COMPARISON OF GENOTYPES AND VIRULENCE FACTORS AMONG CARRIER AND PATHOGENIC STRAINS OF *CANDIDA ALBICANS*

By

ZAIN ILLYAASEEN BIN KHAIRUDIN

A dissertation submitted to the Department of Pre-clinical Sciences,

Faculty of Medicine and Health Sciences,

Universiti Tunku Abdul Rahman,

in partial fulfillment of the requirements for the degree of

Master of Medical Science

January 2021

ABSTRACT

COMPARISON OF GENOTYPES AND VIRULENCE FACTORS AMONG CARRIER AND PATHOGENIC STRAINS OF *CANDIDA ALBICANS*

ZAIN ILLYAASEEN BIN KHAIRUDIN

Candida albicans is a prominent global fungal pathogen capable of causing fatal systemic infections in humans. Presently in Malaysia, there is little data on genetic diversity of this organism and behavioural characteristics of local organisms. In this project, three genotyping methods; 25S rDNA genotyping, ALT sequence typing and MLST were applied to study genetic diversity whereas antifungal susceptibility test, gene expression assay and biofilm formation assay were used to assess some behavioural criteria. This study found that the most common genotype was A, accounting for approximately 70% of the 111 isolates tested. MLST analysis showed a statistically significant association between pathogenicity and a group of closely related isolates. All isolates tested were shown to be susceptible to both nystatin and fluconazole. No statistically significant difference was found between the biofilm forming ability of isolates of pathogenic or commensal origins. qPCR assays revealed the upregulation of HWP1 as a virulence gene in the representative isolates from pathogenic origin. While the distribution of 25S rDNA genotypes mostly matched previous studies done elsewhere, MLST data showed that there may yet be

many novel sequence types in Malaysia which have yet to be catalogued. In the future, greater efforts in genotyping for proper documentation of genetic diversity would be beneficial with focus on greater research into the genomes of isolates exhibiting enhanced virulence.

ACKNOWLEDGEMENT

I would like to express my thanks and gratitude to:

(a) my supervisor, Prof. Ngeow Yun Fong, for her support, guidance and advice throughout the entire duration of my study.

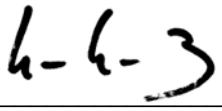
(b) my co-supervisor, Prof. Yap Sook Fan, for her support, advice and guidance.

(c) UTAR (specifically UTAR Strategic Research Fund (UTARSRF): IPSR/RMC/UTARSRF/PROJECT 2014-C1/002 (6274/002).

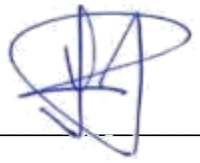
APPROVAL SHEET

This dissertation entitled **"COMPARISON OF GENOTYPES AND VIRULENCE FACTORS AMONG CARRIER AND PATHOGENIC STRAINS OF *CANDIDA ALBICANS*"** was prepared by ZAIN ILLYAASEEN BIN KHAIRUDIN and submitted as partial fulfillment of the requirements for the degree of Master of Medical Science at Universiti Tunku Abdul Rahman.

Approved by:



(Prof. Dr. Ngeow Yun Fong) Date:…1/12/2020……..
Professor/Supervisor
Department of Pre-clinical Sciences
Faculty of Medicine and Health Sciences
Universiti Tunku Abdul Rahman



(Prof. Dr. Yap Sook Fan) Date:…5/12/2020…..
Professor/Co-supervisor
Department of Pre-clinical Sciences
Faculty of Medicine and Health Sciences
Universiti Tunku Abdul Rahman

SUBMISSION SHEET

FACULTY OF MEDICINE AND HEALTH SCIENCES

UNIVERSITI TUNKU ABDUL RAHMAN

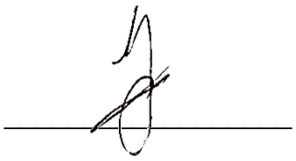
Date: _29/1/2021____

SUBMISSION OF DISSERTATION

It is hereby certified that Zain Illyaaseen Bin Khairudin (ID No: 1805954) has completed this dissertation entitled "COMPARISON OF GENOTYPES AND VIRULENCE FACTORS AMONG CARRIER AND PATHOGENIC STRAINS OF *CANDIDA ALBICANS*" under the supervision of Prof. Dr. Ngeow Yun Fong from the Department of Pre-Clinical Sciences, Faculty of Medicine and Health Sciences, and Prof. Dr. Yap Sook Fan from the Department of Pre-Clinical Sciences, Faculty of Medicine and Health Sciences.

I understand that the University will upload softcopy of my dissertation in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

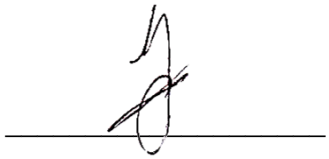
Yours truly,



(Zain Illyaaseen Bin Khairudin)

DECLARATION

I, Zain Illyaaseen Bin Khairudin hereby declare that the dissertation is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously of concurrently submitted for any other degree at UTAR or other institutions.



(Zain Illyaaseen Bin Khairudin)

Date: 5/12/2020

TABLE OF CONTENTs

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CLSI: Clinical and Laboratory Standards Institute

DST: Diploid sequence type

HDU: High dependency unit

HGDI: Hunter Gaston Discriminatory Index

HVS: High vaginal swab

IBM SPSS: International Business Machines Statistical Product and Service Solutions

ICU: Intensive care unit

ISHAM: International Society for Human & Animal Mycology

ITS: Internal transcribed spacer

IUPAC: International Union of Pure and Applied Chemistry

NCBI: National Center for Biotechnology Information

OD_{595}: Optical density reading at 595 nm wavelength

SNPs: Single nucleotide polymorphisms

ST: Sequence type

UPGMA: Unweighted pair group method with arithmetic mean

UV: Ultraviolet

CHAPTER 1

INTRODUCTION

Candida albicans is one of hundreds of infectious fungi which are known to be pathogenic to humans (Taylor et al., 2001). It is considerably prominent, given that 50% to 75% of women will experience at least one episode of vulvovaginal candidiasis and 5% to 8% will have four occurrences annually (Brown et al., 2012). Besides vulvovaginal infections, *C. albicans* can also cause superficial infections of the skin, mucosa of the oral cavity and other sites of the gastrointestinal tract and in some cases potentially fatal systemic infections which include visceral infections and candidemia. The more serious life threatening cases of disseminated candidiasis are more likely to occur in immunocompromised patients, especially those suffering from malignant diseases undergoing cytotoxic therapy and radiological treatment, HIV positive individuals, and those with chronic illnesses who are undergoing immunosuppression therapy and receiving steroids. Disseminated infections originating from biofilms of *C. albicans* can also occur in non-immunocompromised individuals who have long-standing in-dwelling intravascular catheters such as patients requiring long-term parenteral nutrition.

The significance of this pathogen is further compounded by data showing that more than 90% of all reported fungal related deaths are caused by species belonging to one of four genera; *Crytococcus*, *Candida*, *Aspergillus*, and

Pneumocystis. Polymorphic in nature, *C. albicans* is able to switch between different forms allowing quick adaptation to current environmental changes and stresses induced, such as a human immune response. The yeast can survive in a variety of different conditions, for example those involving limited nutrient availability, increased temperatures, altered pH, high CO_2 concentrations and the presence of serum (Gow and Yadav, 2017). A particularly important characteristic is the formation of germ tubes, structures protruding from the ovoid yeast cell in high protein environments such as serum. Germ tube formation and complex hyphal structures aid the organism in expression of virulence by penetrating and damaging host cells. A local study identified renal insufficiency, prior antibacterial and antifungal therapy and urinary catheterization as independent risk factors for candidiasis. The use of catheters in particular can be of concern in long term care facilities as they provide a surface for biofilm formation (Tzar et al., 2015).

Various genotyping methods across the globe have been employed to identify strains of organisms with specific characteristics such as enhanced virulence or antimicrobial resistance and to study outbreaks by identifying genetic similarities between organisms. In this study, 25S rDNA genotyping, ALT repeat sequence typing and multi-locus sequence typing were used. 25S rDNA genotyping is also known as ABC genotyping, a name derived from the three most common genotypes of *C. albicans* found using this method. There were formerly five genotypes identifiable by this method; genotypes A, B, C, D and E. Genotype D has been reclassified into a distinct species, *Candida dubliniensis* while genotype E appears exceptionally rare (McCullough, K. V. Clemons, et al., 1999; Tamura et

al., 2001). The differentiating factor in this method is the size of the PCR amplicon obtained using primers targeting the 25S rDNA region, where genotype A will yield an amplicon size of 450 bp, genotype B an 840 bp amplicon, genotype C 450 and 840 bp amplicons and genotype E a 962 bp amplicon (Tamura et al., 2001). This method is resource efficient compared to some other methods as it only requires PCR and subsequent agarose gel electrophoresis to determine the genotype. The next method is ALT sequence genotyping which is based on the repetition of ALT sequences of 172 bp in length in the repetitive sequences of the *C. albicans* chromosome (Iwata et al., 2006). This method is especially useful for supplementing the discriminatory power of the aforementioned 25S rDNA genotyping since it is able to differentiate substantially more than the initial four true *C. albicans* genotypes. ALT sequences are determined not only by the presence of a band in agarose gel electrophoresis but also by the intensity of the respective bands. From the results of electrophoresis, ALT repeat numbers of 1, 2, 3, 4, 5 or 6, or any combination of the numbers in ascending order separated by '/', such as 2/3 or 3/4/5 in cases where multiple bands share the same intensity which is greater than others. The drawback of this method is that determination of band intensity remains a subjective matter and an assessment which should be made in reasonable time and with respect to reagent integrity to avoid degradation of either the sample or light sensitive reagents. Despite this, as mentioned, is has been combined with ABC genotyping to provide more specific data sets for analysis (Sawadogo et al., 2019). The third and final method of genotyping used was MLST. This technique was initially developed as an unambiguous and portable method for genotyping of

Neisseria meningitidis (Maiden et al., 1998). It is unambiguous in the sense that the results are sequence based as opposed to electrophoresis based, the interpretation of which could be subjective in nature, and portable in the sense that sequencing data can be stored and transported electronically which guarantees preservation of data integrity. Since its inception, MLST schemes have been developed for many pathogenic organisms, both prokaryotes and eukaryotes, including *C. albicans*. Different schemes will make use of different housekeeping genes for sequence type determination. In the case of *C. albicans*, seven housekeeping genes are used and the data is stored in a database which is accessible via the internet.

As a defence mechanism, organisms have been known to develop resistance to antimicrobial agents. Antifungals are a common treatment method for various kinds of fungal infections, both superficial and systemic. A study which showed *C. albicans* as the most common causative agent of fungaemia also found a positive correlation between the use of antifungal medications such as itraconazole, fluconazole and echinocandins with an increase in healthcare-associated candidaemia (Lai et al., 2012). Three classes of antifungal medications have been developed for life-saving treatment in cases of systemic fungal infections; polyenes, azoles and echinocandins (Roemer and Krysan, 2014). Given the absence of novel antifungal drugs, the treatment options for exceptionally severe fungal infections have remained the same for over five decades (Krysan, 2017). As a consequence, doctors have very limited options when planning treatment for particularly virulent and resistant fungal infections. These challenges highlight the need for broad monitoring of incidences of antifungal resistance globally and in-

depth investigation of the mechanisms supporting said resistance to aid the development of novel antifungal agents.

This study encompassed the concepts mentioned above, studying *C. albicans* genotypes using three methods, determination of antifungal susceptibility, virulence gene expression assays and biofilm formation assays. The genetic diversity of 111 *C. albicans* Malaysian isolates, comprising 50 isolates of commensal origin and 61 isolates of pathogenic origin, were analysed using 25S rDNA genotyping supplemented with ALT sequence typing and MLST. Results were analysed to determine if any trends were present in terms of genetic relatedness within pathogenic and commensal isolates of *C. albicans*. The outcomes of other studies conducted across the globe were compared with this study, to determine if there is any uniqueness in genotype distribution in these *C. albicans* Malaysian isolates. In addition, statistical analysis is used to determine concordance between the different genotyping methods as well as their respective discriminatory powers. Finally, an association between expression of virulence genes and biofilm formation with specific isolate groups was sought. This work will contribute to the existing pool of knowledge relating to genotypes of *C. albicans* as determined using the aforementioned methods, as well as any association between these genotypes and pathogenicity and antifungal resistance.

CHAPTER 2

LITERATURE REVIEW

C. albicans is an opportunistic pathogenic yeast capable of causing superficial or systemic infections in humans. This formidable pathogen has been studied the world over but remains a significant threat with mortality rates from systemic infections reaching 50% (Mayer et al., 2013). To date, there have been few research papers on genotype distribution of *C. albicans* from Malaysia. Local researches include an analysis of genetic relatedness of *C. albicans* obtained from blood infections (Saranpal et al., 2015), molecular subtyping of clinical isolates (Tay et al., 2005) and genotyping of *C. albicans* isolates obtained from patients with vaginal candidiasis (Chong et al., 2003). In other regions, genotyping efforts have contributed to global knowledge of genotype distribution with respect to geographic location. A particularly interesting finding was a greater abundance of genotype C isolates in Israel (McCullough, K. Clemons, et al., 1999) since genotype A was found to be the most common in various countries including the United States and United Kingdom (McCullough, K. Clemons, et al., 1999), Burkina Faso (Zida et al., 2018) and Iran (Dalvand et al., 2018). Moreover, it has been shown that ABC genotype and geographic location were statistically significant factors in clade distribution in multilocus sequence typing (Odds et al., 2007). Other genotyping methods have also been used to show that there exists a genetic diversity with respect to geographical location among *C. albicans* isolates (Clemons et al., 1997). ABC genotyping which targets the 25S rDNA is highly reproducible and requires minimal labour. Despite this, the results are somewhat

ambiguous as the genotype determination is done by electrophoresing PCR products. The wide variations in materials used for electrophoresis and different conditions for visualizing and interpreting the results could be prone to error.

ALT subtyping is commonly used as a supplement to 25S rDNA genotyping to increase discriminatory power of the methods. In Ouagadougou researchers used ALT subtyping combined with 25S rDNA genotyping and found genotype A3 to be the most common genotype, also noting a high incidence of fluconazole resistance in this group though it did not carry statistical significance (Sawadogo et al., 2019). A similar study in Japan found genotype A3 to be the most common (Iwata et al., 2006)

Multi locus sequence typing (MLST) has been widely used for epidemiological studies of *C. albicans*. It was first proposed as a tool for better understanding nosocomial transmissions of yeast infections (Bougnoux et al., 2002). As the PCR products are sequenced, the data produced is unambiguous and highly transportable. The prevailing benefit is that data can be shared globally in a study of genetic diversity of *C. albicans* strains. A general consensus has been adopted for the standardization of genes used for MLST of *C. albicans*. These seven genes are *AAT1a, ACC1, ADP1, MPIb, SYA1, VPS13* and *ZWF1b* (Bougnoux et al., 2003). All of these genes were chosen because they were the smallest size set able to provide high discriminatory power in MLST. Properties of the genes are listed in table 2.1. Principally, the selected genes should also be stable under selective pressure to maintain their original function (Bougnoux et al., 2003). MLST works on finding single nucleotide polymorphisms (SNPs) in selected genes. Sequencing

chromatograms are analysed visually and compared with existing data in the online database (Odds and Jacobsen, 2008). The end product is the identification of the sequence type (ST) for haploid organisms or diploid sequence type (DST) for diploid organisms (Odds and Jacobsen, 2008). It is important to note that just because two strains have an identical ST or DST, this does not mean that they are identical strains, merely that they are indistinguishable by MLST.

Table 2.1 Gene products and locus positions of MLST genes.

Gene name	**Gene product**	**Locus position**
AAT1a	Aspartate aminotransferase	Chromosome:2; NC_032090.1 (1077700..1079013)
ACC1	Acetyl-CoA carboxylase	Chromosome:R; NC_032096.1 (155196..162011)
ADP1	Putative ATP-dependent permease	Chromosome:R; NC_032096.1 (1245828..1248944)
MPIb	Mannose 6-phosphate isomerase	Chromosome:2; NC_032090.1 (1972394..1973719)
SYA1	Alanine t-RNA ligase	Chromosome:6; NC_032094.1 (804198..807107)
VPS13	Vascular protein sorting-associated protein	Chromosome:4; NC_032092.1 (1341803..1351054)
ZWF1b	Glucose-6-phosphate 1-dehydrogenase	Chromosome:1; NC_032089.1 (1953260..1954783)

One method to analyse MLST data is the construction of a phylogenetic tree. Given that MEGA software which is utilized for phylogenetic analysis cannot process heterozygous code data, the sequences must be edited to allow only A, C, G and T bases. For example, if a heterozygous locus was denoted by W, this letter would be replaced with A and T. To ensure sequences remained the same length, any locus which is homozygous in nature would be rewritten twice (Tavanti, Davidson, Johnson, et al., 2005). Also common is the use of allelic profiles for assigning clonal clusters to respective strains (Jolley et al., 2001). This method applies an algorithm called BURST to examine internal relationships of clonal complexes in which isolates are grouped based on the number of locus differences in their profiles.

The literature cited supports that the methods chosen for this comparative study of genotyping are indeed well established and recognized. It is expected that a genotype trend similar to what has been shown in other research can be observed in Malaysia. A local study has shown that genotype A is the most predominant genotype in clinical isolates (Tay et al., 2005). What will set this study apart from other local studies of genotyping of *C. albicans* is the use of MLST with its high discriminatory power and the inclusion of healthy carrier isolates for comparison and more definitive genotype trend identification.

Antifungal resistance of *C. albicans* is of major concern given its high prominence in infections presenting in a clinical setting. As with genotyping data, information on trends of antifungal resistance of *C. albicans* in Malaysia is scarce. This work is conducted based on internationally recognized guidelines issued by

CLSI in document M44 which details the method for antifungal disk diffusion susceptibility testing of yeasts (NCCLS, 2004). Of significant importance is a recent study which reported that antifungal resistant candidaemia is significantly more likely to develop in patients who have previously been exposed to echinocandins and those who have received invasive ventilation (Aldardeer et al., 2020). Recurring infections and long term admissions in specialized care units, given the aforementioned study, are likely to further encourage the development of antifungal resistant strains of *Candida spp*., an occurrence particularly critical given the prominence of the organism in nosocomial outbreaks (Behzadi et al., 2015).

A biofilm is a multicellular cluster of microorganisms with a self-produced matrix (Gebreyohannes et al., 2019) and can consist of a single species or a combination of different species. Many bacteria and yeast species are able to form biofilms as a means of survival and adaptation to a new environment and that often helps in the exhibition of pathogenicity within an infected host. *C. albicans* biofilms are capable of forming on abiotic surfaces such as dentures and catheters as well as biotic surfaces such as mucosal c ell surfaces (Mayer et al., 2013). While the biofilm itself serves as a favourable initiator for an infection, it has been shown that *C. albicans* cells in yeast form which are disseminated from the biofilm are crucial to enhancement of the infection and spreading to other body sites, resulting in candidemia or disseminated invasive candidiasis (Uppuluri et al., 2010). These disseminated cells display greater pathogenic properties such as better biofilm forming capabilities and enhanced surface adhesion. Given the importance of Candida biofilms and disseminated cells originating from them in pathogenicity

and cell survival, the biofilm is a prime target for antifungal agents. Differentiation of biofilm forming ability between strains has been demonstrated using MLST, where the DST659 genotype of *C. albicans* was shown to have a greater ability compared to DST693 (Lu et al., 2018).

At the base of cellular pathways and functions are genes, expressed to encode various proteins with many functions, such as defense mechanisms and inflicting damage to host tissues. In this study, five genes; *CAT1*, *EFG1*, *SAP1*, *ALS3* and *HWP1*, have been selected for a gene expression assay. These genes were chosen for their significance to *C. albicans* survival and pathogenicity. Catalase, an antioxidant which in *C. albicans* is derived from the *CAT1* gene, plays a key role in defending organisms against the effects of hydrogen peroxide and can be considered a virulence factor (Wysong et al., 1998). Some antifungals are known to cause oxidative stress in pathogenic fungi, wherein catalase would offer protection against oxidants and by extension, some antifungals (Román et al., 2016). The *EFG1* gene plays a role in filamentous growth (Braun and Johnson, 2000) which is essential for colonization of host tissues and biofilm formation. EFG1 is so crucial that a deletion of solely this gene resulted in a *C. albicans* strain exhibiting reduced adhesion capabilities and inability to penetrate the stratum corneum (Dieterich et al., 2002). It has also been shown that expression levels of EFG1 will differ depending on the immune status of the host, since selective pressures from the respective environments will determine population composition of the pathogen (Pierce and Kumamoto, 2012).

In addition to *EFG1*, the *HWP1* gene also has a role in filamentous growth, where its expression is exclusive to hyphae and the resultant protein acts as an adhesin (Naglik et al., 2006). Given these characteristics, researchers hypothesized and proved that the *HWP1* gene is crucial for biofilm formation and adhesion, where a mutant with a defective *HWP1* gene generated a thin biofilm and overexpression of the gene resulted in enhanced adherence which is critical for biofilm formation (Nobile et al., 2006). Further evidence for the importance of this gene was provided in a study demonstrating in an *in vivo* experiment that a knockout of *HWP1* renders *C. albicans* cells unable to maintain an infection despite normal behavior in the initiation phase (Tsuchimori et al., 2000).

The *ALS3* gene, which belongs to the agglutinin-like sequence gene family, also plays roles similar to *EFG1* and *HWP1*. The encoded protein acts as an adhesin, mediating binding of *C. albicans* to various surfaces within a host and also to extracellular matrix proteins present in biofilms (Liu and Filler, 2011). Besides this, the protein also serves as an invasin which binds to cadherins of human epithelial cells and induces endocytosis (Phan et al., 2007). The Als3 protein has also been shown able to bind ferritin and utilize it as an iron source for *C. albicans* cellular activity and that the susceptibility of human oral epithelial cells to damage from *C. albicans* was directly related to the ferritin concentration within the oral epithelial cells (Almeida et al., 2008).

SAP1 is a member of the secreted aspartyl proteinases family and is a recognized virulence factor of *C. albicans* which contributes to tissue damage as it degrades proteins (Hube et al., 1997; Alonso et al., 2018). There exist some

conflicting ideas on the role of *SAP1* in virulence. For instance, a study noted that increased expression of the *SAP1* gene, among others of the same family, was associated with significant epithelial damage in reconstituted human vaginal epithelium (Schaller et al., 2003) while a contradictory standpoint claims the role is not as significant as thought given that complications of the gene-knockout methods used by Hube et al. (1997) resulted in misinterpretation of results (Correia et al., 2010), an outcome which has been characterized in other works (Bain et al., 2001; Brand et al., 2004).

CHAPTER 3

MATERIALS AND METHODS

3.1 Identification of *C. albicans*

A total of 61 isolates obtained from inpatients of a local hospital constituted the pathogenic group, while 50 isolates obtained from buccal swabs of residents, who showed no symptoms of oral candidiasis, residing in old folk's homes made up the commensal group. All isolates were retrieved from archived samples obtained in previous studies, stored in our biobank. Retrieved isolates were streaked on CHROMagar Candida (Becton Dickinson) and incubated at 37°C for 48 hours. Single green colonies were selected for a germ tube test, performed by culturing the respective isolates in 1.0 ml of human serum for 3 hours followed by a gram stain and microscopic examination. Germ tube positive isolates were subject to DNA extraction using the bacterial/fungal DNA extraction kit (Zymo Research). ITS PCR was done using GoTaq® Green Master Mix (Promega), with components prepared as per Table 3.1 and according to protocols previously described (Saranpal et al., 2015). The PCR products were sequenced and the results checked in the ISHAM Barcoding Database (https://its.mycologylab.org/) to confirm species identity.

Table 3.1 PCR master mix components for PCR.

Component	**Volume (µl)**
Master Mix	12.5
Forward primer	0.2
Reverse primer	0.2
DNA	1.0
H_2O	11.1

3.2 ABC genotyping and subtyping of *C. albicans*

DNA of isolates confirmed to be *C. albicans* by molecular identification was used in the following procedures. PCR for 25S rDNA genotyping was performed using the primers CA-INT-L: ATA AGG GAA GTC GGC AAA ATA GAT CCG TAA and CA-INT-R: CCT TGG CTG TGG TTT CGC (Sawadogo et al., 2019), with other reaction components as listed in Table 3.1. PCR products were electrophoresed in 2% agarose gel (80V, 30 minutes) and the results visualised with a UV transilluminator. Genotype was determined based on the size of the DNA band, or bands, depending on the genotype. The PCR procedure for ALT sequence typing was performed in a similar fashion but instead using the primers ASDcF: TGA ACC TGA ACT TGT GCT ACA AAG and pCSCR: CGC CTC TAT TGG AGC TCG AGT AGT C (Sawadogo et al., 2019), with other reaction components as listed in Table 3.1. PCR products were electrophoresed in 2% agarose gel (80V, 30 minutes) for determination of the subtype. In this instance, typing is determined

by the presence of a band or bands and their respective intensities as they appear under UV transillumination.

3.3 MLST of *C. albicans*

MLST was performed using primers described in other works (Bougnoux et al., 2002; Tavanti et al., 2003). Briefly, PCR was performed using seven sets of primers targeting seven housekeeping genes of *C. albicans*, as listed in Table 3.2. Primer design took into account the necessary region of the gene which would need to be sequenced and provided sufficient length in excess to do so. Components of the PCR master mix were as listed in Table 3.1. PCR products were sequenced by Sanger sequencing outsourced to a private company and the results analysed for SNPs. Where applicable, heterogeneous SNPs determined from the chromatograph were marked appropriately according to the respective IUPAC nucleotide codes. Edited sequences were copied from the chromatograph and pasted into the PubMLST website (https://pubmlst.org/organisms/candida-albicans/) for a database query. Any matches to existing entries for any of the seven genes used were noted. If all seven genes had recorded allele numbers, the allelic profile was formed and cross-referenced in the same database to obtain the sequence type. Any matches were recorded. To make analysis using MEGA7 (Kumar et al., 2016) possible, heterozygous bases were substituted for both nucleotide bases constituting the aforementioned heterozygous IUPAC code (Tavanti, Davidson, Fordyce, et al., 2005). For example, W was replaced with AT and M with AC, etc. Where homozygous bases were present at the respective loci in sequences of different

isolates, that same base was duplicated; A replaced with AA, T with TT, etc. to ensure equal length of the sequences for alignment and phylogenetic analysis.

Table 3.2 Primer sequences for MLST.

Name	**Orientation**	**Sequence (5' to 3')**	**Amplicon size (bp)**
AAT1a	Forward	ACTCAAGCTAGATTTTTGGC	478
	Reverse	CAGCAACATGATTAGCCC	
ACC1	Forward	GCAAGAGAAATTTTAATTCAATG	519
	Reverse	TTCATCAACATCATCCAAGTG	
ADP1	Forward	GAGCCAAGTATGAATGATTTG	537
	Reverse	TTGATCAACAAACCCGATAAT	
MPIb	Forward	ACCAGAAATGGCCATTGC	486
	Reverse	GCAGCCATGCATTCAATTAT	
SYA1	Forward	AGAAGAATTGTTGCTGTTACTG	543
	Reverse	GTTACCTTTACCACCAGCTTT	
VPS13	Forward	TCGTTGAGAGATATTCGACTT	741
	Reverse	ACGGATGGATCTCCAGTCC	
ZWF1b	Forward	GTTTCATTTGATCCTGAAGC	703
	Reverse	GCCATTGATAAGTACCTGGAT	

3.4 Antifungal susceptibility testing.

Antifungal susceptibility testing was performed by CLSI described methods (NCCLS, 2004). Isolates were cultured on CHROMagar Candida for two days at 37°C in an incubator. Cells were suspended in sterile saline to create a suspension with a turbidity of 0.5 McFarland, resulting in a cell suspension of 1 x 10^6 to 5 x 10^6 cells per mL (NCCLS, 2004). A sterile cotton swab was dipped into the suspension and excess liquid wrung out by pressing the cotton swab against the inner wall of the suspension container. The cotton swab was streaked across the surface of a Mueller Hinton agar plate (Becton Dickinson) supplemented with 2% glucose and 0.5µg/ml methylene blue, prepared as per NCCLS (2004) guidelines. The plate was rotated 60° and the streaking was repeated twice to ensure formation of a uniform lawn of cells on the surface of the agar. Using sterile forceps, antifungal impregnated discs containing 25µg fluconazole and 100 units nystatin (Oxoid) respectively were placed onto the surface of the agar, with equal distance between both discs and the edge of the petri dish. The plates were incubated for 21 hours at 37°C in an incubator. After the incubation period, the diameter of the clear zone surrounding the discs was measured using a ruler and the results recorded. Zone diameter interpretation was performed based on values in published articles, as listed in table 3.3. (NCCLS, 2004; Khan et al., 2018).

Table 3.3 Zone diameter result interpretation for antifungal susceptibility testing.

Antifungal	**Disc concentration**	**Zone diameter (mm)**		
		Resistant	Susceptible dose-dependent	Susceptible
Nystatin	100 U	≤10	10-14	≥15
Fluconazole	25 µg	≤14	15-18	≥19

3.5 Virulence gene expression

Virulence genes were identified from previous literature and five; *CAT1, EFG1, SAP1, ALS3* and *HWP1* were selected for analysis (Jahanshiri et al., 2019; Fonzi and Calderone, 2001; Alonso et al., 2018). Primer sequences were either obtained from published papers or self-designed using NCBI Primer Blast, as listed in Table 3.4. Three housekeeping genes; *ACT1, RPP2B* and *PMA1* were selected for generating a standard curve. qPCR was performed using a Qiagen Rotor Gene Q and the results were analysed using REST 2009. SYBR® Green qPCR master mix (Qiagen) was used to prepare PCR reactions with respective primers in 0.1 mL tubes (Qiagen) suitable for use with the Qiagen Rotor Gene Q. The reaction component volumes are as described in Table 3.5. 22 *C. albicans* isolates were selected for the qPCR assay based on variability in the MLST profiles with the aim of minimizing genetic diversity in this context.

Table 3.4 Primer sequences for qPCR virulence gene expression assay.

Name	**Orientation**	**Sequence (5' to 3')**	**Product size (bp)**
ACT1 **(Alves et al., 2014)**	Forward	TGCTGAACGTATGCAAAAGG	186
	Reverse	TGAACAATGGATGGACCAGA	
RPP2B	Forward	ATATCACCGCTTTATTGGAATCCG	136
	Reverse	GATGGGACAGAAGCTAATTTGG	
PMA1	Forward	CCAGAGCTCAAGGTCCATTCT	257
	Reverse	AGTGTGTTGTTGTGGTTTTCTAC	
CAT1	Forward	GACTGCTTACATTCAAAC	117
	Reverse	AACTTACCAAATCTTCTCA	
EFG1	Forward	ACGAGTAACAACTACCAT	89
	Reverse	TATCTGCTCTTCTGACAA	
SAP1	Forward	CCTCGTCCTGGTCAATCAGC	258
	Reverse	ACCAGCAGCCTCATTGGTTT	
ALS3	Forward	TATTTGTCGGTTGCGACTGC	108
	Reverse	TGGGGTTCCTGGTCCCTTAT	
HWP1	Forward	CCGGAATCTAGTGCTGTCGT	86
	Reverse	TTGGCAGATGGTTGCATGAG	

Table 3.5 qPCR reaction components.

Component	Volume (µl)
Master mix	5.0
Forward primer	0.2
Reverse primer	0.2
cDNA	1.0
H_2O	3.6
Total	10.0

3.6 Biofilm formation assay

22 *C. albicans* isolates were selected for the biofilm formation assay based on variability in the MLST profiles with the aim of minimizing genetic diversity in this context. *C. albicans* isolates retrieved from glycerol stocks stored in a -80°C ultra-low temperature freezer were thawed on a cold block for 10 minutes prior to culturing on CHROMagar Candida for 2 days to ensure viability. Single colonies of *C. albicans* were inoculated in 7 ml Mueller Hinton broth for 48 hours in a shaker incubator at 37°C with a rotation speed of 200 rpm. After incubation, the culture tubes were centrifuged at 3000 x *g* for ten minutes. Broth was poured out of the tubes and the pellet was resuspended in phosphate buffered saline for washing. The culture tubes were again centrifuged and the washing process was repeated once more. After the second centrifugation, the pellet was resuspended in fresh Mueller Hinton Broth to create a suspension with a turbidity of 5.0 McFarland. 100µl of this

suspension was pipetted into a well of a 96-well plate. Each isolate was tested in triplicates, with blanks consisting of sterile Mueller Hinton Broth. Replications were performed in a single run to prevent variation between batches in terms of quality of culture medium and other reagents as well as prevention of changes in incubation conditions. Subsequent steps of the assay were also applied to blank wells. The inoculated 96-well plate was incubated in an incubator at 37°C for 90 minutes. After incubation, the plate was removed and each well was washed twice with 150µl of phosphate buffered saline to remove any unadhered cells. 100µl of fresh Mueller Hinton broth was then pipetted into the wells and the plate was incubated at 37°C for 48 hours. Broth was replaced daily. After the incubation period, each well was washed twice with 200µl of phosphate buffered saline and left to air dry for 45 minutes. 110µl of 0.1% crystal violet was pipetted into each well and the plate was incubated at room temperature for 45 minutes. Each well was then washed with 200µl of sterile distilled water to remove excess stain and immediately destained with 200µl of 95% ethanol. The plate was incubated at room temperature for 45 minutes. 100µl of the ethanol was transferred to a new 96-well plate and the optical density at a wavelength of 595 nm of the suspension in each well was measured using a microplate reader.

3.7 Data analysis

Results were analysed using Graphpad PRISM and Microsoft Excel. The statistical analysis performed were the T-test and Fisher's Exact Test. MEGA7 (Kumar et al., 2016) was used for UPGMA phylogenetic analysis with bootstrapping (1000

iterations). The Hunter Gaston Discriminatory Index (HGDI) (Hunter and Gaston, 1988) and 95% confidence intervals of the respective genotyping methods was determined by using an online software, Comparing Partitions, available at http://www.comparingpartitions.info/?link=Tool (Carriço et al., 2006). The adjusted Rand coefficient was also calculated using the same software to investigate the congruence of the results between the genotyping methods applied, with 1 indicating both methods generate completely concordant genotyping results and 0 indicating both methods generate completely discordant findings.

CHAPTER 4

RESULTS

4.1 Identification of *C. albicans*

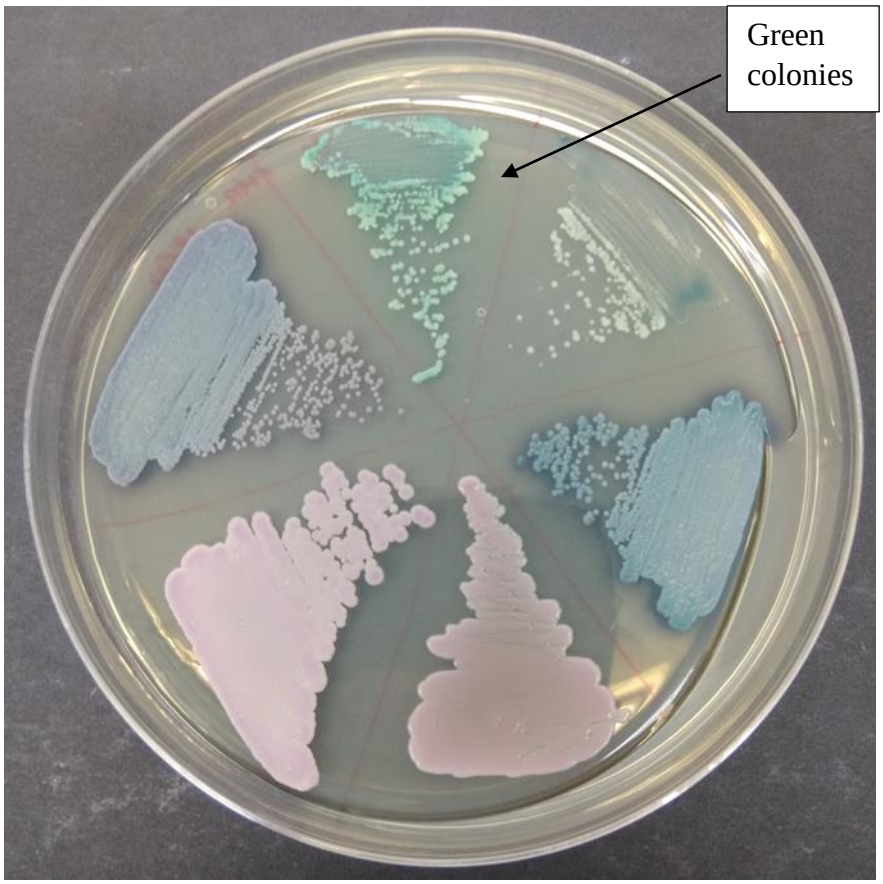


Figure 4.1 Various *Candida spp.* streaked on CHROMagar Candida. Green colonies, as seen in the top division, were selected for subsequent analysis.

As described by the product manufacturer, *C. albicans* formed green colonies on CHROMagar Candida. Besides *C. albicans*, *C. dubliniensis* also formed green colonies, thus necessitating further identification techniques.

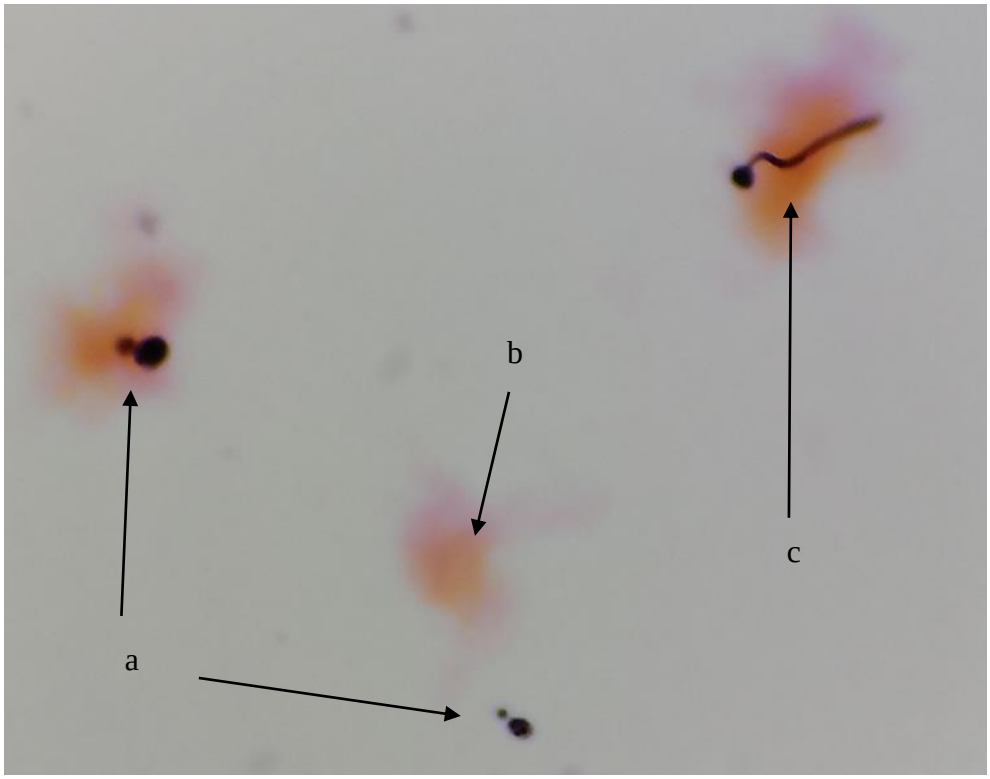


Figure 4.2 *C. albicans* cultured in human serum at 37°C for 3 hours. (a) *C. albicans* in budding yeast form. (b) Artifact from human plasma. (c) *C. albicans* germ tube.

The germ tube technique was applied to isolates which had formed green colonies on CHROMagar Candida. As seen in figure 4.2(c), *C. albicans* formed germ tubes. Given the genetic similarity between *C. albicans* and *C. dubliniensis*, both are known to form germ tubes. As such, molecular identification continued with ITS PCR.

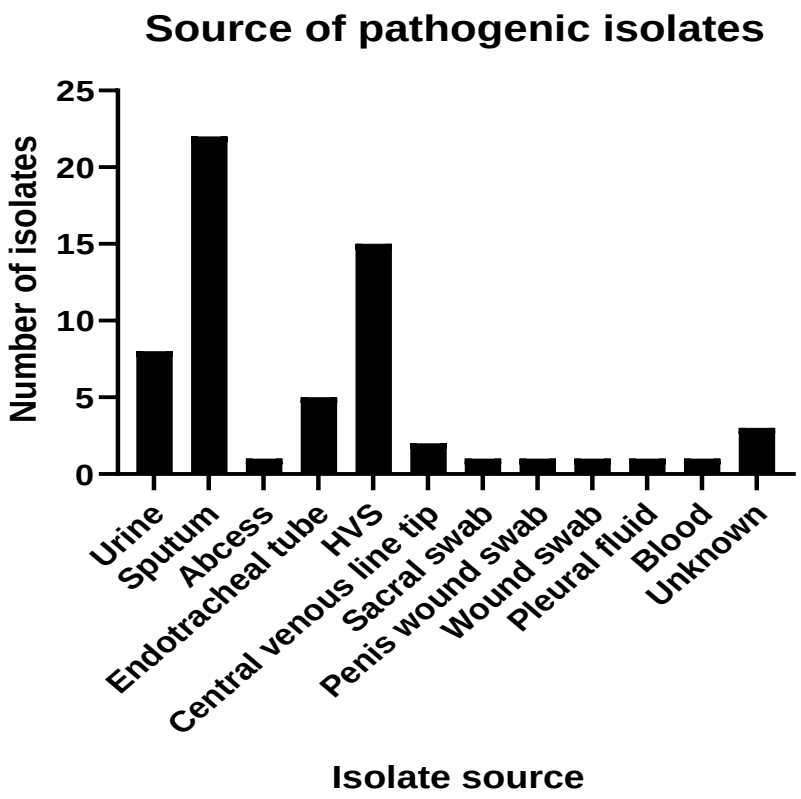


Figure 4.3 Distribution of pathogenic isolates by source.

Isolates of pathogenic origin were sampled from a variety of body sites as demonstrated in Figure 4.3. The most common source of pathogenic isolates in this study was from sputum (22), followed by high vaginal swabs (HVS) (15) and urine (8). All 50 of the commensal origin isolates were obtained via buccal swabs.

4.2 Distribution of ABC genotypes of *Candida albicans*

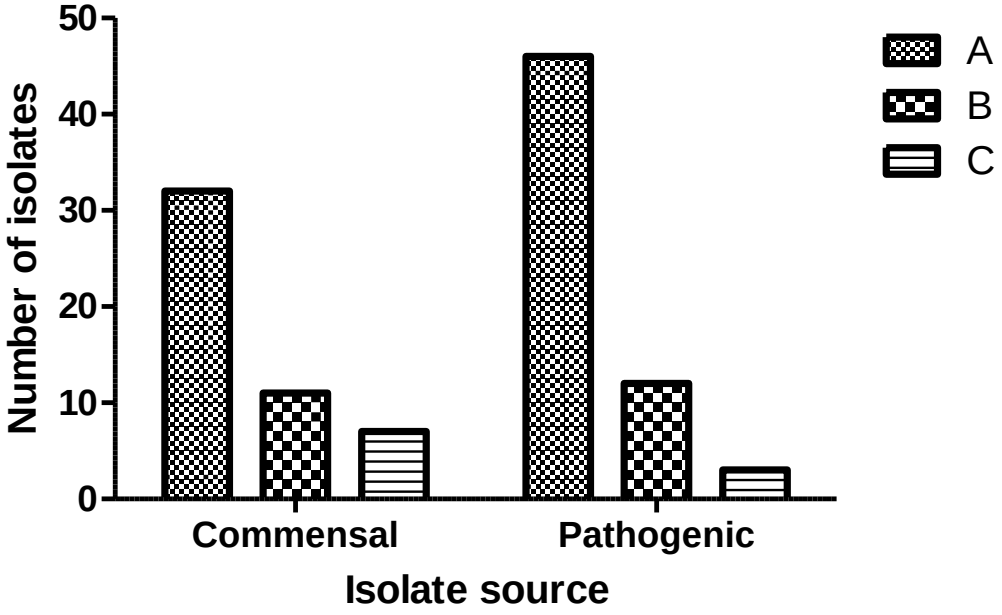


Figure 4.4 Distribution of ABC genotypes of C. *albicans* commensal and pathogenic isolates.

As seen in figure 4.4, genotype A was the most common among all 111 isolates studied, making up 70.27%. Of the 50 commensal isolates studied, 32 were genotype A, 11 genotype B and 7 genotype C. In the pathogenic group, 46 were genotype A, 12 genotype B and 3 genotype C.

4.3 Distribution of ABC genotypes and subtypes of *Candida albicans*

Table 4.1 Distribution of ABC genotype and ALT sequence subtype of *C. albicans*.

Genotype/Subtype	Pathogenic	Commensal
A1/3	2	0
A2	2	0
A3	16	20
A4	1	0
A2/3	3	1
A3/4	15	6
A4/5	1	0
A2/3/4	5	5
A3/4/5	1	0
B3	2	4
B4	3	0
B2/3	1	0
B3/4	5	7
B4/5	1	0
C3	1	1
C4	0	1
C3/4	1	5
C3/4/5	1	0
Total	61	50

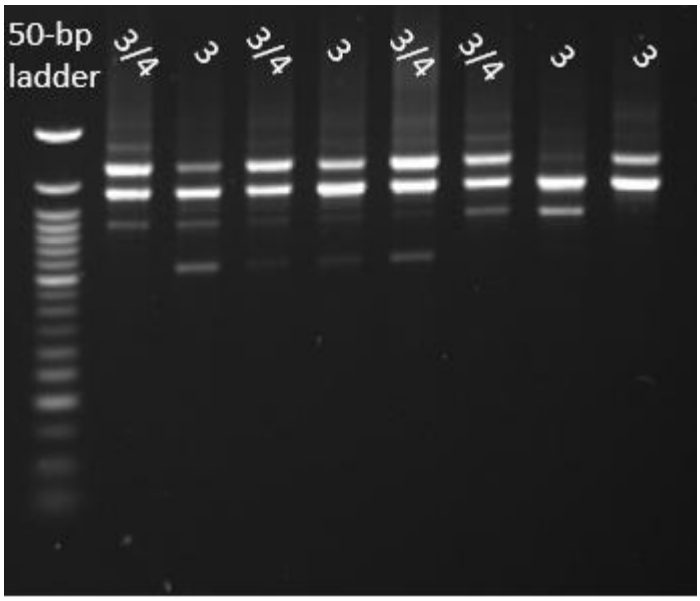


Figure 4.5 Agarose gel electrophoresis results showing some ALT subtypes of *C. albicans.*

As displayed in Table 4.1, nine different ALT sequence based genotypes were found in this study. The most common in the commensal group was subtype 3 (25 isolates) while subtype 3/4 had the highest number in the pathogenic group (21 isolates). The second highest subtypes in the commensal and pathogenic groups were 3/4 and 3 respectively. Greater variation in this genotyping method was found in the pathogenic group with nine different genotypes whereas the commensal group only had five genotypes. A combination of ABC genotyping data and ALT sequence genotyping indicated that genotype A3 was highest in both commensal and pathogenic groups, representing 20 and 16 isolates respectively. Genotype B3/4 represented 7 commensal isolates (11.47%), the second most abundant genotype in the group. In the pathogenic group A3/4 was the second most common genotype comprising of 15 isolates or 30%.

4.4 MLST: UPGMA phylogenetic tree using concatenated sequences of 7 housekeeping genes

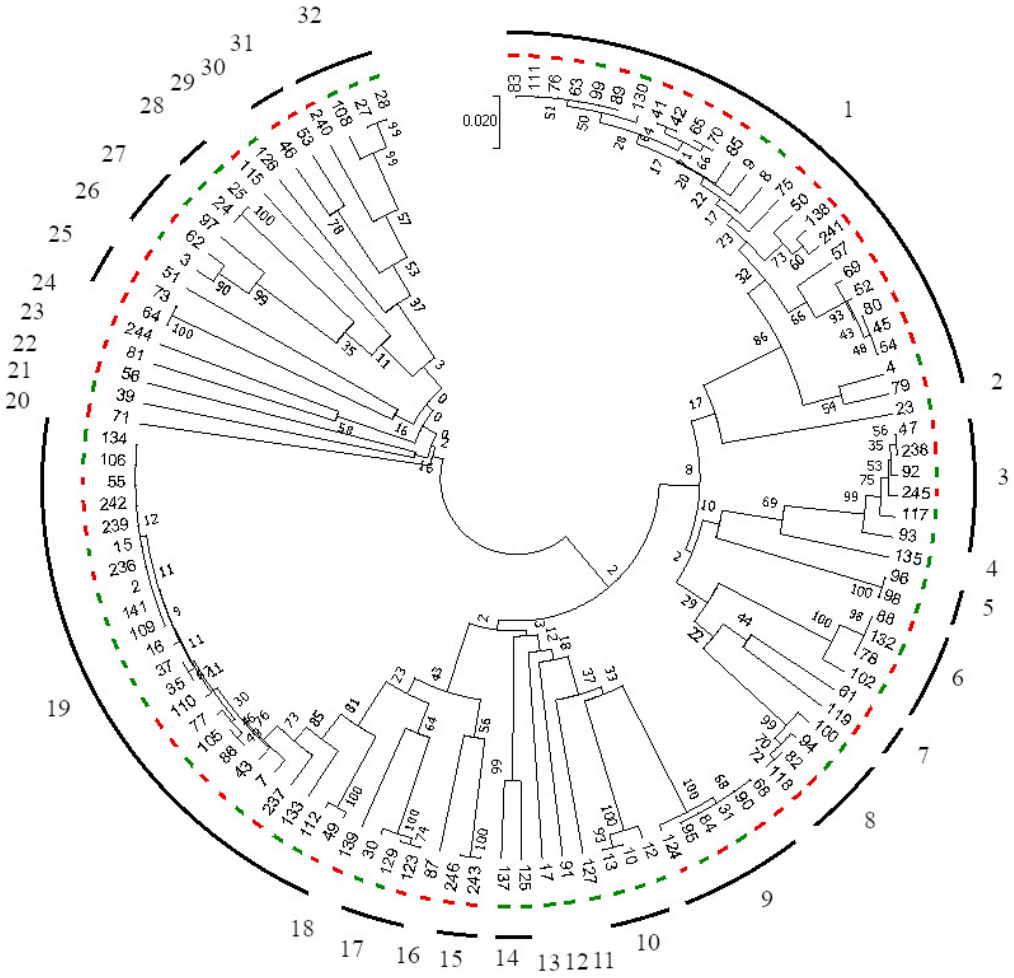


Figure 4.6 UPGMA phylogenetic tree of MLST data. Red bars indicate isolate of pathogenic origin, green bars indicate isolate of commensal origin and black bars indicate clade designation.

MLST has shown great variety in discrimination between strains of *C. albicans*. Of the 111 isolates, only 45 isolates or 40.54% could be matched to existing entries in the international online database. In contrast, a total of 66 isolates analysed had either a novel allelic profile (50 isolates) or novel SNPs in one or more alleles (16 isolates). Based on concatenated sequences of the 7 housekeeping genes used in MLST, the phylogenetic tree in figure 4.6 was constructed. The black bars surrounding the tree indicate separate clades which were determined by a p-distance of p=0.04 (Odds et al., 2007). Any isolates without a black bar were not closely related to neighbouring isolates and were considered singletons in this analysis constituting their own respective clades. In figure 4.6, clade numbers were assigned arbitrarily. Clade 1 was the largest found in this analysis, consisting of 26 isolates in total (21 pathogenic and 5 commensal) followed by clade 19 which consists of 23 isolates (10 pathogenic and 13 commensal). The most common sequence type found was ST365, belonging to 10 isolates (134, 106, 55, 242, 239, 15, 236, 2, 141, 109), which in the phylogenetic tree is located in clade 19.

Table 4.2. Genotype profiles of isolates selected for biofilm and qPCR assays.

ID	Type	ABC/Subtype	MLST	AAT1a	ACC1	ADP1	MPIb	SYA1	VPS13	ZWF1b
7	Commensal	A3	2716	55	14	4	3	24	45	15
9	Commensal	A2/3/4	184	3	5	5	2	2	6	5
16	Commensal	A3	3286	55	14	4	3	155	45	15
35	Commensal	A/3	1901	55	14	14	3	6	45	15
37	Commensal	A/3	1901	55	14	14	3	6	45	15
105	Commensal	A/3	1754	55	3	4	3	6	45	15
112	Commensal	A3/4	90	25	7	6	3	6	27	37
132	Commensal	B3	443	59	5	21	2	80	108	15
106	Commensal	A3	365	55	14	4	3	6	45	15
109	Commensal	A3	365	55	14	4	3	6	45	15
134	Commensal	A3	365	55	14	4	3	6	45	15
236	Pathogenic	A3	365	55	14	4	3	6	45	15
237	Pathogenic	A3	365	55	14	4	3	6	45	12
238	Pathogenic	A1/3	1097	13	26	5	3	93	53	12
239	Pathogenic	A3	365	55	14	4	3	6	45	15
240	Pathogenic	B3	No match	60	13	21	1	7	114	New
241	Pathogenic	A2/3/4	572	8	5	5	2	2	6	5
242	Pathogenic	A3	365	55	14	4	3	6	45	15
243	Pathogenic	A2/3	No match	New	4	54	3	31	47	12
244	Pathogenic	B2/3	No match	87	35	4	106	13	261	244
245	Pathogenic	A1/3	No match	13	26	5	3	93	22	12
246	Pathogenic	A2/3	No match	New	4	54	3	31	47	12

4.5 Association between MLST clade 1 and pathogenicity

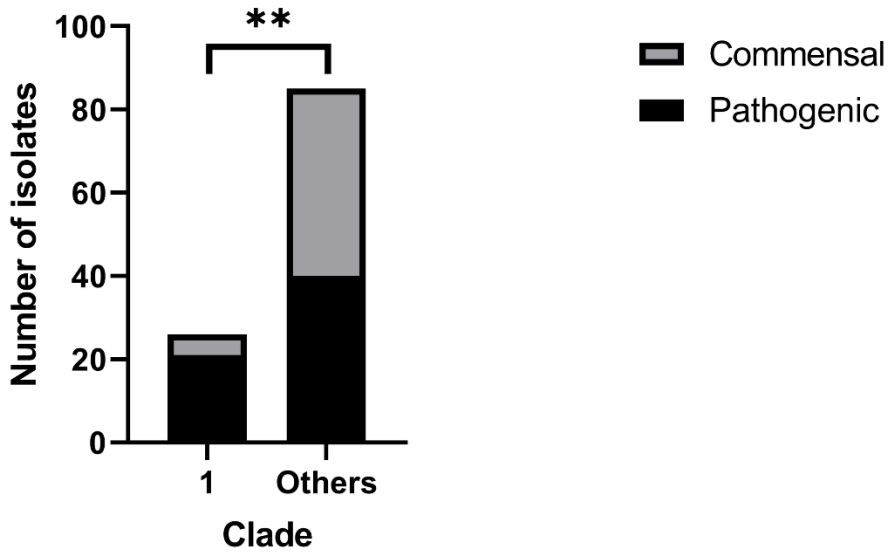


Figure 4.7 Bar chart depicting the distribution of commensal and pathogenic isolates in clade 1 and other clades.

Based on the genetic relatedness determined from the phylogenetic tree in figure 4.6, statistical analysis was done to investigate the possibility of association between the most abundantly populated clade 1 and prevalence in a clinical setting. As shown in figure 4.7, a statistically significant association was found. The clade 1 consists of 21 pathogenic isolates and 5 commensal isolates. For comparison, the second most abundantly occupied clade 19 consisted of 10 pathogenic and 13 commensal isolates, as described in figure 4.6. The statistical analysis applied to distribution in clade 1 also yielded an Odds ratio of 4.725.

4.6 Average zone diameters from antifungal susceptibility testing

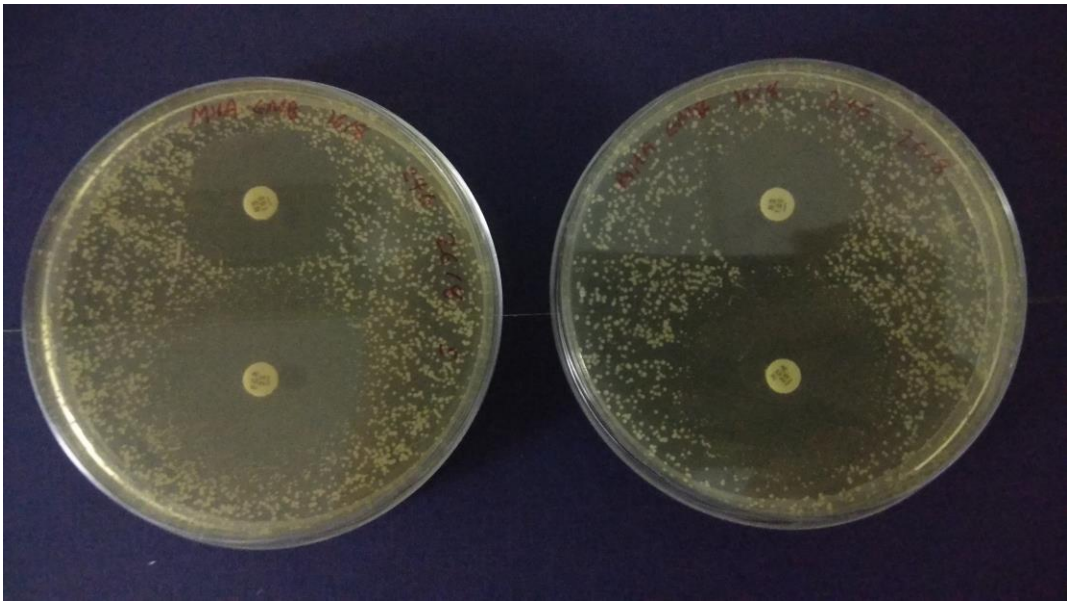


Figure 4.8 Example of results of *C. albicans* antifungal susceptibility testing using the disk diffusion method.

As seen in figure 4.8, *C. albicans* will form creamy colonies on Mueller Hinton Agar supplemented with glucose and methylene blue. Diffusion of the antifungal compounds from the impregnated discs placed on the agar surface results in inhibition of growth of susceptible strains of *C. albicans* in the zone surrounding the discs. This area is termed the zone of inhibition and is the measure of susceptibility for this method of antifungal susceptibility testing.

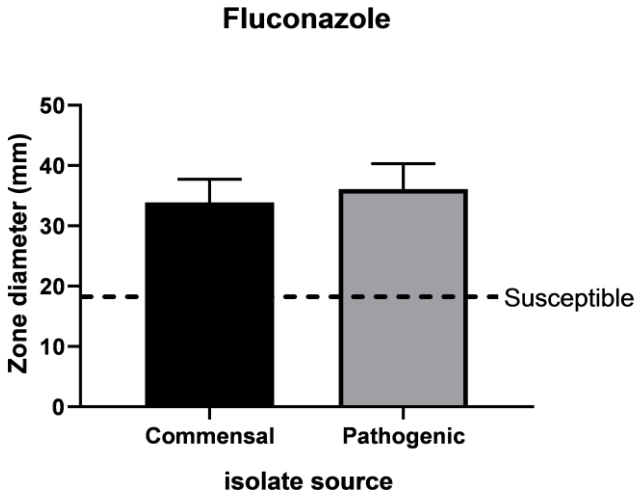


Figure 4.9 Bar charts showing the average zone diameters obtained in antifungal susceptibility testing using fluconazole.

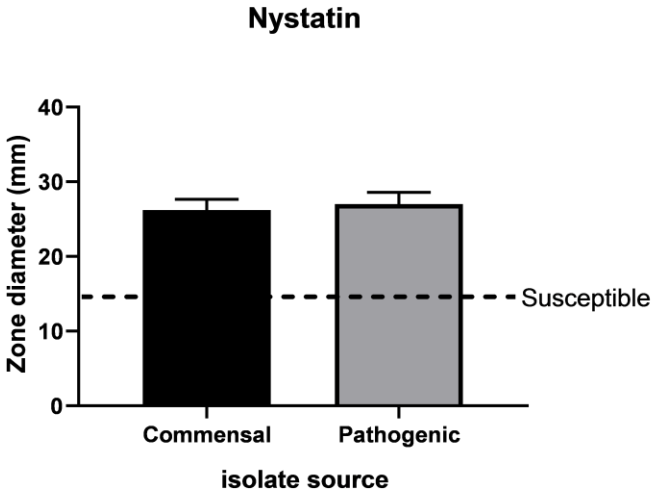


Figure 4.10 Bar chart showing the average zone diameters obtained in antifungal susceptibility testing using nystatin.

As indicated in figure 4.9 and figure 4.10, all isolates tested for antifungal susceptibility were found to be susceptible to both nystatin and fluconazole. Commensal isolates, on average, were found to be less susceptible to both antifungals compared to the pathogenic isolates.

4.7 Relative gene expression of *HWP1* gene in commensal and pathogenic isolates

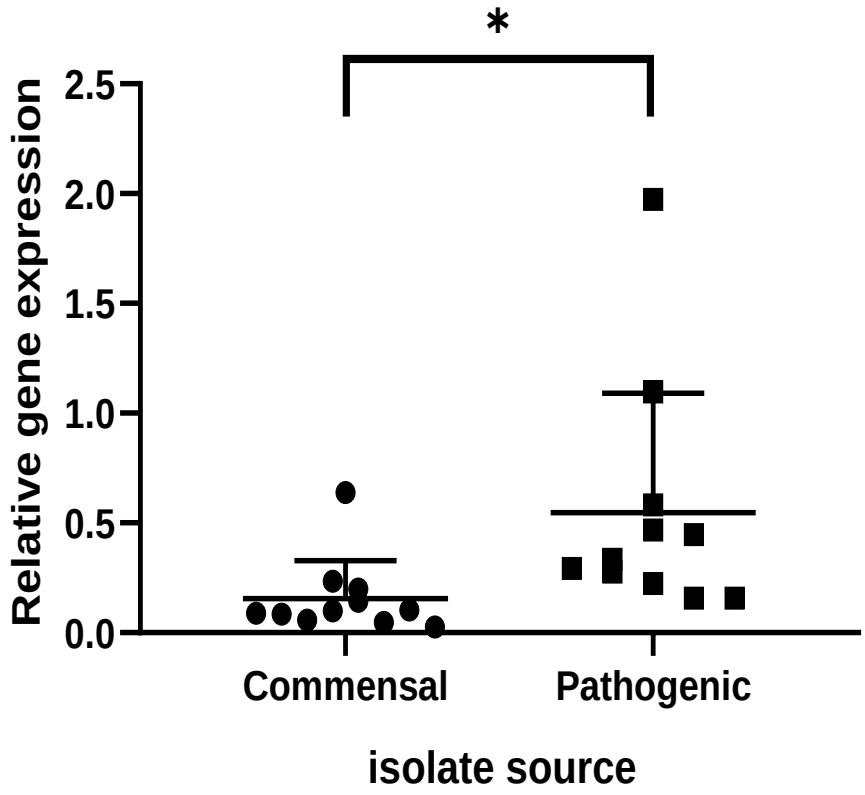


Figure 4.11 Scatter plot of relative gene expression of *HWP1* in commensal and pathogenic isolates. T-test p-value <0.05.

qPCR data indicated an upregulation of expression of *HWP1* in the pathogenic group, as shown in figure 4.11, with statistical significance compared to the commensal group. The individual dots on the scatter plot represent separate data points for their respective isolates and visual representation of the data showed a large number of the commensal isolates have reduced expression of *HWP1* compared to the pathogenic isolates. For the genes *CAT1*, *EFG1*, *ALS3* and *SAP1*, no statistically significant difference was found in relative expression levels of both these groups.

4.8 Biofilm formation assay for commensal and pathogenic isolates

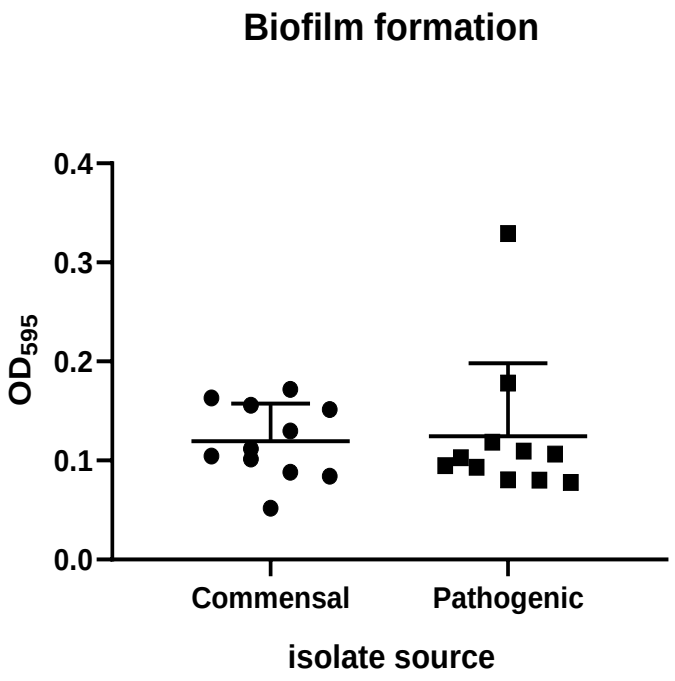


Figure 4.12 Scatter plot of biofilm quantification by optical density measurement of crystal violet staining in commensal and pathogenic isolates.

OD_{595} readings of crystal violet stains are shown in figure 4.12. Dots in the scatter plot represent individual data points which were the average readings of 3 different culture wells of the same isolate in a 96-well microplate. No statistically significant difference was found in the *in vitro* biofilm forming abilities between the two groups.

CHAPTER 5

DISCUSSION

Of the 61 pathogenic isolates included in this study, 46 (75%) were genotype A, 12 (20%) were genotype B and 3 (4%) were genotype C. The 50 isolates in the commensal group consisted of 32 (64%) genotype A, 11 (22%) genotype B and 7 (14%) genotype C. Genotype C was the least common genotype found in this study, as shown in Figure 4.4, which is consistent with findings in other regions across the globe (Dalvand et al., 2018; Zida et al., 2018; Tamai et al., 2014). There was no significant association found between genotype and isolate source, whether pathogenic or commensal. This may be due to the wide variety of infected sites sampled for the pathogenic group, as seen in Figure 4.3 such as high vaginal swabs, blood, urine and sputum, among others, while the commensal group was isolated only from buccal swabs of individuals without symptoms of candidiasis. Samples were taken from asymptomatic individuals as this is an indicator that *C. albicans* present in their oral cavities are not expressing pathogenicity, thus providing a clear differentiation in host dwelling conditions between the two groups. It is possible that the genetic diversity of commensal *C. albicans* present in other body sites such as the vagina or intestinal tract would not match that of the oral cavity. The grouping of *C. albicans* ABC genotypes and subtypes in this study is consistent with the findings of other workers, where genotype A3 was found to be the most common (Sawadogo et al., 2019; Iwata et al., 2006). In a more targeted study conducted in Thailand, researchers discovered that *C. albicans* genotype B was more commonly present that *C. albicans* genotype

A in oral cavities of patients experiencing oral candidiasis as well as asymptomatic carriers (Tantivitayakul et al., 2019). This indicates that certain genotypes may have a tendency to exhibit pathogenicity in particular ways, such as a systemic or superficial infection, or in varying sites of the host body. Such focused studies are beneficial for determination of this association, which could not be done in this study due to the varying sources of isolates as mentioned earlier.

All 111 isolates were subjected to multi-locus sequence typing using the globally established standard of 7 housekeeping genes. Of these isolates, 45 had a match to an existing database entry for allelic profiles while the remaining 66 were found to have either a novel allelic profile or a completely new allele number for at least one of seven alleles studied, indicating a novel sequence type. Typically, a clonal cluster diagram is constructed based on the allelic profiles of the isolates which can be obtained by matching DNA sequencing results with database entries available online. Since obtaining a new allele number or sequence type from the database curator requires submission of the isolates, the typical method of analysis was not possible. Instead, a second method was employed which has also been used to analyse this type of data. SNP loci were identified and sequence data extracted, edited, aligned and used for phylogenetic analysis (Tavanti, Davidson, Fordyce, et al., 2005). An UPGMA phylogenetic tree was constructed to illustrate genetic relatedness between the different isolates with a p-distance cut-off of 0.04 to define clades (Odds et al., 2007). The largest clades in this study, numbered arbitrarily, are 1 and 19, with clade 1 containing the largest number of pathogenic isolates (21). A Fisher's Exact test found a statistically significant association between clade 1

and prevalence in a clinical setting in this study. This statistical analysis also yielded an Odds ratio of 4.725, indicating that *C. albicans* isolates in clade 1 were 4.725 times more likely to be observed in a clinical setting which suggests that they could have a greater tendency to cause disease.

MLST generated a significantly higher score (p-value<0.001) in the HGDI (Hunter and Gaston, 1988), at 0.985 (95% confidence interval: 0.976-0.995), compared to ABC and ALT sequence typing which scored only 0.837. This indicates that MLST is a very powerful genotyping method capable of capturing the great genetic diversity of *C. albicans*. Ideally, concordant genotyping results are expected to have strains characterized as one genotype by a less discriminatory method (ABC+ALT) to be further subdivided into more genotypes by another more discriminatory method (MLST). However, the findings from both genotyping methods appear to be discordant (adjusted Rand coefficient = 0.221). For example, the largest ABC+ALT genotype, A3 (36 members), can be further divided into 11 MLST clades (clade 1, 5, 8, 12, 14, 17, 19, 21, 25, 29, and 30). In contrast, the largest clade (clade 1 with 26 members) identified by the supposedly more discriminatory MLST method, can be further divided into 8 ABC+ALT genotypes (A2, A2/3, A2/3/4, A3, A3/4, A3/4/5, A4/5, and C3/4/5). This is not entirely unexpected as both methods utilized different genes and interpretations to genotype the *C. albicans* strains. Differentiation by clades in the UPGMA phylogenetic tree, utilizing a p-distance cut-off of 0.04 yielded a HGDI value of 0.895 (95% confidence interval: 0.858-0.932). As such, even using the alternative method of SNP analysis instead of clustering by allelic profile, the MLST is still a better

discriminatory tool than ABC and ALT sequence typing combined (p-value = 0.027). In addition, the unambiguity of using sequencing data for analysis as compared to visualisation of band intensity via agarose gel electrophoresis is a significant strengthening factor for the use of MLST.

The antifungal susceptibility testing showed that all isolates in this study are susceptible to nystatin and fluconazole. It is possible that resistance levels may have changed over time due to differences in sample handling and source. As the isolates are from the laboratory bio-bank, it could not be determined whether the isolates obtained from the hospital originated in patients already on medication or the number of sub-cultures performed by the hospital laboratory staff for identification of the pathogen.

For qPCR, 11 isolates originating from an intensive care unit and a high dependency unit as well as 11 isolates from old folk's homes that had MLST genotypes similar to the ICU/HDU group were selected to determine whether there was a difference in expression of virulence by isolates of the same or similar genotype from either group. Based on the qPCR results, it appears that sequence type alone may be insufficient as a discriminatory factor of expression of virulence. Of the five virulence genes chosen for analysis, only *HWP1* yielded a statistically significant difference in relative gene expression between the two groups. The pathogenic group showed a significant upregulation in relative gene expression compared to the commensal group, strengthening the statement that hyphal wall proteins are essential for expression of virulence. Germ tube formation and complex hyphal structures are excellent morphological structures of *C. albicans* for

penetrating host cells and causing localized damage to host tissues (Desai, 2018). The other four virulence genes analysed yielded no statistically significant difference in the relative gene expression between the two groups. This too may be due to handling methods of the isolates obtained from hospital patients. It is unclear how many passages were performed by the hospital laboratory technicians prior to confirming diagnosis of candidiasis and what incubation and sample storage conditions were. The use of archived isolates is a major limitation in this study.

A biofilm formation assay was conducted using the same 22 isolates as in the qPCR experiment. There was no statistically significant difference in the *in vitro* biofilm forming abilities of isolates, as measured by cell mass, in both groups. Biofilm formation is common on catheters, endotracheal tubes or other external medical devices inserted into the body for extended periods of time. Devices such as ventricular shunts, when contaminated, also have the risk of allowing pathogens to access patient's bodies, potentially resulting in life-threatening systemic infections. In this study, a modified version of different types of assays was implemented based on reagent availability. This was performed since there is no international standard for biofilm assays for *C. albicans*, where procedures have a wide range of growth mediums, incubation temperatures, shaking speeds, carbon dioxide concentration, measurement methods and biofilm adherence surfaces. Given that *C. albicans* biofilms may form on mucosal surfaces and inert materials present in medical devices (Pereira et al., 2020), it is possible that the surface used for cell adherence in this assay was unsuitable. In addition, no other published works have used Mueller Hinton Broth for culturing *C. albicans* biofilms. It may

however be a suitable medium to explore given its high protein content which has been used to produce *C. albicans* biofilms (Rimek et al., 2008). Further analysis is necessary for determination of the suitability of Mueller Hinton Broth as a medium for propagation of *C. albicans* biofilms.

CHAPTER 6

CONCLUSION

In conclusion, this study has shown that there is vast diversity in *C. albicans* genotypes using MLST. Using 25S rDNA genotyping and ALT sequence typing, the diversity was shown to be similar to that in other regions. As shown by the Hunter Gaston Discriminatory Index, MLST is a more powerful genotyping method for documenting genetic diversity. Considering that there are over 3000 sequence types documented in the online database, primarily from areas outside of Southeast Asia, it is possible that local regions have undocumented loci variations. While there was no statistically significant association between 25S rDNA genotype and expression of pathogenicity, there was a statistically significant association between expression of pathogenicity and a specific clade of *C. albicans* sequence types. In general, there was very little variation in antifungal susceptibility and expression of virulence genes, with the exception of the *HWP1* gene. As such, in the future it would be of particular importance to conduct further investigation into different parts of the genomes of the single clade isolates to determine possible genetic markers which enhance pathogenicity.

REFERENCES

Aldardeer, N.F. et al., 2020. Antifungal resistance in patients with Candidaemia: A retrospective cohort study. *BMC Infectious Diseases*, 20(1), pp.1–7.

Almeida, R.S. et al., 2008. The hyphal-associated adhesin and invasin Als3 of Candida albicans mediates iron acquisition from host ferritin. *PLoS Pathogens*, 4(11), pp.1–17.

Alonso, G.C., Pavarina, A.C., Sousa, T. V. and Klein, M.I., 2018. A quest to find good primers for gene expression analysis of Candida albicans from clinical samples. *Journal of Microbiological Methods*, 147(August 2017), pp.1–13. Available at: https://doi.org/10.1016/j.mimet.2018.02.010.

Alves, C.T. et al., 2014. Candida albicans promotes invasion and colonisation of Candida glabrata in a reconstituted human vaginal epithelium. *Journal of Infection*. Available at: http://dx.doi.org/10.1016/j.jinf.2014.06.002.

Bain, J.M., Stubberfield, C. and Gow, N.A.R., 2001. Ura-status-dependent adhesion of Candida albicans mutants. *FEMS Microbiology Letters*, 204(2), pp.323–328.

Behzadi, P., Behzadi, E. and Ranjbar, R., 2015. Urinary tract infections and candida albicans. *Central European Journal of Urology*, 68(1), pp.96–101.

Bougnoux, M.E. et al., 2003. Collaborative Consensus for Optimized Multilocus Sequence Typing of Candida albicans. *Journal of Clinical Microbiology*, 41(11), pp.5265–5266.

Bougnoux, M.E., Morand, S. and D'Enfert, C., 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of Candida albicans. *Journal of Clinical Microbiology*, 40(4), pp.1290–1297.

Brand, A. et al., 2004. Ectopic expression of URA3 can influence the virulence phenotypes and proteome of Candida albicans but can be overcome by targeted reintegration of URA3 at the RPS10 locus. *Eukaryotic Cell*, 3(4), pp.900–909.

Braun, B.R. and Johnson, A.D., 2000. TUP1, CPH1 and EFG1 make independent contributions to filamentation in Candida albicans. *Genetics*, 155(1), pp.57–67.

Brown, G.D. et al., 2012. Hidden killers: Human fungal infections. *Science Translational Medicine*, 4(165), pp.1–10.

Carriço, J.A. et al., 2006. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant Streptococcus pyogenes. *Journal of Clinical Microbiology*, 44(7), pp.2524–2532.

Chong, P.P., Lee, Y.L., Tan, B.C. and Ng, K.P., 2003. Genetic relatedness of Candida strains isolated from women with vaginal candidiasis in Malaysia. *Journal of Medical Microbiology*, 52(8), pp.657–666.

Clemons, K. V., Feroze, F., Holmberg, K. and Stevens, D.A., 1997. Comparative analysis of genetic variability among Candida albicans isolates from different geographic locales by three genotypic methods. *Journal of Clinical Microbiology*, 35(6), pp.1332–1336.

Correia, A. et al., 2010. Limited role of secreted aspartyl proteinases Sap1 to Sap6 in Candida albicans virulence and host immune response in murine hematogenously disseminated candidiasis. *Infection and Immunity*, 78(11), pp.4839–4849.

Dalvand, A., Katiraee, F., Joozani, R.J. and Shokri, H., 2018. Genotyping of Candida albicans isolated from animals using 25S ribosomal DNA and ALT repeats polymorphism in repetitive sequence. *Current Medical Mycology*, 4(4), pp.12–19.

Desai, J. V., 2018. Candida albicans hyphae: From growth initiation to invasion. *Journal of Fungi*, 4(1).

Dieterich, C. et al., 2002. In vitro reconstructed human epithelia reveal contributions of Candida albicans EFG1 and CPH1 to adhesion and invasion. *Microbiology*, 148(2), pp.497–506.

Fonzi, W.A. and Calderone, R.A., 2001. Virulence factors of Candida albicans. *Trends in Microbiology*, 9(7), pp.327–335.

Gebreyohannes, G., Nyerere, A., Bii, C. and Sbhatu, D.B., 2019. Challenges of intervention, treatment, and antibiotic resistance of biofilm-forming microorganisms. *Heliyon*, 5(8), p.e02192. Available at: https://doi.org/10.1016/j.heliyon.2019.e02192.

Gow, N.A.R. and Yadav, B., 2017. Microbe profile: Candida albicans: A shape-changing, opportunistic pathogenic fungus of humans. *Microbiology*, 163(8), pp.1145–1147.

Hube, B. et al., 1997. Disruption of each of the secreted aspartyl proteinase genes SAP1, SAP2, and SAP3 of Candida albicans attenuates virulence. *Infection and Immunity*, 65(9), pp.3529–3538.

Hunter, P.R. and Gaston, M.A., 1988. Numerical index of the discriminatory ability of typing systems: An application of Simpson's index of diversity. *Journal of Clinical Microbiology*, 26(11), pp.2465–2466.

Iwata, T. et al., 2006. Genotyping of Candida albicans on the basis of polymorphisms of ALT repeats in the repetitive sequence (RPS). *Journal of Dermatological Science*, 41(1), pp.43–54.

Jahanshiri, Z. et al., 2019. Genotyping of Candida albicans isolates from oropharyngeal candidiasis in head and neck cancer patients in Iran: Molecular epidemiology and SAP2 gene expression. *Journal de Mycologie Medicale*, 29(4), pp.310–316.

Jolley, K.A., Feil, E.J. and Maiden, M.C.J., 2001. Sequence type analysis and

recombinational tests (START). *Bioinformatics*, 17(12), pp.1230–1231.

Khan, M. et al., 2018. Antifungal susceptibility testing of vulvovaginal Candida species among women attending antenatal clinic in tertiary care hospitals of Peshawar. *Infection and Drug Resistance*, 11, pp.447–456.

Krysan, D.J., 2017. The unmet clinical need of novel antifungal drugs. *Virulence*, 8(2), pp.135–137. Available at: http://dx.doi.org/10.1080/21505594.2016.1276692.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, 33(7), pp.1870–1874.

Lai, C.C. et al., 2012. Association between incidence of candidaemia and consumption of antifungal agents at a medical centre in Taiwan. *International Journal of Antimicrobial Agents*, 40(4), pp.349–353. Available at: http://dx.doi.org/10.1016/j.ijantimicag.2012.05.024.

Liu, Y. and Filler, S.G., 2011. Candida albicans Als3, a multifunctional adhesin and invasin. *Eukaryotic Cell*, 10(2), pp.168–173.

Lu, J.J. et al., 2018. DST659 genotype of Candida albicans showing positive association between biofilm formation and dominance in Taiwan. *Medical mycology*, 56(8), pp.972–978.

Maiden, M. et al., 1998. Multilocus sequence typing : A portable approach to the identification of clones within populations of. *Proc. Natl. Acad. Sci. USA*, 95(1), pp.3140–3145.

Mayer, F.L., Wilson, D. and Hube, B., 2013. Candida albicans pathogenicity mechanisms. *Virulence*, 4(2), pp.119–128.

McCullough, M.J., Clemons, K. and Stevens, D., 1999. Molecular epidemiology of the global and temporal diversity of Candida albicans. *Clinical Infectious Diseases*, 40(10), pp.827–834.

McCullough, M.J., Clemons, K. V. and Stevens, D.A., 1999. Molecular and phenotypic characterization of genotypic Candida albicans subgroups and comparison with Candida dubliniensis and Candida stellatoidea. *Journal of Clinical Microbiology*, 37(2), pp.417–421.

Naglik, J.R. et al., 2006. Candida albicans HWP1 gene expression and host antibody responses in colonization and disease. *Journal of Medical Microbiology*, 55(10), pp.1323–1327.

NCCLS, 2004. Method for antifungal disk diffusion susceptibility testing of yeasts: Approved guideline. *NCCLS document M44-A*, 24(15), p.23.

Nobile, C.J., Nett, J.E., Andes, D.R. and Mitchell, A.P., 2006. Function of Candida albicans adhesin hwp1 in biofilm formation. *Eukaryotic Cell*, 5(10), pp.1604–1610.

Odds, F.C. et al., 2007. Molecular phylogenetics of Candida albicans. *Eukaryotic Cell*, 6(6), pp.1041–1052.

Odds, F.C. and Jacobsen, M.D., 2008. Multilocus sequence typing of pathogenic Candida species. *Eukaryotic Cell*, 7(7), pp.1075–1084.

Pereira, R., dosSantos Fontenelle, R.., DeBrito, E.H.. and DeMorais, S.., 2020. Biofilm of Candida albicans formation, regulation and resistance. *Journal of Applied Microbiology*.

Phan, Q.T. et al., 2007. Als3 is a Candida albicans invasin that binds to cadherins and induces endocytosis by host cells. *PLoS Biology*, 5(3), pp.0543–0557.

Pierce, J. V. and Kumamoto, C.A., 2012. Variation in Candida albicans EFG1 expression enables host-dependent changes in colonizing fungal populations. *mBio*, 3(4), pp.1–8.

Rimek, D., Fehse, B. and Göpel, P., 2008. Evaluation of Mueller-Hinton-agar as a simple medium for the germ tube production of Candida albicans and Candida dubliniensis. *Mycoses*, 51(3), pp.205–208.

Roemer, T. and Krysan, D.J., 2014. Antifungal Drug Development: Challenges, Unmet Clinical Needs , and New Approaches. *Cold Spring Harb Perspect Med*, 4, p.a019703.

Román, E. et al., 2016. Role of catalase overproduction in drug resistance and virulence in Candida albicans. *Future Microbiology*, 11(10), pp.1279–1297.

Saranpal, C.S., Chong, P.P., Ng, K.P. and Lim, C.S.Y., 2015. Genetic relatedness of Candida albicans bloodstream infection clinical isolates in Malaysia. *Malaysian Journal of Microbiology*, 11(3), pp.294–299.

Sawadogo, P.M. et al., 2019. Genotype analysis of clinical candida albicans isolates using pcrs targeting 25s rdna and alt repeat sequences of the rps and antifungal susceptibility in ouagadougou (Burkina Faso). *Infection and Drug Resistance*, 12, pp.3859–3866.

Schaller, M. et al., 2003. The secreted aspartyl proteinases Sap1 and Sap2 cause tissue damage in an in vitro model of vaginal candidiasis based on reconstituted human vaginal epithelium. *Infection and Immunity*, 71(6), pp.3227–3234.

Tamai, I.A. et al., 2014. Repetitive sequences based on genotyping of candida albicans isolates obtained from iranian patients with human immunodeficiency virus. *Iranian Journal of Basic Medical Sciences*, 17(11), pp.831–835.

Tamura, M. et al., 2001. Molecular characterization of new clinical isolates of Candida albicans and C. dubliniensis in Japan: Analysis reveals a new genotype of C. albicans with group I intron. *Journal of Clinical Microbiology*, 39(12), pp.4309–4315.

Tantivitayakul, P. et al., 2019. Genotyping of Candida albicans and Candida dubliniensis by 25S rDNA analysis shows association with virulence attributes in

oral candidiasis. *Archives of Oral Biology*, 97(September 2018), pp.18–24. Available at: https://doi.org/10.1016/j.archoralbio.2018.10.006.

Tavanti, A., Davidson, A.D., Johnson, E.M., et al., 2005. Multilocus Sequence Typing for Differentiation of Strains of Candida tropicalis. *Journal of Clinical Microbiology*, 43(11), pp.5593–5600.

Tavanti, A. et al., 2003. Optimization and Validation of Multilocus Sequence Typing for Candida albicans. *Journal of Clinical Microbiology*, 41(8), pp.3765–3776.

Tavanti, A., Davidson, A.D., Fordyce, M.J., et al., 2005. Population Structure and Properties of Candida albicans,as Determined by Multilocus Sequence Typing. *Journal of Clinical Microbiology*, 43(11), pp.5601–5613.

Tay, S.T. et al., 2005. Molecular subtyping of clinical isolates of Candida albicans and identification of Candida dubliniensis in Malaysia. *Mycopathologia*, 159(3), pp.325–329.

Taylor, L.H., Latham, S.M. and Woolhouse, M.E.J., 2001. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1411), pp.983–989.

Tsuchimori, N. et al., 2000. Reduced virulence of HWP1-deficient mutants of Candida albicans and their interactions with host cells. *Infection and Immunity*, 68(4), pp.1997–2002.

Tzar, M.N., Norazlah, B. and Shamsul, A.S., 2015. Risk factors for candidaemia in a Malaysian tertiary hospital. *Sains Malaysiana*, 44(5), pp.735–740.

Uppuluri, P. et al., 2010. Dispersion as an important step in the Candida albicans biofilm developmental cycle. *PLoS Pathogens*, 6(3).

Wysong, D.R. et al., 1998. Cloning and sequencing of a Candida albicans catalase gene and effects of disruption of this gene. *Infection and Immunity*, 66(5), pp.1953–1961.

Zida, A. et al., 2018. 25S rDNA genotype and antifungal susceptibility of clinical Candida albicans in Ouagadougou (Burkina Faso). *Medical Mycology*, 56(7), pp.907–910.

APPENDICES

APPENDIX A

		***AAT1a* basepair position**											
Clade	**Unique Sequence**	**87**	**108**	**120**	**150**	**169**	**204**	**244**	**405**	**430**	**432**	**441**	**453**
1	1(a)	G	T	R	T	R	Y	G	C	A	C	C	C
	1(b)	R	T	R	T	R	Y	G	C	A	C	C	C
	1(c)	G	T	R	T	R	Y	G	C	A	C	C	C
	1(d)	R	T	R	T	R	Y	G	C	A	C	C	C
	1(e)	G	T	R	T	R	Y	G	C	A	C	C	C
	1(f)	G	T	A	T	A	C	G	C	A	C	C	C
	1(g)	G	T	R	T	R	Y	G	C	A	C	C	C
	1(h)	G	T	R	T	R	Y	G	C	A	C	C	C
	1(i)	A	T	G	T	G	T	G	C	A	C	C	C
	1(j)	A	T	G	T	G	T	G	C	A	C	C	C
	1(k)	A	T	G	T	G	T	G	C	A	C	C	C
	1(l)	G	T	A	T	A	C	G	C	A	C	C	C
	1(m)	R	T	R	T	R	Y	G	C	A	C	C	C
	1(n)	G	T	R	T	R	Y	G	C	A	C	C	C
	1(o)	G	T	R	T	R	Y	G	C	A	C	C	C
	1(p)	G	T	R	T	R	Y	G	C	A	C	C	C
	1(q)	G	T	R	T	R	Y	G	C	A	C	C	C
2	2(a)	A	C	G	T	A	T	G	C	A	C	C	C
3	3(a)	A	T	A	T	A	C	G	C	A	C	C	C
	3(b)	A	T	A	T	A	C	G	C	A	C	C	C
	3(c)	A	T	A	T	A	C	G	C	A	C	C	C
	3(d)	T	T	A	T	A	C	G	C	A	C	C	C
	3(e)	A	T	A	T	A	C	G	C	A	C	C	C
4	4(a)	G	T	A	T	A	C	G	C	A	C	C	C
5	5(a)	G	T	A	T	A	C	G	C	A	C	C	C
6	6(a)	A	T	R	Y	A	Y	G	Y	A	C	C	C
	6(b)	A	T	R	Y	A	Y	G	Y	A	C	C	C
7	7(a)	A	T	A	T	A	C	G	C	A	C	C	C
	7(b)	A	T	A	C	A	C	G	C	A	C	C	C
8	8(a)	A	T	A	T	A	C	G	C	A	C	C	C
	8(b)	A	T	A	T	A	C	G	C	A	C	C	C

		AAT1a basepair position											
Clade	**Unique Sequence**	**87**	**108**	**120**	**150**	**169**	**204**	**244**	**405**	**430**	**432**	**441**	**453**
	8(c)	A	T	A	T	A	C	G	C	A	C	C	C
9	9(a)	A	Y	R	Y	A	Y	G	Y	A	C	C	C
	9(b)	A	Y	R	Y	A	Y	G	C	A	C	C	C
10	10(a)	A	T	R	T	R	Y	G	C	A	C	C	C
	10(b)	A	T	R	T	R	Y	G	C	A	C	C	C
11	11(a)	A	Y	R	Y	A	Y	G	Y	A	C	C	C
12	12(a)	A	T	G	T	G	T	G	C	A	C	C	C
13	13(a)	G	T	A	T	A	C	G	C	A	C	C	C
14	14(a)	A	Y	R	Y	A	Y	G	C	R	Y	Y	Y
	14(b)	A	Y	R	Y	A	Y	G	C	A	Y	Y	Y
15	15(a)	A	C	G	C	A	C	G	C	A	C	C	C
16	16(a)	A	T	A	C	A	T	G	C	A	C	C	C
17	17(a)	A	T	A	C	A	C	G	C	A	C	C	C
	17(b)	A	T	A	C	A	C	G	C	A	C	C	C
18	18(a)	A	T	A	T	A	C	G	C	A	C	C	C
19	19(a)	A	Y	G	T	A	T	R	C	A	C	C	C
	19(b)	A	Y	G	T	A	T	G	C	A	C	C	C
	19(c)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(d)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(e)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(f)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(g)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(h)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(i)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(j)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(k)	A	C	G	T	A	Y	G	C	A	C	C	C
	19(l)	A	C	G	T	A	Y	G	C	A	C	C	C
20	20(a)	A	C	G	T	A	T	G	T	A	C	C	C
21	21(a)	A	T	R	T	A	Y	G	C	A	C	C	C
22	22(a)	A	T	A	C	A	C	G	C	A	C	C	C
23	23(a)	A	Y	G	T	A	T	G	Y	A	Y	Y	Y
24	24(a)	A	Y	R	Y	A	Y	G	C	A	Y	Y	Y
25	25(a)	A	T	R	T	R	Y	G	C	A	C	C	C
26	26(a)	G	Y	R	T	A	Y	G	C	A	Y	Y	Y
27	27(a)	G	T	A	T	A	C	G	C	A	C	C	C
	27(b)	G	T	A	T	A	C	G	C	A	C	C	C
	27(c)	G	T	A	T	A	C	G	C	A	C	C	C
28	28(a)	A	T	A	Y	A	C	G	C	A	C	C	C

		AAT1a basepair position											
Clade	**Unique Sequence**	**87**	**108**	**120**	**150**	**169**	**204**	**244**	**405**	**430**	**432**	**441**	**453**
29	29(a)	G	T	R	T	R	Y	G	C	A	C	C	C
30	30(a)	R	Y	R	T	A	Y	G	C	A	Y	Y	Y
31	31(a)	A	Y	R	T	A	Y	G	C	A	Y	Y	Y
	31(b)	A	C	R	T	A	Y	G	C	A	Y	Y	Y
32	32(a)	A	Y	R	T	A	Y	G	C	A	Y	Y	Y
	32(b)	A	Y	R	T	A	Y	G	C	A	Y	Y	Y
	32(c)	A	Y	R	T	A	Y	G	C	A	Y	Y	Y

		ACC1 basepair position								
Clade	**Unique sequence**	**3258**	**3279**	**3325**	**3333**	**3452**	**3461**	**3531**	**3567**	**3642**
1	1(a)	G	T	A	G	A	M	C	Y	C
	1(b)	G	T	A	G	A	M	C	Y	C
	1(c)	G	T	A	G	A	C	C	T	C
	1(d)	G	T	A	G	A	C	C	T	C
	1(e)	G	T	A	G	A	A	C	C	C
	1(f)	G	T	A	G	A	M	C	Y	C
	1(g)	G	T	A	G	A	M	C	Y	C
	1(h)	G	T	A	G	A	M	C	Y	C
	1(i)	G	T	A	G	A	M	C	Y	C
	1(j)	G	T	A	G	A	A	C	C	C
	1(k)	G	T	A	G	A	M	C	Y	C
	1(l)	G	T	A	G	A	A	C	C	C
	1(m)	G	T	A	G	A	M	C	Y	C
	1(n)	G	T	A	G	A	M	C	Y	C
	1(o)	G	T	A	G	A	M	C	Y	C
	1(p)	G	T	A	G	A	M	C	Y	C
	1(q)	G	T	A	G	A	M	C	Y	C
2	2(a)	G	T	A	G	A	A	C	Y	C
3	3(a)	G	T	A	G	A	A	C	Y	C
	3(b)	G	T	M	G	A	A	C	Y	C
	3(c)	G	T	A	G	A	A	C	Y	C
	3(d)	G	T	A	G	A	A	C	Y	C
	3(e)	G	T	A	G	A	A	C	C	C
4	4(a)	G	T	A	G	A	A	C	C	C
5	5(a)	G	T	A	G	A	A	C	C	C
6	6(a)	G	T	A	G	A	M	C	Y	C
	6(b)	G	T	A	G	A	M	C	Y	C
7	7(a)	G	T	A	G	A	A	C	T	C
	7(b)	G	T	A	G	A	M	Y	Y	C
8	8(a)	G	T	A	G	A	A	C	T	C
	8(b)	G	T	A	G	A	A	Y	Y	C
	8(c)	G	T	A	G	A	A	Y	Y	C
9	9(a)	R	T	A	G	A	A	C	Y	Y
	9(b)	R	T	A	G	A	A	C	Y	Y
10	10(a)	G	T	A	G	A	A	C	C	C
	10(b)	R	T	A	G	A	A	C	C	Y
11	11(a)	R	T	A	G	A	A	C	C	Y
12	12(a)	G	T	A	G	A	A	C	C	C

Clade	Unique sequence	ACC1 basepair position								
		3258	3279	3325	3333	3452	3461	3531	3567	3642
13	13(a)	G	T	A	G	A	A	Y	Y	C
14	14(a)	G	T	A	G	R	A	C	Y	C
	14(b)	G	T	A	G	R	A	C	Y	C
15	15(a)	G	T	A	G	A	M	Y	Y	C
16	16(a)	G	T	A	G	A	A	T	C	C
17	17(a)	G	T	A	G	A	A	Y	Y	C
	17(b)	G	T	A	G	A	A	Y	Y	C
18	18(a)	G	T	A	G	A	A	C	T	C
19	19(a)	G	T	A	G	A	A	T	C	C
	19(b)	G	T	A	G	A	A	T	C	C
	19(c)	G	T	A	G	A	A	C	C	C
	19(d)	G	T	A	G	A	A	Y	C	C
	19(e)	G	T	A	G	A	A	Y	C	C
	19(f)	G	T	A	G	A	A	Y	C	C
	19(g)	G	T	A	G	A	A	C	C	C
	19(h)	G	T	A	G	A	A	C	C	C
	19(i)	G	T	A	G	A	A	T	C	C
	19(j)	G	T	A	G	A	A	Y	C	C
	19(k)	G	T	A	G	A	A	Y	C	C
	19(l)	G	T	A	G	A	A	Y	C	C
20	20(a)	G	T	A	G	A	A	C	T	C
21	21(a)	G	T	A	G	A	A	C	C	C
22	22(a)	G	T	A	G	A	A	C	C	C
23	23(a)	G	T	A	G	A	A	C	C	C
24	24(a)	R	T	A	G	A	A	C	Y	Y
25	25(a)	R	T	A	G	A	A	C	C	Y
26	26(a)	G	T	A	G	A	A	C	C	C
27	27(a)	G	T	A	G	A	A	C	C	C
	27(b)	G	T	A	G	A	A	C	C	C
	27(c)	G	T	A	G	A	A	C	C	C
28	28(a)	R	T	A	G	A	A	C	Y	Y
29	29(a)	R	T	A	R	A	A	Y	C	C
30	30(a)	G	W	A	G	A	A	C	C	C
31	31(a)	A	T	A	G	A	A	C	C	T
	31(b)	R	T	A	G	A	A	Y	C	Y
32	32(a)	A	T	A	G	A	A	C	C	T
	32(b)	A	T	A	G	A	A	C	C	T
	32(c)	A	T	A	G	A	A	C	C	T

		ADP1 basepair position															
Clade	**Unique sequence**	**934**	**952**	**957**	**963**	**1004**	**1026**	**1042**	**1083**	**1122**	**1142**	**1149**	**1199**	**1269**	**1287**	**1312**	**1337**
1	1(a)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(b)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(c)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(d)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(e)	T	G	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(f)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(g)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(h)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(i)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(j)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(k)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(l)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	Y
	1(m)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(n)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(o)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(p)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	1(q)	T	G	C	C	G	A	A	G	G	T	C	G	A	A	A	C
2	2(a)	T	A	C	C	G	A	A	G	G	T	C	G	R	A	A	C
3	3(a)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	3(b)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	3(c)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	3(d)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	3(e)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
4	4(a)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
5	5(a)	Y	A	Y	Y	G	R	R	R	R	W	Y	G	A	A	A	C
6	6(a)	T	A	Y	Y	G	R	R	R	R	W	Y	G	A	A	A	C
	6(b)	T	A	Y	Y	G	R	R	R	R	Y	Y	G	A	A	A	C

		ADP1 basepair position															
Clade	**Unique sequence**	**934**	**952**	**957**	**963**	**1004**	**1026**	**1042**	**1083**	**1122**	**1142**	**1149**	**1199**	**1269**	**1287**	**1312**	**1337**
7	7(a)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	7(b)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
8	8(a)	T	A	C	C	G	A	R	R	R	W	Y	G	A	A	A	C
	8(b)	T	A	C	C	G	A	R	R	R	W	Y	G	A	A	A	C
	8(c)	T	A	C	C	G	A	R	R	R	W	Y	G	A	A	A	C
9	9(a)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
	9(b)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
10	10(a)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	10(b)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
11	11(a)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
12	12(a)	T	R	C	C	G	A	A	G	G	T	C	G	A	A	A	C
13	13(a)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
14	14(a)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	14(b)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
15	15(a)	T	A	C	C	G	A	R	R	R	W	Y	G	A	A	A	C
16	16(a)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
17	17(a)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
	17(b)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
18	18(a)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
19	19(a)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	19(b)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	19(c)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
	19(d)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
	19(e)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
	19(f)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C
	19(g)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
	19(h)	T	A	C	C	G	A	A	G	G	T	C	G	A	A	A	C

Clade	Unique sequence	ADP1 basepair position															
		934	**952**	**957**	**963**	**1004**	**1026**	**1042**	**1083**	**1122**	**1142**	**1149**	**1199**	**1269**	**1287**	**1312**	**1337**
	19(i)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
	19(j)	T	A	C	C	G	A	G	G	G	T	C	G	A	A	A	C
	19(k)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
	19(l)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
20	20(a)	T	A	Y	Y	G	R	G	R	R	W	Y	S	A	A	W	C
21	21(a)	T	A	T	T	A	G	A	G	G	A	T	G	G	G	A	C
22	22(a)	T	A	T	T	G	G	G	A	A	A	T	C	A	A	A	C
23	23(a)	T	A	Y	Y	G	R	R	R	R	W	Y	G	A	A	A	C
24	24(a)	T	A	C	C	G	A	R	G	G	T	C	G	A	A	A	C
25	25(a)	T	A	T	T	G	G	G	A	A	A	T	G	A	A	A	C
26	26(a)	T	A	T	T	G	G	R	A	A	A	T	G	A	A	A	C
27	27(a)	T	A	Y	Y	G	R	G	R	R	W	Y	G	A	A	A	C
	27(b)	T	A	Y	Y	G	R	G	R	R	W	Y	G	A	A	A	C
	27(c)	T	A	T	T	G	G	G	A	A	A	T	G	A	A	A	C
28	28(a)	T	A	Y	Y	G	R	R	R	R	W	Y	G	A	A	A	C
29	29(a)	T	A	Y	Y	G	R	R	R	R	W	Y	S	A	A	W	C
30	30(a)	T	A	Y	Y	G	R	R	R	R	W	Y	G	A	A	A	C
31	31(a)	T	A	Y	Y	G	R	R	R	R	W	Y	G	A	A	A	C
	31(b)	T	A	Y	Y	G	R	R	R	R	W	Y	G	A	A	A	C
32	32(a)	T	A	Y	Y	G	R	R	R	R	W	Y	G	A	A	A	C
	32(b)	T	A	T	T	G	G	G	A	A	A	T	G	A	A	A	C
	32(c)	T	A	T	T	G	G	G	A	A	A	T	G	A	A	A	C

		MPIb basepair position																	
Clade	**Unique sequence**	**474**	**480**	**487**	**489**	**519**	**525**	**541**	**547**	**560**	**591**	**612**	**683**	**687**	**690**	**700**	**729**	**730**	**742**
1	1(a)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(b)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(c)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(d)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(e)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(f)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(g)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(h)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(i)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(j)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(k)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	1(l)	G	A	G	T	T	A	A	G	C	C	A	C	C	A	G	G	G	A
	1(m)	G	A	G	T	T	A	A	G	C	C	A	C	C	A	G	G	G	A
	1(n)	G	A	G	T	T	A	A	G	C	C	A	C	C	A	G	G	G	A
	1(o)	G	A	G	T	T	A	A	G	C	C	A	C	C	A	G	G	G	A
	1(p)	G	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
	1(q)	G	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
2	2(a)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
3	3(a)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	3(b)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	3(c)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	3(d)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	3(e)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
4	4(a)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
5	5(a)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
6	6(a)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A

		MPIb basepair position																	
Clade	**Unique sequence**	**474**	**480**	**487**	**489**	**519**	**525**	**541**	**547**	**560**	**591**	**612**	**683**	**687**	**690**	**700**	**729**	**730**	**742**
	6(b)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
7	7(a)	G	R	R	Y	K	W	R	G	M	C	A	C	C	A	G	G	G	A
	7(b)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
8	8(a)	G	A	G	T	T	A	A	G	C	C	A	C	C	A	G	G	G	A
	8(b)	G	A	G	T	T	A	A	G	C	C	A	C	C	A	G	G	G	A
	8(c)	G	A	G	T	T	A	A	G	C	C	A	C	C	A	G	G	G	A
9	9(a)	G	R	R	Y	G	T	G	G	C	C	A	C	C	A	G	G	G	A
	9(b)	G	R	R	Y	G	T	G	G	C	C	A	C	C	A	G	G	G	A
10	10(a)	G	G	A	C	G	T	G	G	C	C	A	M	C	A	G	G	G	A
	10(b)	G	G	A	C	G	T	G	G	C	C	A	M	C	A	G	G	G	A
11	11(a)	R	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
12	12(a)	R	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
13	13(a)	G	G	A	C	G	T	G	R	M	C	A	C	C	A	G	G	G	A
14	14(a)	G	R	R	Y	K	W	G	G	C	C	A	C	C	A	G	G	G	A
	14(b)	G	A	G	T	T	A	G	G	C	C	A	C	C	A	G	G	G	A
15	15(a)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
16	16(a)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
17	17(a)	R	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	17(b)	R	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
18	18(a)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
19	19(a)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(b)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(c)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(d)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(e)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(f)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A

		MPIb basepair position																	
Clade	**Unique sequence**	**474**	**480**	**487**	**489**	**519**	**525**	**541**	**547**	**560**	**591**	**612**	**683**	**687**	**690**	**700**	**729**	**730**	**742**
	19(g)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(h)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(i)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(j)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(k)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
	19(l)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
20	20(a)	A	G	A	C	G	T	G	G	C	C	A	C	C	A	A	G	G	A
21	21(a)	G	G	A	C	G	T	G	R	M	C	A	C	C	A	G	G	G	A
22	22(a)	G	G	A	C	G	T	G	G	C	C	A	C	A	G	G	A	G	G
23	23(a)	G	R	R	Y	K	W	R	G	C	Y	R	C	M	R	G	R	R	R
24	24(a)	G	R	R	Y	K	W	R	G	C	Y	R	C	M	R	G	R	G	R
25	25(a)	G	R	R	Y	K	W	G	G	C	C	A	C	M	R	G	R	G	R
26	26(a)	G	A	G	T	T	A	R	G	C	C	A	C	C	A	G	G	G	A
27	27(a)	G	A	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	27(b)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
	27(c)	G	R	R	Y	K	W	R	G	C	C	A	C	C	A	G	G	G	A
28	28(a)	G	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
29	29(a)	R	R	R	Y	G	T	G	G	C	C	A	C	C	A	G	G	G	A
30	30(a)	R	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
31	31(a)	A	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
	31(b)	A	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
32	32(a)	R	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
	32(b)	R	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A
	32(c)	R	G	A	C	G	T	G	G	C	C	A	C	C	A	G	G	G	A

		SYA1 basepair position									
Clade	**Unique sequence**	**2400**	**2424**	**2460**	**2499**	**2541**	**2559**	**2584**	**2675**	**2706**	**2750**
1	1(a)	T	A	G	C	A	T	G	C	C	C
	1(b)	T	A	G	C	A	T	G	C	C	C
	1(c)	T	A	G	C	A	T	G	C	C	C
	1(d)	T	A	G	C	A	T	G	C	C	C
	1(e)	T	A	G	C	A	T	G	C	C	C
	1(f)	T	A	G	C	A	T	G	C	C	C
	1(g)	T	A	G	C	A	T	G	C	C	C
	1(h)	T	A	G	C	A	T	G	C	C	C
	1(i)	T	A	G	C	A	T	G	C	C	C
	1(j)	T	A	G	C	A	T	G	C	C	C
	1(k)	T	A	G	C	A	T	G	C	C	C
	1(l)	T	A	G	C	A	T	G	C	C	C
	1(m)	T	A	G	C	A	T	G	C	C	C
	1(n)	T	A	G	C	A	T	G	C	C	C
	1(o)	T	A	G	C	A	T	G	C	C	C
	1(p)	T	A	G	C	A	T	G	C	C	C
	1(q)	T	A	G	C	A	T	G	C	C	C
2	2(a)	T	A	G	C	A	T	G	C	C	C
3	3(a)	T	A	R	Y	R	Y	G	C	M	C
	3(b)	T	A	R	Y	R	Y	G	C	M	C
	3(c)	T	A	R	Y	R	Y	G	C	M	C
	3(d)	T	A	R	Y	R	Y	G	C	M	C
	3(e)	T	A	R	Y	R	Y	G	C	M	C
4	4(a)	Y	M	R	C	A	Y	K	C	C	C
5	5(a)	Y	M	R	T	G	C	G	C	M	C
6	6(a)	T	M	R	C	A	T	G	C	C	C
	6(b)	T	M	R	C	A	T	G	C	C	C
7	7(a)	T	A	G	C	A	T	G	C	C	C
	7(b)	T	A	G	C	A	T	G	C	C	C
8	8(a)	T	A	R	C	A	T	G	C	C	C
	8(b)	T	A	R	C	A	T	G	C	C	C
	8(c)	T	A	R	C	A	T	G	C	C	C
9	9(a)	T	M	A	Y	R	Y	G	C	C	Y
	9(b)	T	M	A	Y	R	Y	G	C	C	Y
10	10(a)	T	M	R	Y	R	Y	G	C	C	Y
	10(b)	T	M	R	Y	R	Y	G	C	C	Y
11	11(a)	T	M	A	T	G	C	G	C	M	Y
12	12(a)	T	C	A	T	G	C	G	C	C	T

		SYA1 basepair position									
Clade	**Unique sequence**	**2400**	**2424**	**2460**	**2499**	**2541**	**2559**	**2584**	**2675**	**2706**	**2750**
13	13(a)	T	A	A	Y	R	Y	G	Y	C	Y
14	14(a)	T	C	A	T	G	C	G	C	C	T
	14(b)	T	M	R	Y	R	Y	G	C	C	Y
15	15(a)	T	M	A	T	G	C	G	C	M	Y
16	16(a)	Y	M	R	T	G	C	G	C	M	Y
17	17(a)	Y	C	R	T	G	C	G	C	C	T
	17(b)	Y	C	A	T	G	C	G	C	C	T
18	18(a)	Y	C	R	T	G	C	G	C	M	Y
19	19(a)	Y	C	R	T	G	C	G	C	C	T
	19(b)	Y	C	R	T	G	C	G	C	C	T
	19(c)	Y	C	R	T	G	C	G	C	C	T
	19(d)	Y	C	R	T	G	C	G	C	C	T
	19(e)	C	C	G	T	G	C	G	C	C	T
	19(f)	Y	C	R	T	G	C	G	C	C	T
	19(g)	Y	C	R	T	G	C	G	C	C	T
	19(h)	Y	C	R	T	G	C	G	C	C	T
	19(i)	Y	C	R	T	G	C	G	C	C	T
	19(j)	Y	C	R	T	G	C	G	C	C	T
	19(k)	Y	C	A	T	G	C	G	C	C	T
	19(l)	Y	C	R	T	G	C	G	C	C	T
20	20(a)	C	C	G	T	G	C	K	C	C	Y
21	21(a)	Y	M	R	Y	R	Y	G	C	C	Y
22	22(a)	Y	C	A	Y	R	C	K	C	C	Y
23	23(a)	Y	C	A	Y	R	C	K	C	C	Y
24	24(a)	Y	C	A	Y	R	C	K	C	C	Y
25	25(a)	T	M	R	Y	R	Y	G	C	C	Y
26	26(a)	Y	C	R	Y	R	Y	G	C	C	C
27	27(a)	T	C	A	T	G	C	G	C	C	T
	27(b)	T	C	A	T	G	C	G	C	C	T
	27(c)	T	C	A	T	G	C	G	C	C	T
28	28(a)	T	C	A	T	G	C	G	C	C	T
29	29(a)	T	C	A	T	G	C	G	C	M	Y
30	30(a)	Y	C	R	T	G	C	G	C	C	T
31	31(a)	T	C	A	T	G	C	G	C	C	T
	31(b)	T	C	A	T	G	C	G	C	C	T
32	32(a)	T	M	R	Y	R	Y	G	C	C	Y
	32(b)	T	M	R	Y	R	Y	G	C	C	Y
	32(c)	T	M	R	Y	R	Y	G	C	C	Y

		VPS13 basepair position															
Clade	**Unique sequence**	**5183**	**5199**	**5284**	**5318**	**5320**	**5362**	**5367**	**5391**	**5431**	**5432**	**5470**	**5472**	**5476**	**5478**	**5484**	**548**
1	1(a)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(b)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(c)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(d)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(e)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(f)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(g)	G	C	A	G	C	G	C	A	R	G	G	K	G	G	G	G
	1(h)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	1(i)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(j)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(k)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(l)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(m)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(n)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(o)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	1(p)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	1(q)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
2	2(a)	G	C	A	G	C	G	Y	A	R	G	G	T	G	G	G	G
3	3(a)	G	M	A	G	C	G	Y	A	G	G	R	T	G	G	G	G
	3(b)	G	M	A	G	C	G	Y	A	G	G	R	T	G	G	G	G
	3(c)	G	M	A	G	C	G	T	A	G	G	R	T	G	G	G	G
	3(d)	G	M	A	G	C	G	Y	A	G	G	R	T	G	G	G	G
	3(e)	G	C	A	G	C	G	C	A	G	G	G	T	G	G	G	G
4	4(a)	G	C	A	G	C	G	Y	A	R	G	G	T	G	G	G	G
5	5(a)	G	M	A	G	C	G	T	A	G	G	A	T	G	G	G	G
6	6(a)	G	C	A	G	C	G	Y	A	G	G	G	T	G	G	G	S
	6(b)	G	C	A	G	C	G	Y	A	G	G	G	T	G	G	G	S

		VPS13 basepair position															
Clade	**Unique sequence**	**5183**	**5199**	**5284**	**5318**	**5320**	**5362**	**5367**	**5391**	**5431**	**5432**	**5470**	**5472**	**5476**	**5478**	**5484**	**548**
7	7(a)	G	C	A	G	C	G	T	G	G	R	G	T	G	G	G	G
	7(b)	G	M	A	G	C	G	T	A	G	G	R	T	G	G	G	S
8	8(a)	G	C	A	G	C	G	T	G	G	R	G	T	G	G	G	G
	8(b)	G	C	A	G	C	G	T	G	G	G	G	T	G	G	G	G
	8(c)	G	C	A	G	C	G	T	G	G	R	G	T	G	G	G	G
9	9(a)	G	C	A	G	C	G	Y	A	R	G	G	T	R	G	G	G
	9(b)	G	C	A	G	C	G	Y	A	R	G	G	T	R	G	G	G
10	10(a)	G	M	A	G	Y	G	T	A	G	G	G	T	G	G	G	G
	10(b)	G	M	A	G	Y	G	T	A	G	G	G	T	G	G	G	G
11	11(a)	G	M	A	G	Y	G	T	R	G	R	G	T	G	G	G	G
12	12(a)	G	C	A	R	C	G	T	G	G	R	G	T	G	G	G	G
13	13(a)	G	C	T	G	C	G	T	A	A	G	G	T	G	G	G	G
14	14(a)	R	M	A	G	Y	G	T	A	G	G	G	T	G	R	G	G
	14(b)	R	M	A	G	Y	G	T	A	G	G	G	T	G	R	G	G
15	15(a)	G	C	A	G	C	G	T	A	A	G	G	T	A	G	G	G
16	16(a)	G	C	A	G	C	G	T	A	A	G	G	K	R	G	G	G
17	17(a)	G	C	A	G	C	G	T	R	R	G	G	K	G	G	G	G
	17(b)	G	C	A	G	C	G	T	R	R	G	G	K	G	G	G	G
18	18(a)	G	C	A	R	C	G	T	R	G	G	G	T	G	G	G	S
19	19(a)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	19(b)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
	19(c)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	19(d)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	19(e)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	19(f)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	19(g)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	19(h)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G

		VPS13 basepair position															
Clade	**Unique sequence**	**5183**	**5199**	**5284**	**5318**	**5320**	**5362**	**5367**	**5391**	**5431**	**5432**	**5470**	**5472**	**5476**	**5478**	**5484**	**548**
	19(i)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	19(j)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	19(k)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
	19(l)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
20	20(a)	G	A	A	G	C	G	T	A	G	G	G	T	G	G	G	G
21	21(a)	G	C	A	G	C	G	T	A	A	G	G	G	G	G	G	G
22	22(a)	G	C	A	G	C	G	T	A	A	G	G	T	A	G	G	G
23	23(a)	G	C	A	G	C	G	Y	A	R	G	G	T	G	G	G	G
24	24(a)	G	C	A	G	C	G	T	R	R	R	G	K	G	G	G	G
25	25(a)	G	C	W	G	C	G	T	R	R	A	G	T	G	G	G	G
26	26(a)	G	M	A	G	C	G	T	A	R	G	R	T	G	G	G	G
27	27(a)	G	C	A	G	C	G	T	A	G	G	G	T	G	G	G	G
	27(b)	G	C	A	G	C	G	Y	A	G	G	G	T	G	G	G	G
	27(c)	G	C	A	G	C	G	Y	A	G	G	G	T	G	G	G	G
28	28(a)	G	C	A	G	C	G	Y	A	R	G	G	K	G	G	G	G
29	29(a)	G	C	A	G	C	G	Y	R	G	R	G	T	G	G	G	G
30	30(a)	G	C	A	G	C	G	Y	R	G	G	G	T	G	G	G	G
31	31(a)	G	C	A	G	C	R	T	G	G	R	G	T	G	G	R	G
	31(b)	G	C	A	G	C	A	T	G	G	G	G	T	G	G	A	G
32	32(a)	G	C	A	G	C	R	T	G	G	R	G	T	G	G	G	G
	32(b)	G	C	A	G	C	R	T	G	G	R	G	T	G	G	R	G
	32(c)	G	C	A	G	C	G	T	G	G	A	G	T	G	G	G	G

		ZWF1b basepair position														
Clade	**Unique sequence**	**908**	**925**	**933**	**945**	**951**	**957**	**1060**	**1077**	**1089**	**1164**	**1176**	**1239**	**1281**	**1341**	**1384**
1	1(a)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(b)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(c)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(d)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(e)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(f)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(g)	C	A	T	T	T	T	A	C	G	C	A	T	Y	A	W
	1(h)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(i)	C	A	Y	W	T	T	A	C	G	C	A	T	T	A	A
	1(j)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(k)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(l)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(m)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(n)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	A
	1(o)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(p)	C	A	C	W	T	T	A	C	G	C	A	Y	Y	A	W
	1(q)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
2	2(a)	M	C	C	T	C	C	A	C	G	C	A	C	T	A	A
3	3(a)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
	3(b)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
	3(c)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
	3(d)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
	3(e)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
4	4(a)	C	A	Y	W	T	Y	A	C	G	C	R	T	T	A	A
5	5(a)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
6	6(a)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	6(b)	C	M	T	T	T	T	A	C	G	C	A	Y	T	A	A

		ZWF1b basepair position														
Clade	**Unique sequence**	**908**	**925**	**933**	**945**	**951**	**957**	**1060**	**1077**	**1089**	**1164**	**1176**	**1239**	**1281**	**1341**	**1384**
7	7(a)	C	C	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	7(b)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
8	8(a)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	8(b)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	8(c)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
9	9(a)	C	A	Y	T	Y	Y	A	C	G	C	A	T	T	A	A
	9(b)	C	A	Y	T	Y	Y	A	C	G	C	A	T	T	A	A
10	10(a)	C	M	Y	T	Y	Y	R	C	G	C	A	Y	Y	A	W
	10(b)	C	M	Y	T	Y	Y	R	C	G	C	A	Y	Y	A	W
11	11(a)	C	C	C	T	C	C	A	C	G	C	A	C	T	A	A
12	12(a)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
13	13(a)	C	C	C	T	C	C	A	C	G	C	A	C	T	A	A
14	14(a)	C	A	C	A	Y	Y	A	C	G	C	R	C	Y	A	W
	14(b)	C	A	C	A	Y	Y	A	C	G	C	R	C	Y	A	W
15	15(a)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
16	16(a)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
17	17(a)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	17(b)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
18	18(a)	C	A	Y	W	Y	Y	A	Y	G	C	A	Y	T	A	A
19	19(a)	C	M	C	W	Y	C	A	C	G	C	R	Y	T	A	A
	19(b)	C	M	C	W	Y	C	A	C	G	C	R	Y	T	A	A
	19(c)	C	C	C	T	C	C	A	C	T	C	A	C	T	A	A
	19(d)	C	A	T	T	T	T	A	C	G	C	A	T	T	A	A
	19(e)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	19(f)	C	M	Y	T	Y	Y	A	C	G	C	A	C	T	A	A
	19(g)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	19(h)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A

		ZWF1b basepair position														
Clade	**Unique sequence**	**908**	**925**	**933**	**945**	**951**	**957**	**1060**	**1077**	**1089**	**1164**	**1176**	**1239**	**1281**	**1341**	**1384**
	19(i)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	19(j)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	19(k)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	19(l)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
20	20(a)	C	A	C	W	C	C	A	C	G	C	R	Y	T	A	A
21	21(a)	C	A	Y	W	Y	Y	A	C	G	M	A	Y	T	R	A
22	22(a)	C	M	C	W	Y	C	A	C	G	C	R	Y	T	A	A
23	23(a)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
24	24(a)	C	A	C	A	C	C	A	C	G	C	R	C	T	A	A
25	25(a)	C	A	Y	W	T	T	A	C	G	C	A	Y	Y	A	W
26	26(a)	C	A	Y	W	Y	Y	A	Y	G	M	R	Y	T	A	A
27	27(a)	C	C	C	T	C	C	A	C	G	C	A	C	T	A	A
	27(b)	C	C	C	T	C	C	A	C	G	C	A	Y	T	A	A
	27(c)	C	C	C	T	C	C	A	C	G	C	A	Y	T	A	A
28	28(a)	C	M	C	W	C	C	A	C	G	C	A	C	T	A	A
29	29(a)	C	M	C	W	Y	Y	A	Y	G	C	R	T	T	A	A
30	30(a)	C	M	C	W	Y	Y	A	C	G	C	A	C	Y	A	W
31	31(a)	C	M	Y	T	Y	Y	A	C	G	C	A	T	T	A	A
	31(b)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
32	32(a)	C	A	T	T	T	Y	A	C	G	C	A	Y	T	A	A
	32(b)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A
	32(c)	C	M	Y	T	Y	Y	A	C	G	C	A	Y	T	A	A