological Chemistry. Vol.281 Nº1 (2006): p.90-98.

BLOT, S.; VANDEWOUDE, K. - Management of Invasive Candidiasis in Critically Ill Patients. Drugs. Vol.64 Nº19 (2004): p.2159-2175.

BOCCIA, S.; POSTERARO, B.; LA SORDA, M.; VENTO, G.; MATASSA, P.; TEMPERA, A.; PETRUCCI, S.; FADDA, G. - Genotypic Analysis by 27A DNA Fingerprinting of *Candida albicans* Strains Isolated During an Outbreak in a Neonatal Intensive Care Unit. Concise Communications. Vol.23 Nº5 (2002): p.281-284.

BOTTEREL, F.; CESTERKE, C.; COSTA, C.; BRETAGNE, S. - Analysis of microsatellite markers of *Candida albicans*used for rapid typing. Journal Of Clinical Microbiology. Vol.39 (2001): p.4076-4081.

BOUGNOUX, M.; MORAND, S.; D'ENFERT, C. - Usefulness of Multilocus Sequence Typing for Characterization of Clinical Isolates of *Candida albicans*. Journal of Clinical Microbiology. Vol.40 Nº4 (2002): p.1290-1297.

BRETAGNE, S.; COSTA, J. M.; BESMOND, C.; CARSIQUE, R.; CALDERONE, R. - Microsatellite Polymorphism in the Promoter Sequence of the Elongation Factor 3 Gene of *Candida albicans* as the Basis for a Typing System .Journal Of Clinical Microbiology. Vol.35, Nº 7 (1997): p.1777–1780.

BUSCHIAZZO, E.; GEMMELL, N. J. - The rise, fall and renaissance of microsatellites in eukaryotic genomes. BioEssays. Vol.28 (2006): p.1040–1050.
CALDERONE, R.; FONZI, W. - Virulence factors of *Candida albicans*. Trends in Microbiology. Vol.9 Nº7 (2001): p.327-335.

CHEN, K.; LO, H.; LIN, Y.; LI, S. - Comparison of four molecular typing methods to assess genetic relatedness of *Candida albicans* clinical isolates in Taiwan. Journal of Medical Microbiology. Vol.54 (2005): p.249-258.

CHONG, P.; HADI, A.; LEE, Y.; PHAN, C.; TAN, B.; PENG, K.; SEOW, H. - Genotyping and drug resistance profile of *Candida spp.* In recurrent and one-off vaginitis, and high association of non-*albicans* species with non-pregnant status. Infection, Genetics and Evolution. Vol.7 Nº4 (2007): p.449-456.

CHOWDHARY, A.; LEE-YANG, W.; LASKER, B.; BRANDT, M.; WARNOCK, D.; ARTHINGTON-SKAGGS, B. - Comparison of multilocus sequence typing and Ca3 fingerprinting for molecular subtyping epidemiologically-related clinical isolates of Candida albicans. Medical Mycology. Vol. 44 (2006): p.405-417.

CHRISTENSEN, M.; SUNDE, L.; BOLUND, L.; ÉRNTOFT, T. F. - Comparison of three methods of microsatellite detection. J Clin Lab Invest. Vol.59 (1999): p.167-178.

COLOMBO, A.; NUCCI, M.; PARK, B.; NOUER, S.; ARTHINGTON-SKAGGS, B.; DA MATTA, D.; WARNOCK, D.; MORGAN, J. - Epidemiology of Candidemia in Brazil: a Nationwide Sentinel Surveillance of Candidemia in Eleven Medical Centers. Journal of Clinical Microbiology. Vol.44 Nº8 (2006): p.2816-2823.

CORREIA, A.; SAMPAIO, P.; ALMEIDA, J.; PAIS, C. - Study of Molecular Epidemiology of Candidiasis in Portugal by PCR Fingerprinting of *Candida* Clinical Isolates. Journal of Clinical Microbiology. Vol. 42 No. 12 (2004): p. 5899–5903.

DALLE, F.; DUMONT, L.; FRANCO, N.; MESMACQUE, D.; CAILLOT,D.; BONNIN, P.; MOIROUX, C.; VAGNER, O.; CUISENIER, B.; LIZARD, S.; BONNIN, A. - Genotyping of *Candida albicans* Oral Strains from Healthy Individuals by Polymorphic Microsatellite Locus Analysis. Journal of Clinical Microbiology Vol. 41, No. 5 (2003): p. 2203–2205.

DALLE, F.; FRANCO, N.; LOPEZ, J.; VAGNER, O.; CAILLOT, D.; CHAVANET, P.; CUISENIER, B.; AHO, S.; LIZARD, S.; BONNIN, A. - Comparative Genotyping of *Candida albicans* Bloodstream and Nonbloodstream Isolates at a Polymorphic Microsatellite Locus. Journal of Clinical Microbiology. Vol. 38 Nº12 (2000), p.4554–4559.

DE GROOT, P. W. J.; HELLINGWERF, K. J.; KLIS, F. M. - Genome-wide identification of fungal GPI proteins. Yeast. Vol.20 (2003): p.781–796.

DEFORCE, D.; MILLECAMPS, R.; VAN HOOFSTAT, D.; VAN DEN EECKHOUT, E. - Comparison of slab gel electrophoresis and capillary electrophoresis for the detection of the fluorescently labelled polymerase chain reaction products of short tandem repeat fragments. Journal of Chromatography A. Vol. 806 (1998): p. 149–155.

DIERINGER, D.; SCHLÖTTERER, C. - Two Distinct Modes of Microsatellite Mutation Processes: Evidence from the Complete Genomic Sequences of Nine Species. Genome Res. Vol.13 (2003): p.2242-2251.

ELOY, O.; MARQUE, S.; BOTTEREL, F.; STEPHAN, F.; COSTA, J.; LASSERRE, V.; BRETAGNE, S. - Uniform distribution of three Candida albicans microsatellite markers in two French ICU populations supports a lack of nosocomial cross-contamination. BMC Infectious Diseases. Vol. 6 (2006): p. 162-169.

ERNST, J. - Transcription factors in *Candida albicans* – environmental control of morphogenesis. Microbiology. Vol.146 (2000): p.1763–1774.

FIELD, D.; WILLS, C. - Long, polymorphic microsatellites in simple organisms. Proc Biol Sci. Vol. 263, Nº1367 (1996): p. 209-215.

FIELD, D.; WILLS, C. - Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. PNAS. Vol. 95 (1998): p.1647–1652.

FRIDKIN S. K., JARVIS R. - Epidemiology of Nosocomial Fungal Infections. CLINICAL MICROBIOLOGY REVIEWS. Vol. 9 Nº 4 (1996): p. 499–511.

FUNDYGA, R. E.; LOTT, T. J.; ARNOLD, J. - Population structure of *Candida albicans*, a member of the human flora, as determined by microsatellite loci. Infection, Genetics and Evolution. Vol.2 (2002): p.57–68.

GHANNOUM, A. - Potential role of phospholipases in virulence and fungal pathogenesis. Clin. Microbiol. Vol. 13 (2000): p. 122–143.

JERHME G., RAYMOND M., DE MEEIIS T., ROUSSETT F. - Testing Differentiation in Diploid Populations. Genetics. Vol. 144 (1998): p. 1933-1940.

GOZALBO, D.; ROIG, P.; VILLAMÓN, E.; GIL, M. L. - Candida and Candidiasis: The Cell Wall as a Potential Molecular Target for Antifungal Therapy. Current Drug Targets – Infectious Disorders. Vol.4 (2004): p.117-135.

GUO, B.; STYLES, C. A.; FENG, Q.; FINK, G. R. - A Saccharomyces gene family involved in invasive growth, cell – cell adhesion, and mating. Pnas. Vol.97 Nº22 (2000):p.12158–12163.

GUO, S.; THOMPSON, E. - A Monte Carlo Method for Combined Segregation and Linkage Analysis. Am. J. Hum. Genet. Vol.51 (1992): p. 1111-1126.

GUSMÃO, L.; PRATA, M.J.; AMORIM, A.; SILVA, F.; BESSA, I. - Characterization of four short tandem repeat loci (THO1, VWA31/A, CD4, and TP53) in northern Portugal. Hum Biol. Vol.69 Nº1 (1997): p.31-40.

HANCOCK, J. M.; SIMON, M. - Simple sequence repeats in proteins and their significance for network evolution. Gene. Vol.345 (2005): p.113– 118.

HOYER, L. - The ALS gene family of *Candida albicans*. Trends Microbiol. Vol.9 Nº4 (2001): p.176-80.

HUNTER, P. R.; GASTON, M.A. - Numerical Index of the Discriminatory Ability of Typing Systems: an Application of Simpson's Index of Diversity. Journal of Clinical Microbiology, Vol. 26, No. 11 (1988): p. 2465-2466.

IBRAHIM, A. S.; MIRBOD, F.; FILLER, S.G.; BANNO, Y.; COLE, G.T.; KITAJIMA, Y.; EDWARDS JR, J.E.; NOZAWA, Y.; GHANNOUM, M.A. - Evidence Implicating Phospholipase as a Virulence Factor of *Candida albicans*. Infection and immunity. Vol. 63, No. 5 (1995): p.1993–1998.

IVANOVSKA, N. - Phospholipases as a factor of pathogenicity in microorganisms. Journal of Molecular Catalysis. Vol.22 (2003): p. 357–361.

KAISER, C.; MICHAELIS, S.; MITCHEL, A. - Methods in yeast genetics. Cold Spring Harbour Laboratory, Cold Spring Harbour NY, 1994.

KASHI, Y.; King, D. G. - Simple sequence repeats as advantageous mutators in evolution. Trends in Genetics. Vol.22 Nº5 (2006): p.253-259.

KATTI, M. V.; RANJEKAR, P. K.; GUPTA, V. S. - Differential Distribution of Simple Sequence Repeats in Eukaryotic Genome Sequences. Mol. Biol. Evol. Vol.18 Nº7 (2001): p.1161–1167.

KOELSCH, G.; TANG, J.; LOY, J.; MONOD, M.; JACKSON, K.; FOUNDLING, S.; LIN, X. - Enzymic characteristics of secreted aspartic proteases of *Candida albicans*. Biochimica et Biophysica Acta. Vol. 1480 (2000): p. 117-131.

LAN, C.; NEWPORT, G.; MURILLO, L.A.; JONES, T.; SCHERER, S.; DAVIS, R.W.; AGABIAN, N. - Metabolic specialization associated with phenotypic switching in *Candida albicans.* PNAS. Vol.12 (2002): p.14907-14912.

LEVINSON, G.; GUTNAN, G. A. - High frequencies of short fameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escheychia coli K-12. Nucleic Acids Research. Vol.15 Nº13 (1987): p.5323-5338.

LI, Y.; KOROL, A. B.; FAHIMA, T.; NEVO, E. - Microsatellites Within Genes: Structure, Function, and Evolution. Molecular Biology and Evolution. Vol.21, Nº6 (2004): p.991-1007.

LOTT, T. J.; EFFAT, M. M. - Evidence for a more recently evolved clade within a *Candida albicans* North American population. Microbiology. Vol.147 (2001): p.1687–1692.

LOTT, T. J.; HOLLOWAY, B. P.; LOGAN, D. A.; FUNDYGA, R.; ARNOLD, J. - Towards understanding the evolution of the human commensal yeast *Candida albicans*. Microbiology. Vol.145 (1999): p.1137–1143.

LUNEL, F.V.; LICCIARDELLO, L.; STEFANI, S.; VERBRUGH, H.A.; MELCHERS, W.J.; MEIS, J.; SCHERER, S.; VAN BELKUM,,A. - Lack of Consistent Short Sequence Repeat Polymorphisms in Genetically Homologous Colonizing and Invasive *Candida albicans* Strains. Journal of bacteriology. Vol.180, Nº15 (1998): p. 3771–3778.

MAIDEN, M. C.; BYGRAVES, J. A.; FEIL, E.; MORELLI, G.; RUSSELL, J. E.; URWIN, R.; ZHANG, Q.; ZHOU, J.; ZURTH, K.; CAUGANT, D. A.; FEAVERS, I. M.; ACHTMAN, M.; SPRATT B. G. - Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. PNAS. Vol. 95 (1998): p.3140–3145.

MARCO, F.; LOCKHART, S.; PFALLER, M.; PUJOL, C.; RANGEL-FRAUSTO, M.; WIBLIN, T.; BLUMBERG, H.; EDWARDS, J.; JARVIS, W.; SAIMAN, L.; PATTERSON, J.; RINALDI, M.; WENZEL, R.; THE NEMIS STUDY GROUP; SOLL, D. - Elucidating the Origins of Nosocomial Infections with *Candida albicans* by DNA Fingerprinting with the Complex Probe Ca3. Journal of Clinical Microbiology. Vol. 37, No. 9 (1999): p. 2817–2828.

MAVOR, A.; THEWES, S.; HUBE, B. - Systemic Fungal Infections Caused by *Candida* Species: Epidemiology, Infection Process and Virulence Attributes. Current Drug Target. Vol. 6, Nº8 (2005): p. 1-12.

METZGAR, D.; BYTOF, J.; WILLS, C. - Expansion in Coding DNA Selection Against Frameshift Mutations Limits Microsatellite. Genome Res.Vol. 10. (2000): p. 72-80.

METZGAR, D.; VAN BELKUM, A.; FIELD, D.; HAUBRICH, R.; WILLS, C. - Random Amplification of Polymorphic DNA and Microsatellite Genotyping of Pre- and Posttreatment Isolates of Candida spp from Human Immunodeficiency Virus-Infected Patients on Different Fluconazole Regimens. Journal of Clinical Microbiology. Vol.36 Nº8 (1998): p.2308-2313.

NAGLIK, J.R.; CHALLACOMBE, S.J.; HUBE, B. - *Candida albicans* Secreted Aspartyl Proteinases in Virulence and Pathogenesis. Microbiology and molecular biology reviews. Vol. 67, No. 3 (2003): p. 400–428.

NAIDOO, R.; CHETTY, R. - The application of Microsatellites in molecular Pathology. Pathology Oncology Research. Vol.4 Nº4 (1998): p.310-315.

NG, K.; MADASAMY, M.; SAW, T.; BAKI, A.; HE, J.; SOO-HOO, T. - *Candida* biotypes isolated from clinical specimens in Malaysia. Mycopathologia. Vol. 144 (1999): p. 135–140.

ODDS, F.; DAVIDSON, A.; JACOBSEN, M.; TAVANTI, A.; WHYTE, J.; KIBBLER, C.; ELLIS, D.;MAIDEN, M.; SHAW, D.; GOW, N. - *Candida albicans* Strain Maintenance, Replacement, and Microvariation Demonstrated by Multilocus Sequence Typing. Journal of Clinical Microbiology. Vol. 44, No. 10 (2006): p. 3647–3658.

PERLROTH, J.; CHOI, B.; SPELLBERG, B. – Nosocomial fungal infections: epidemiology, diagnosis, and treatment. Medical Mycology. Vol.45 Nº4 (2007): p.321-346.

PUJOL, S.; JOLY,S.; LOCKHART, R.; NOEL, S.; TIBAYRENC, M.; SOLL; D. R.- Parity among the Randomly Amplified Polymorphic DNA Method, Multilocus Enzyme Electrophoresis, and Southern Blot Hybridization with the Moderately Repetitive DNA Probe Ca3 for Fingerprinting *Candida albicans.* Journal of Clinical Microbiology. Vol. 35, No. 9 (1997): p. 2348–2358.

RAYMOND, M.; ROUSSET, F. - GENEPOP Version 1.2: Population genetics software for exact tests and ecumenicist. J Heredity. (1995): p.248-249.

RICHARD, G.; HENNEQUIN, C.; THIERRY, A.; DUJON, B. - Trinucleotide repeats and other microsatellites in yeasts. Res. Microbiol. Vol.150 (1999): p.589–602.

RICHARD, M.; IBATA-OMBETTA, S.; DROMER, F.; BORDON-PALLIER, F.; JOUAULT, T.; GAILLARDIN, C. - Complete glycosylphosphatidylinositol anchors are required in *Candida albicans* for full morphogenesis, virulence and resistance to macrophages. Molecular Microbiology. Vol.44 Nº3 (2002): p.841–853.

RICHARD, M.L.; PLAINE, A. - A comprehensive analysis of GPI anchored proteins in *Candida albicans.* American Society for Microbiology. Vol.6, Nº2 *(*2006): p.119-133.

ROBLES, J.; LARRY KOREEN,1,2 STEVEN PARK,1 AND DAVID S. PERLIN - Multilocus Sequence Typing Is a Reliable Alternative Method to DNA Fingerprinting for Discriminating among Strains of *Candida albicans.* Journal of clinical microbiology. Vol. 42, No. 6 (2004): p. 2480–2488.

ROMANI, L.; FRANCESCO BISTONI AND PAOLO PUCCETTI - Adaptation of Candida albicans to the host environment: the role of morphogenesis in virulence and survival in mammalian hosts. Current Opinion in Microbiology. Vol.6 (2003): p.338–343.

RUIZ-DIEZ, B.; MARTINEZ, V.; ALVAREZ, M.; RODRIGUEZ-TUDELA, J.; MARTINEZ-SUAREZ, J. - Molecular Tracking of *Candida albicans* in a Neonatal Intensive Care Unit: Long-Term Colonizations versus Catheter-Related Infections. Journal of Clinical Microbiology. Vol.35 Nº12 (1997): p.3032-3036.

SAMARANAYAKE, Y. H.; SAMARANAYAKE, L. P.; DASSANAYAKE, R. S.; YAU, J. Y. Y.; TSANG, W. K.; CHEUNG, B. P. K.; YEUNG, K. W. S. - Genotypic shuffling of sequential clones of *Candida albicans* in HIV-infected individuals with and without symptomatic oral candidiasis. Journal of Medical Microbiology. Vol. 52 (2003): p.349–359.

SAMPAIO, P.; GUSMÃO, L.; ALVES, C.; PINA-VAZ, C.; AMORIM, A.; PAIS, C. - Highly Polymorphic Microsatellite for Identification of *Candida albicans* Strains. Journal of clinical microbiology. Vol.41, Nº2 (2003): p. 552–557.

SAMPAIO, P.; GUSMÃO, L.; CORREIA, A.; ALVES, C.; RODRIGUES, A.; PINA-VAZ, C.; AMORIM, A.; PAIS, C. - New Microsatellite Multiplex PCR for *Candida albicans* Strain Typing Reveals Microevolutionary Changes. Journal of clinical microbiology. Vol. 43, Nº8 (2005): p. 3869–3876.

SAVILLE, S.P. ; LAZZELL, A.L. ; BRYANT, A.P. ; FRETZEN, A. ; MONREAL, A. ; SOLBERG, E.O. ; MONTEAGUDO, C. ; LOPEZ-RIBOT, J.L. ; MILNE, G.T. - Inhibition of Filamentation Can Be Used To Treat Disseminated Candidiasis. Antimicrobial agents and chemotherapy. Vol. 50, Nº 10 (2006): p. 3312–3316.

SCHLÖTTERER, C. - Evolutionary dynamics of microsatellite DNA. Chromosoma. Vol.109 (2000): p.365–371.

SCHMID, J.; ROTMAN,'M.; REED, B.; PIERSON, C.L.; SOLL, D.R. – Genetic Similarity of *Candida albicans* Strains from Vaginitis Patients and Their Partners. Journal of clinical microbiology. Vol.31, Nº1 (1993): p. 39-46.

SCHROPPEL, K.; ROTMAN, M.; GALASK, R.; MAC, K.; SOLL, D.R. - Evolution and Replacement of Candida albicans Strains during Recurrent Vaginitis Demonstrated by DNA Fingerprinting. Journal of clinical microbiology. Vol.32, Nº11 (1994): p. 2646-2654.

SHIN, J.; OG, Y.; CHO, D.; KEE, S.; SHIN, M.; SUH, S.; RYANG, D. - Molecular Epidemiological Analysis of Bloodstream Isolates of Candida albicans from a University Hospital over a Five-Year Period. The Journal of Microbiology. Vol.43, Nº6 (2005): p.546-554.

SNYDMAN, D. - Shifting Patterns in the Epidemiology of Nosocomial Candida Infections. Chest. Vol.123 (2003): p.500-503.

SOLL, D.R - Gene regulation during high-frequency switching in *Candida albicans* Microbiology. Vol.143 (1997): p. 279–288.

SOLL, D.R. - The Ins and Outs of DNA Fingerprinting the Infectious Fungi. Clinical microbiology reviews. Vol. 13, Nº2 (2000): p. 332–370.

STEHR, F.; KRETSCHMAR, M.; KRÖGER, C.; HUBE, B.; SCHÄFER, W. – Microbial lipases as virulence factors. Journal of Molecular Catalysis B: Enzymatic. Vol.22 (2003): p. 347–355.

STÉPHAN, F.; BAH, M. S.; DESTERKE, C.; RÉZAIGUIA-DELCLAUX, S.; FOULET, F.; DUVALDESTIN, P.; BRETAGNE, S. - Molecular Diversity and Routes of Colonization of *Candida albicans* in a Surgical Intensive Care Unit, as Studied Using Microsatellite Markers. Clinical Infectious Diseases. Vol.35 (2002): p.1477–1483.

TAUTZ, D. - Hypervariability of simple sequences as a general source for polymorphic DNA markers.Nucleic. Acids Research. Vol.17 Nº16 (1989): p.6463-6471.

TAVANTI, A.; GOW, N.; SENESI, S.; MAIDEN, M.; ODDS, F.C. - Optimization and Validation of Multilocus Sequence Typing for *Candida albicans*. Journal of clinical microbiology. Vol. 41, Nº8 (2003): p. 3765–3776.

TAY,S.T.; CHAI, H.C ; NA, S.L. ; NG, K.P. ; SOO-HOO, T.S. - Molecular subtyping of clinical isolates of *Candida albicans* and identification of Candida dubliniensis in Malaysia. Mycopathologia. Vol. 159 (2005): p.325–329.

Taylor, B.N. ; Harrer, T.; Pscheidl, E. ; Schweizer, A.; Rollinghoff, M.; Schroppel, K. -Surveillance of nosocomial transmission of Candida albicans in an intensive care unit by DNA fingerprinting. Journal of Hospital Infection. Vol. 55 (2003): p. 283–289.

TORTORANO, A.; KIBBLER, C.; PEMANC, J.; BERNHARDT, H.; KLINGSPOR, L.; GRILLOT , R. - Candidaemia in Europe: epidemiology and resistance. International Journal of Antimicrobial Agents. Vol.27 (2006): p.359-366.

TÓTH, G.; GÁSPÁRI, Z.; JURKA, J. - Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. Genome Res. Vol.10 (2000): p.967-981.

VARGAS, K.; JOLY, S. - Carriage Frequency, Intensity of Carriage, and Strains of Oral Yeast Species Vary in the Progression to Oral Candidiasis in Human Immunodeficiency Virus-Positive Individuals. J. clin. microbiology. Vol.40 Nº2 (2002): p.341-350.

VAZQUEZ, J.A.; BECKLEY, A.; SOBEL, J.D.; ZERVOS,M.J. - Comparison of Restriction Enzyme Analysis and Pulsed-Field Gradient Gel Electrophoresis as Typing Systems for Candida albicans. Journal of clinical microbiology. Vol. 29, Nº5 (1991): p. 962-967.

VERSTREPEN, K. J.; JANSEN, A.; LEWITTER, F.; FINK, G. R. - Intragenic tandem repeats generate functional variability. Nat Genet. Vol. 37, Nº9 (2005): p. 986–990.

VOSS, A.; PFALLER, M.A.; HOLLIS, R.J.; RHINE-CHALBERG, J.; DOEBBELING, N. - Investigation of *Candida albicans* Transmission in a Surgical Intensive Care Unit Cluster by Using Genomic DNA Typing Methods. Journal of clinical microbiology. Vol. 33, No. 3 (1995): p. 576–580.

WHITEWAY, M.; OBERHOLZER, U. - *Candida* morphogenesis and host–pathogen interactions. Current Opinion in Microbiology. Vol. (2004): p. 350–357.

YANG, Y. – Virulence factors of *Candida* species. J Microbiol Immunol Infect. Vol.36 (2003): p. 223-228.

Annexe

CAI and CAIV genotypes of all isolates analysed and respective origin

Strain	Provenance	Collection Site	CAI genotype	CAIV genotype
63	HSJ	Blood	16-25	26-26
69	HSJ	Blood	16-25	26-26
90	HSJ	Blood	21-25	26-26
93	HSJ	Blood	21-25	26-26
114	HSJ	Blood	12-12	27-27
124a	HSJ	Blood	18-34	23-39
140	HSJ	Blood	18-34	23-39
140a	HSJ	Blood	34-34	23-39
144	HSJ	Blood	18-34	23-39
130	HSJ	Blood	12-12	16-23
141	HSJ	Blood	12-17	23-23
143	HSJ	Blood	26-26	23-23
150	HSJ	Blood	26-26	23-23
154	HSJ	Blood	26-26	23-23
154b	HSJ	Blood	26-26	23-23
155	HSJ	Blood	29-29	16-16
164	HSJ	Blood	26-28	16-23
164b	HSJ	Blood	26-28	16-23
165	HSJ	Blood	21-25	26-26
168	HSJ	Blood	21-26	26-26
168a	HSJ	Blood	21-26	26-26
168b	HSJ	Blood	21-25	26-26
7	HSJ	Vagina	30-30	35-35
8	HSJ	Vagina	30-32	35-35
9	HSJ	Vagina	30-32	29-29
22	HSJ	Vagina	23-27	29-29
23	HSJ	Vagina	23-27	32-32
22	IPO Porto	Blood		26-26
24	IPO Porto	Blood		16-26
1	IPO Lisboa	Blood	25-25	23-23
2	IPO Lisboa	Blood	25-25	23-23

3	IPO Lisboa	Blood		23-32
4	IPO Lisboa	Blood		32-32
5	IPO Lisboa	Blood		23-23
6	IPO Lisboa	Blood		32-32
7	IPO Lisboa	Blood	26-26	23-23
8	IPO Lisboa	Blood		26-29
10	IPO Lisboa	Blood	12-12	26-48
11	IPO Lisboa	Blood	23-27	29-29
13	IPO Lisboa	Blood	40-40	26-26
14	IPO Lisboa	Blood	20-28	26-26
15	IPO Lisboa	Blood	35-44	26-26
16	IPO Lisboa	Blood		26-26
17	IPO Lisboa	Blood	16-38	26-29
18	IPO Lisboa	Blood		26-29
19	IPO Lisboa	Blood	18-28	26-29
20	IPO Lisboa	Blood	21-25	26-26
21	IPO Lisboa	Blood	19-34	32-32
22	IPO Lisboa	Blood	21-26	26-28
23	INSA	Vagina	27-30	27-30
24	INSA	Vagina		24-31
26	INSA	Vagina		26-26
27	INSA	Vagina	12-21	26-26
28	INSA	Vagina		31-35
29	INSA	Vagina	27-50	26-29
30	INSA	Vagina		31-35
31	INSA	Vagina		27-27
32	INSA	Vagina		26-29
33	INSA	Vagina	32-46	29-32
34	INSA	Vagina		26-26
35	INSA	Vagina	30-48	
36	INSA	Vagina	32-45	
37	INSA	Vagina	27-34	26-29
38	INSA	Vagina	26-26	16-32
39	INSA	Vagina	18-24	16-23
40	INSA	Vagina	18-18	16-23

41	INSA	Vagina	20-33	29-55
42	INSA	Vagina	26-35	16-32
43	INSA	Vagina		26-26
44	INSA	Vagina	20-28	16-26
1	CSC	Vagina	17-21	25-31
2	CSC	Vagina	26-33	32-32
3	CSC	Vagina	20-37	16-32
4	CSC	Vagina	18-18	23-23
10	CSC	Vagina	21-26	26-31
11	CSC	Vagina	25-26	16-23
12	CSC	Vagina	24-26	32-32
13	CSC	Vagina	27-49	26-26
16	CSC	Vagina	21-25	25-25
19	CSC	Vagina	20-27	26-26
24	CSC	Vagina	39-46	26-26
25	CSC	Vagina	21-26	26-26
26	CSC	Vagina	27-27	26-29
27	CSC	Vagina	17-21	26-29
29	CSC	Vagina	16-27	26-29
36	CSC	Vagina	21-25	26-26
36E	CSC	Vagina	21-25	26-26
37	CSC	Vagina	11-18	20-20
38	CSC	Vagina	20-28	29-48
44	CSC	Vagina	11-18	16-20
46	CSC	Vagina	27-27	29-48
50	CSC	Vagina	21-26	26-26
52	CSC	Vagina	22-34	26-26
53	CSC	Vagina	27-47	26-29
54	CSC	Vagina	27-42	26-29
55	CSC	Vagina	11-28	16-27
61	CSC	Vagina	28-47	26-29