Marco Alexandre Figueira Rodrigues Guerreiro

Licenciatura em Biotecnologia

Unveiling the mating system and genetic variability in the yeast *Kwoniella mangroviensis* using molecular approaches

Dissertação para obtenção do Grau de Mestre em
Genética Molecular e Biomedicina

Orientador: Prof. Doutor Álvaro Luís A. M. R. Fonseca,
Professor Auxiliar, FCT/UNL

Júri:

Presidente: Prof. Doutor José Paulo Nunes de Sousa Sampaio
Arguente: Prof. Doutor Artur Jorge da Costa Peixoto Alves

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro de 2012

Marco Alexandre Figueira Rodrigues Guerreiro

Licenciatura em Biotecnologia

Unveiling the mating system and genetic variability in the yeast *Kwoniella mangroviensis* using molecular approaches

Dissertação para obtenção do Grau de Mestre em
Genética Molecular e Biomedicina

Orientador: Prof. Doutor Álvaro Luís A. M. R. Fonseca,
Professor Auxiliar, FCT/UNL

Júri:

Presidente: Prof. Doutor José Paulo Nunes de Sousa Sampaio
Arguente: Prof. Doutor Artur Jorge da Costa Peixoto Alves

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro de 2012

Unveiling the mating system and genetic variability in the yeast Kwoniella mangroviensis using molecular approaches

Part of the results discussed in this thesis was presented in:

Marco A. Guerreiro, Joana A. Rodrigues, Deborah Springer, Keisha Findley, Joseph Heitman, Álvaro Fonseca. Molecular evidence for a tetrapolar mating system and genetic heterogeneity in the basidiomycetous yeast *Kwoniella mangroviensis.* XIX Jornadas de Biologia de Leveduras "Professor Nicolau van Uden". FCT/UNL, Caparica, Portugal, June 15th and 16th 2012. [Oral communication].

ACKNOWLEDGMENTS

I wish to express my deep gratitude to my supervisor, Prof. Álvaro Fonseca, for accepting me as a master student; the opportunity to work with him; the knowledge shared over these months; for the guidance, the wise advice, the support, the trust, the patience, the motivation and for everything else.

I would also like to thank: Prof. João Almeida (CREM) for support in bioinformatics issues and advice during this work; Dr. Deborah Springer (Duke University) for providing strains crucial for this work and some DNA sequences; and CREM for providing the technical support and conditions required for the practical work.

I would like to thank Cláudia Carvalho (PYCC/CREM) for the friendship and companionship, all the technical support, all the conversations, advice, laughter, kindness and for helping me in every single difficulty that I had.

A special thanks is owed to Dr. Andrey Yurkov (CREM) for the friendship, for all the advice, tips and suggestions, support in bioinformatics analyses, the endless discussions, the knowledge and all the stories shared during coffee breaks.

I would like to thank my lab colleagues at CREM for technical advice and for providing a healthy and joyful environment to work in.

I must also mention my family for all the support over the years, specially my mother for all the education and values transmitted. My friends who always supported me, Neuza Sousa, the drama solver and for making everything more bearable, Pedro Penedo, for everything done and also Miguel Lopes and Hélio Fazenda for making life a little bit more funnier and glamorous.

Finally, I would like to acknowledge Fundação para a Ciência e Tecnologia for funding of a BI grant under the project PTDC/BIA-MIC/113051/2009, during the last 6 months of my work.

ABSTRACT

In fungi, sexual reproduction is orchestrated by genomic regions known as mating type loci (*MAT*), which can be defined by two physically unlinked sex determining regions (tetrapolar mating system) or a single locus (bipolar system). *Kwoniella mangroviensis* is a saprobic basidiomycetous yeast, belonging to the Kwoniella clade of the order Tremellales, which was described as possessing a bipolar mating-system, similar to the related pathogenic species *Cryptococcus neoformans* and *Cryptococcus gattii* of the sister Filobasidiella clade. Studies aimed at elucidating the evolution of the *MAT* locus of these two *Cryptococcus* species of clinical importance, targeted several related saprobic species in the Kwoniella and Filobasidiella clades. An evolutionary model ensuing from those studies suggests that the tetrapolar mating-systems found in most species are ancestral and gave rise to the bipolar mating-system of *C. neoformans* by chromosomal rearrangements and fusion events.

The present study comprised strains from the original work describing *K. mangroviensis*, as well as additional isolates from plant substrates in Europe and Africa. A multilocus sequence typing approach revealed genetic variability among those strains and led to the identification of two novel species closely related to *K. mangroviensis*: *Kwoniella* sp. A and *Kwoniella* sp. B. The mating system of *K. mangroviensis* and sibling species was further explored by a genetic approach based on sequencing of two *MAT* genes: *STE20* and the divergently transcribed genes *SXI1* and *SXI2*. The results obtained demonstrated tetrapolar mating systems in *K. mangroviensis*, as well as in *Kwoniella* sp. A and *Kwoniella* sp. B. Additionally, the *MAT* locus structure of *K. mangroviensis* was unveiled by sequencing a 43 kb genomic region containing the *STE20* gene. Twelve genes also present in the *MAT* loci of related species were identified, and full synteny was found between *K. mangroviensis* and *Cryptococcus heveanensis*, a distant member of the Kwoniella clade. These findings provided novel insights into the evolution of *MAT* loci in basidiomycetous yeasts in the Tremellales.

Keywords: Fungi, basidiomycetes, *Cryptococcus*, *MAT* locus, sexual reproduction, MLST

RESUMO

A reprodução sexuada em fungos é regulada por regiões genómicas conhecidas como *loci MAT*, que podem ser definidos por duas regiões fisicamente separadas (sistema de compatibilidade tetrapolar) ou um único *locus* (sistema bipolar). *Kwoniella mangroviensis* é uma levedura basidiomiceta saprófita, pertencente ao clado Kwoniella da ordem Tremellales, descrita originalmente como possuindo um sistema bipolar, similar ao das espécies patogénicas afins *Cryptococcus neoformans* e *Cryptococcus gattii* do clado Filobasidiella. Devido à relevância clínica destas duas espécies de *Cryptococcus*, foram realizados estudos para compreender o processo evolutivo das suas regiões *MAT* em espécies saprófitas afins. Foram encontrados sistemas tetrapolares em todas as espécies estudadas o que sugere que estes sistemas são ancestrais e terão dado origem ao sistema bipolar de *C. neoformans* através de rearranjos cromossómicos e fusões.

O estudo descrito nesta tese compreendeu estirpes do trabalho original que descreveu *K. mangroviensis* bem como novos isolados de substratos vegetais na Europa e África. Uma abordagem de sequenciação *multilocus* demonstrou variabilidade genética entre as estirpes estudadas e levou à identificação de duas novas espécies próximas de *K. mangroviensis*: *Kwoniella* sp. A e *Kwoniella* sp. B. O sistema de compatibilidade de *K. mangroviensis* foi estudado com base na sequencição de dois genes *MAT*: *STE20* e o par *SXI1* e *SXI2*. Os resultados obtidos demonstraram a existência de sistemas tetrapolares em *K. mangroviensis* bem como em *Kwoniella* sp. A e *Kwoniella* sp. B. A estrutura do *locus MAT* de *K. mangroviensis* foi parcialmente revelada por sequenciação dum fragmento genómico de 43 kb contendo o gene *STE20*. Foram identificados 12 genes também presentes nos *loci MAT* de espécies afins e foi encontrada uma sintenia quase completa entre *K. mangroviensis* e *Cryptococcus heveanensis*, membro mais distante do clado Kwoniella. Estes resultados forneceram novas pistas valiosas sobre a evolução dos *loci MAT* em leveduras basidiomicetes da ordem Tremellales.

Palavras-chave: Fungos, basidiomicetes, *Cryptococcus*, locus *MAT*, reprodução sexuada, MLST

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

(T) – Type-strain

***B.* –** *Bullera*

bp – Base pairs

***C.* –** *Cryptococcus*

CDS – Coding Sequence

***E. coli* –** *Escherichia coli*

***F.* –** *Filobasidiella*

h – Hour(s)

HD – Homeodomain

ITS (Internal Transcriber Spacer) **–** ITS1 + 5.8S rRNA + ITS2

JGI – Joint Genome Initiative

***K.* –** *Kwoniella*

kb – Kilobase

LSU – Large Subunit of ribosomal RNA

MAPK – Mitogen-Activated Protein Kinase

***MAT* –** Mating-type

min – Minute(s)

ML – Maximum Likelihood

MLST – Multilocus Sequence Typing

NJ – Neighbor Joining

P/R – Pheromone and pheromone receptor

PCR – Polymerase Chain Reaction

rpm – Revolutions per minute

***SXI* –** *SXI1*, *SXI2* and intergenic region

***T. mesenterica* –** *Tremella mesenterica*

***T. wingfieldii* –** *Tsuchiyaea wingfieldii*

var. – Variety

w/v – weight/volume

1 - INTRODUCTION

1.1 - Fungi and Basidiomycetes

The kingdom Fungi is a diverse group of eukaryotic microorganisms that includes mushrooms, molds and yeasts. The most recent classification of this kingdom (Hibbett *et al*., 2007) includes one subkingdom (Dikarya), seven phyla (Chytridiomycota, Neocallimastigomycota, Blastocladiomycota, Microsporidia, Glomeromycota, Ascomycota, Basidiomycota), ten subphyla, 35 classes, 12 subclasses and 129 orders. According to the Global Catalogue of Microorganisms, more than 77000 species of fungi have been described and some predictions suggest that 1.5 million species may exist (Hawksworth, 1991). The majority of known fungi are included in the subkingdom Dikarya, comprising the two sister phyla Ascomycota and Basidiomycota, which share a monophyletic origin (Hibbett *et al*., 2007). Yeasts are found in the latter two phyla and are mostly saprophytic but include some important human pathogenic fungi, namely *Candida* spp. (ascomycetes), *Cryptococcus neoformans* and *Cryptococcus gattii* (basidiomycetes) (Cooper, 2011). Yeasts are generally characterized by budding or fission as the primary mode of vegetative reproduction and have sexual states that are not enclosed in fruiting bodies (Kurtzman *et al*., 2011a).

The phylum Basidiomycota comprises economically, environmentally and medically important fungi that produce meiotic spores on the surface of often flask-shaped cells, termed basidia. This phylum is divided into three subphyla: the Pucciniomycotina (including rust fungi, anther smuts, some jelly-like fungi and many basidiomycetous yeasts), the Ustilaginomycotina (including the true smut fungi and some yeasts), and the Agaricomycotina (including the mushroom-forming species, jelly fungi and many yeasts in the class Tremellomycetes) (Hibbett *et al*., 2007; Kües *et al*., 2011).

1.2 - Looking for sexual reproduction in Fungi

In nature, there are many organisms that can reproduce asexually and/or sexually. Fungi can reproduce both ways, even though, some fungal species are only known by their asexual form.

There are two main approaches used to look for evidence of sex or sexual recombination in fungi. The direct approach is based on the direct observation of biological mating ability and this is the ultimate biological proof of fertility; however, appropriate mating conditions and mating partners must be identified in order to get successful crosses (Campbell and Carter, 2006). Non-optimal mating conditions may result in an unsuccessful crossing and therefore, be mistaken for lack of fertility. To overcome these problems, indirect approaches are used to find evidence of sex, such as DNA based analyses based on molecular population genetics or the search for functional sex-determining genomic regions (Campbell and Carter, 2006).

Molecular approaches circumvent the need of testing strain compatibility under laboratory conditions and also allow to determine the sexual recombination versus clonal propagation ratio in a population (Campbell and Carter, 2006). Populations are defined as a group of organisms belonging to a single species, present in one place at one time. Sexual reproduction enables genetic exchange; therefore, the choice of the population is decisive for the recombination analysis. If the isolates are sampled too widely, they may belong to genetically isolated subpopulations and therefore, no sexual recombination will be found. If the isolates are sampled too narrowly, there might be an oversampling of clones. In both situations, clonality may be concluded by mistake due to inappropriate sampling. A molecular population genetic analysis requires the development of molecular markers and the analysis of their distribution among the population. For recombination analysis, the most important factor is that all loci arrange independently. This requirement excludes techniques such as DNA fingerprinting to identify sexual recombination, although this technique is useful to identify populations, discriminate and identify species and strains (Xu, 2006).

Multilocus sequence typing (MLST) is a method used for generating independent loci for recombination analyses (Campbell and Carter, 2006). This method, based on sequencing of independent and polymorphic loci, uses nucleotide sequences from several housekeeping genes to characterize genetic diversity. For phylogenetic analysis, the sequenced genes are chosen based on their degree of conservation within a species and among species. The ribosomal RNA (rRNA) gene cluster is the most common sequenced region for species identification (Xu, 2006). This region includes the internal transcribed spacers (ITS) 1 and 2, the intergenic spacer (IGS), 5.8S, 18S (SSU) and 26S (LSU) rRNA genes and it is chosen due to its multi-copy and high conservation within a species but high polymorphism among species along the spacers. Other genes besides the rRNA gene cluster are also used, such as mitochondrial ATPase subunits, beta-tubulin and elongation factor (Xu, 2006). For species to be identified and distinguished, single copy protein-coding genes, such as RNA polymerases (*RPB1* and *RPB2*) (James *et al*., 2006) or DNA replication licensing factor (*MCM7*) (Schmitt *et al*., 2009) are useful, while for strains within a species, non-functional DNA regions are more informative than conserved genes due to evolutionary and selective pressure, whereas non-coding sequences are more variable (Xu, 2006). MLST has the advantage that it is highly reproducible, allowing data from different experiments and studies to be compiled together, continuously being improved and complemented (Campbell and Carter, 2006).

To assess the potential for sexual reproduction, the previous molecular approaches can be complemented with the search for functional sex-determining genomic regions such as mating type loci and the genes therein that are known to regulate and be involved in mating in fungi (Casselton and Challen, 2006). The existence of these regions suggests an ability to mate, even in the absence of an extant sexual cycle (Ni *et al*., 2011), and also provides useful information to identify potential compatible mating partners.

1.3 - Basidiomycete life-cycle and Heterothallism vs. Homothallism

Basidiomycetes have a complex life-cycle, and undergo haploid, diploid and dikaryotic developmental pathways. Many fungi in the Basidiomycota have a dimorphic life-cycle, where monokaryotic yeasts alternate with dikaryotic filaments. Species that switch between single cell-yeasts and a filamentous growth form (most sexual basidiomycetous yeasts) are called "dimorphic", in contrast with "monomorphic" species (such as mushrooms and the rusts) that only have one known growth form (Morrow and Fraser, 2009). Those dimorphic fungi can exist as yeasts that divide by budding, or as filaments that are involved in sexual or monokaryotic fruiting (Figure 1.1). Sexual reproduction is common among fungi and it can happen, usually under nutrient-limited conditions, in two different sexual systems: heterothallism (self-sterility) or homothallism (self-fertility).

Heterothallism requires cell-cell fusion between two morphologically identical cells with different and compatible sexual identities, while in homothallism cell-cell fusion occurs between cells of a single individual in solo culture (Kües *et al*., 2011; Ni *et al*., 2011). Some statistics suggest that approximately 90% of basidiomycetes are heterothallic and 10% are homothallic (Whitehouse, 1949), but sometimes, under some specific environmental conditions, some heterothallic fungi can undergo same-sex mating (for example, *Cryptococcus neoformans*) in the absence of a compatible mating partner (Ni *et al*., 2011).

In a typical heterothallic sexual-cycle of a dimorphic basidiomycete, two self-sterile haploid cells undergo chemoattraction, mediated by pheromones, and fuse (Figure 1.1.A). After cell-cell fusion, nuclear fusion is delayed and the dikaryon grows as hyphae, in most cases, with fused clamp connections. Nuclear migration is mediated also by pheromone signaling and regulated by a process involving clamp cell formation, fusion and the formation of a dikaryon with two different haploid nuclei. A single specialized cell, the basidium, is formed at the termini of the dikaryotic hyphae, where nuclear fusion (karyogamy) and meiosis occur to produce external haploid basidiospores (Kües *et al*., 2011). The reproduction in homothallic basidiomycetes may include uniparental reproduction without a partner (apomixis) or biparental reproduction (amphimixis). In this case mating can involve the fusion between two self-fertile strains (outcrossing) or the fusion of two cells/nuclei of the same strain (inbreeding) (Kües *et al*., 2011).

Under nutrient-limiting conditions and in the absence of a compatible partner, some dimorphic heterothallic species can undergo monokaryotic fruiting (Figure 1.1.B). Haploid cells with the same sexual identity become diploid by endoduplication or by nuclear fusion, following cell fusion. The diploid monokaryotic hyphae form clamp connections that are not fused. When the basidium is formed, meiosis occurs and haploid basidiospores are produced (Lin and Heitman, 2006).

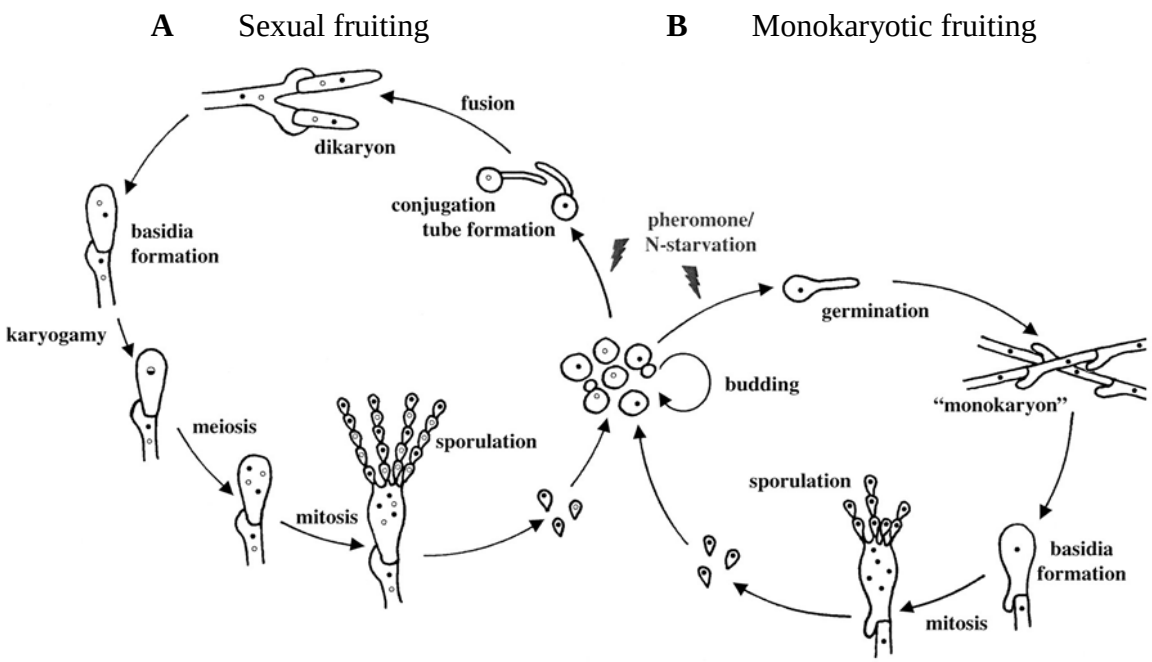


Figure 1.1 - Life cycle of the basidiomycetous yeast *Cryptococcus neoformans* (adapted from Lengeler *et al*., 2000). Heterothallic sexual-cycle (**A**), two self-sterile haploid cells undergo chemoattraction, mediated by pheromones, and fuse. After cell-cell fusion, nuclear fusion is delayed and the dikaryon grows as hyphae with fused clamp connections. The basidium is formed at the termini of the dikaryotic hyphae, where nuclear fusion and meiosis occur to produce haploid basidiospores. Different sexual identities are represented by different nucleus colors (black and white). Monokaryotic fruiting (**B**). Haploid cells with the same sexual identity become diploid and cells with the same sexual identity fuse. The diploid monokaryotic hyphae form clamp connections that are not fused. The basidium is formed, meiosis occurs and haploid basidiospores are produced.

1.4 - Bipolar and Tetrapolar Mating Systems

Sexual identity is always established in the haploid phase and is determined by specialized genomic regions known as mating type loci (*MAT*), responsible for cell identity, cell-cell fusion and/or the viability of the zygote. In spite of the importance of this genomic region to the fungal cell cycle, species identity and fitness, it is extremely plastic and the genes within *MAT* differ considerably between fungal phyla and among species within phyla (Hsueh *et al*., 2011). In general, these regions can be arranged in two different sexual compatibility systems: tetrapolar and bipolar.

In the tetrapolar mating system, exclusive to the basidiomycetes (for example, *Tremella mesenterica* and *Ustilago maydis*), the mating type is defined by two physically unlinked sex determining loci. One locus contains genes encoding pheromones and pheromone receptors (P/R or A locus), and the other locus genes encoding homeodomain transcription factors (HD or B locus). The P/R locus is commonly biallelic (A1, A2), but there are some exceptions, while the HD locus is usually multiallelic (B1, B2, …, Bn) (Kües *et al*., 2011). Each allele combination (for example, A1B1,

A1B3, A2B1) is a different mating-type, and in order to mate, individuals must have compatible mating-types, i.e. different alleles at both loci (Casselton and Challen, 2006). The shared regulation for these loci is to control the mating process by self and nonself recognition during early (cell fusion) and late (nuclear fusion) stages of sexual reproduction (Hsueh *et al*., 2011).

In the bipolar mating system, which is found in ascomycetes and basidiomycetes (for example, *Saccharomyces cerevisiae* and *Cryptococcus neoformans*, respectively), a single locus, most of the times biallelic (designated as, for example, **a** or α), contains genes encoding only transcription factors in ascomycetes or pheromone, pheromone receptors and homeodomain transcription factors in basidiomycetes (Kües *et al*., 2011).

MAT loci vary considerably in size and gene content between mating systems and species. In the basidiomycetous bipolar yeast *Cryptococcus neoformans*, this locus spans more than 100 kb and comprises more than 20 genes (Lengeler *et al*., 2002), while in the ascomycetous bipolar yeast *Saccharomyces cerevisiae* it spans only 642 bp (*MAT***a**) or 747 bp (*MAT*α) and comprises only one or two genes (Figure 1.2) (Fraser *et al*., 2004). In the tetrapolar dimorphic basidiomycete *Tremella mesenterica*, the P/R locus spans about 18 kb, encoding six genes, and the HD locus spans about 3 kb, encoding two genes (*SXI1* and *SXI2*) (Findley, 2010), while in the tetrapolar dimorphic basidiomycete *Ustilago maydis*, the P/R locus spans 4.6 kb or 8.4 kb, encoding two or four genes, and the HD locus spans about 3.6 kb, encoding 2 genes (*bW* and *bE*) as well (Figure 1.2) (Fraser *et al*., 2004).

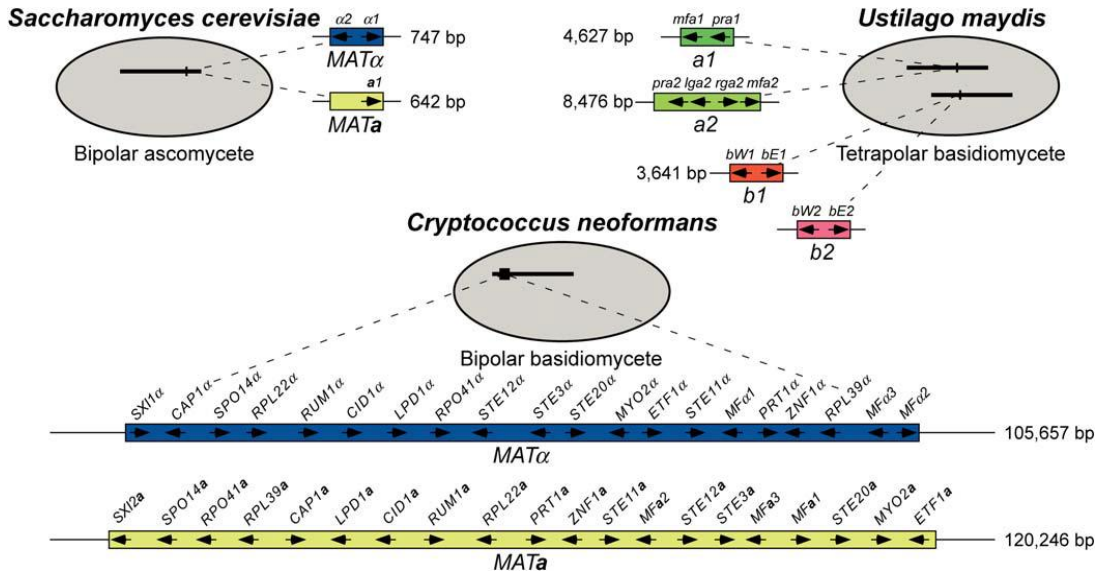


Figure 1.2 - Fungal *MAT* Locus Paradigms (Fraser *et al*., 2004). *MAT* loci of the bipolar ascomycete *Saccharomyces cerevisiae* (upper left), the tetrapolar basidiomycete *Ustilago maydis* (upper right) and the bipolar basidiomycete *Cryptococcus neoformans* (lower graphic). In *U. maydis*, the two *MAT* loci are physically unlinked. One encodes for pheromone and pheromone receptors and the other for homeodomain transcription factors. In *C. neoformans* the same locus encodes both pheromone and pheromone receptors and homeodomain transcription factors. The length is related to the extension of the non-recombining *MAT*-specific region in these organisms.

1.5 - Mating type loci in the Basidiomycota

In basidiomycetes, the mating type loci and their gene content differ between species; however, *MAT* genes comprise at least three classes: those encoding homeodomain transcription factors, pheromone precursors and cognate G protein-coupled 7-transmembrane pheromone receptors (Kües *et al*., 2011). These *MAT* genes are also found in self-fertile, homothallic species and in many anamorphic species. In some species, those genes are absent from the functional mating type loci, but present somewhere else in the genome (Kües *et al*., 2011).

1.5.1 - Pheromone and Pheromone Receptors

The first step of mating in most dimorphic basidiomycetes is the recognition of a compatible mating partner. This recognition occurs through export and sensing of pheromones. For different individuals to be sexually compatible, different pheromones and pheromone receptors must be produced. Pheromone and pheromone receptor genes are usually linked within *MAT* and differ when different alleles of the P/R locus are present in both individuals. The allele sequences of these genes from different mating types of the same species are generally highly divergent, as they regulate the mechanism by which self and non-self pheromones are recognized (Hsueh *et al*., 2011; Kües *et al*., 2011).

Basidiomycete pheromone genes encode precursors of small lipopeptide pheromones, which consist of a 9 to 14 amino acid chain covalently linked to an isoprenoid lipid after post-translational modifications. Pheromones are exported from the cell and interact with G protein-coupled (GPCR) 7-transmembrane (7TM) pheromone receptors, localized in the cellular membranes, inducing a subsequent mating reaction. Due to homology to the lipid-modified mating factor **a** pheromones (MF**a**), and the G protein-coupled 7-transmembrane Ste3 pheromone receptor of *S. cerevisiae*, the pheromone receptor genes of basidiomycetes are also labeled *STE3* (Kües *et al*., 2011). When a pheromone binds to the pheromone receptor of the compatible mating type, it stimulates conformational changes at the C-terminus of the receptor, which then interacts with a heterotrimeric G protein (Figure 1.3) (Kües *et al*., 2011). When activated, these G protein subunits transmit the pheromone-activated signal to a serine/threonine protein kinase intermediate, and then through a mitogen-activated protein (MAP) kinase transduction cascade. The MAP kinase pathway consists of three sequentially acting kinases: the downstream MAP kinase (MAPK) is activated by a MAP kinase kinase (MAPKK), which is activated by a MAP kinase kinase kinase (MAPKKK). The proteins involved in this pathway are encoded by genes, which may or may not be within the *MAT* locus (Casselton and Challen, 2006). In *C. neoformans* genes within the *MAT* locus include those encoding Ste20p, the first kinase in the pathway, Ste11p, the first kinase in the MAPK module (MAPKKK) and Ste12p, the pheromone response pathway target transcription factor (Figure 1.3) (Kozubowski *et al*., 2009; Raudaskoski and Kothe, 2010). Genes outside of the *MAT* locus such as *CPR2*, *STE7*, and

CPK1 also make up part of the pathway (Lin, 2009). The outcome response, generally involves transcriptional activation/repression of many different genes responsible for regulation of filamentous growth during nitrogen starvation, response to osmotic shock and other stresses, cope with cell wall stress or initiation of sporulation. In mating, the final product of this cascade is involved in the regulation performed by the homeodomain transcription factors in the nucleus, and together they regulate the next stages of sexual reproduction (Figure 1.3).

1.5.2 - Homeodomain Transcription Factors

Mating type genes encoding homeodomain (HD) transcription factors are divided into two different classes, *HD1* and *HD2*. These two genes classes are based on structural homologies and distinct sequences of the DNA-binding motifs of their deduced products (Kües *et al*., 2011). In basidiomycetes, these genes are generally arranged as adjacent and divergently transcribed genes within the HD locus. For a successful mating to occur, the products of both genes, but from different alleles, must interact with each other. Sequences of *HD1* and *HD2* alleles are usually highly divergent, and so are the resulting products. These differences are responsible for the mating type specificity of those proteins, and interaction between proteins from the same mating type is unsuccessful, being self-discriminatory. Interaction and dimerization between non-self proteins generates an active heterodimeric transcription factor (HD1-HD2) localized in the nucleus, which binds to specific sequences in the promoters of target genes and regulates the expression of genes involved in dikaryon development and sexual reproduction (Figure 1.3) (Kües *et al*., 2011).

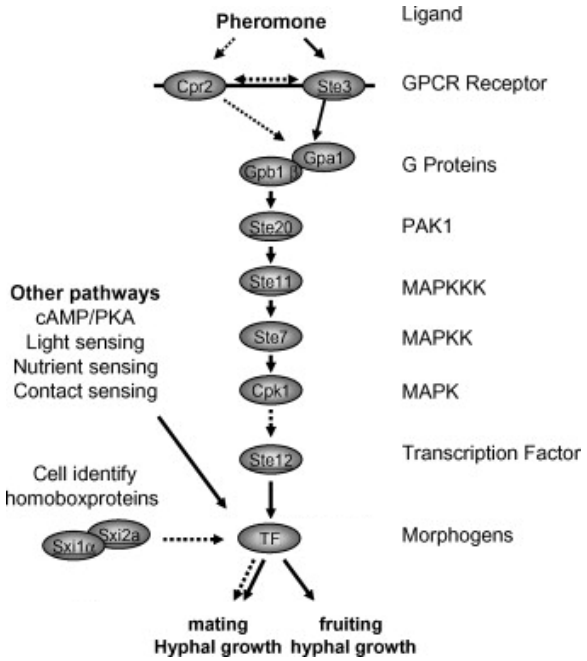


Figure 1.3 - Pheromone-activated MAPK transduction pathway in *C. neoformans* (adapted from Lin, 2009). MAPK signal transduction pathway and interactions with other pathways, leading to hyphal growth. Dotted arrows indicate predicted interactions. TF indicates unknown transcription factors.

1.6 - *Cryptococcus* species complex

The "*Cryptococcus* species complex" is a cluster of human pathogenic and dimorphic basidiomycetous yeasts, also known by the teleomorphic name *Filobasidiella* (based on morphological features of sexual state; Kwon-Chung, 2011), within the Filobasidiella clade, which belongs to the order Tremellales, class Tremellomycetes (Figure 1.4) (Findley *et al*., 2009). This clade comprises two sibling taxa, *Cryptococcus neoformans*, with two recognized varieties, *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii*, and *Cryptococcus gattii* (Campbell and Carter, 2006), which are responsible for respiratory diseases or neurological diseases in immunocompromised and immunocompetent patients, respectively (Casadevall and Perfect, 1998).

C. neoformans and *C. gattii* possess a biallelic bipolar mating system consisting of two opposite mating types, **a** and α (*MAT**a*** and *MATα*) (Lengeler *et al*., 2002). The length and gene content

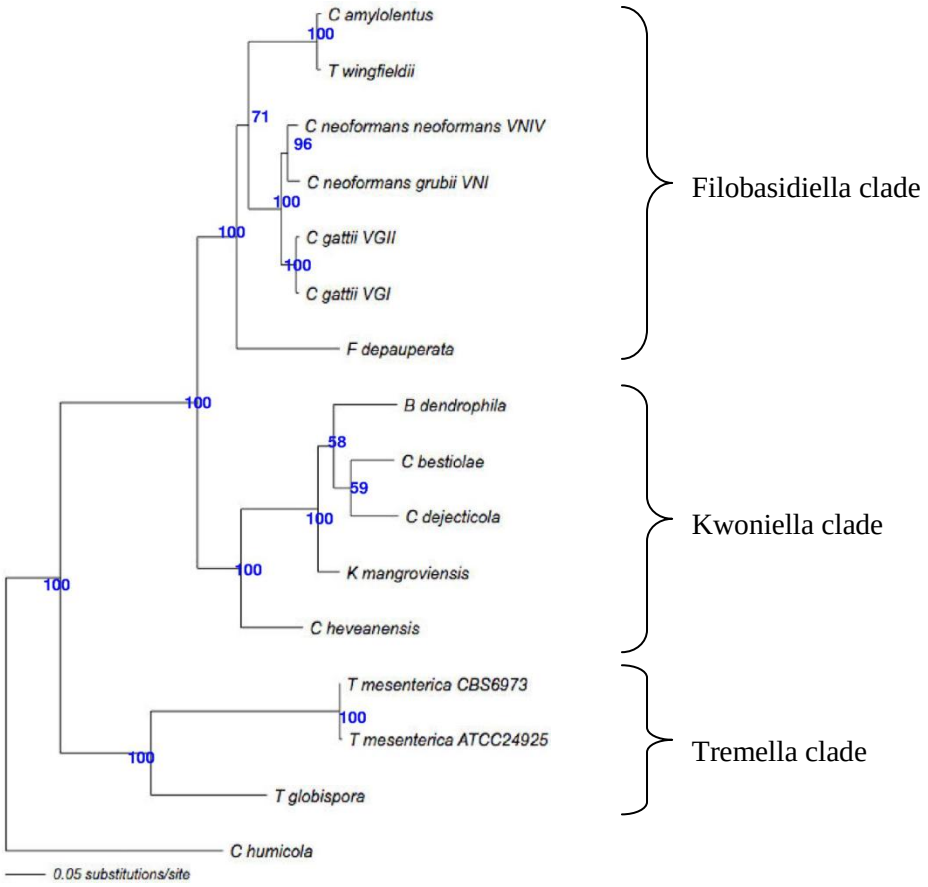


Figure 1.4 - Phylogenetic relationships among selected members of the Tremellales (adapted from Findley *et al*., 2009). Combined data set of six concatenated genomic loci (*RPB1*, *RPB2*, *EF1α*, mitSSU, nucLSU (D1/D2), and ITS) shows the presence of 3 genetically different clades: Filobasidiella (*C. amylolentus*, *T. wingfieldii*, *C. neoformans* var. *neoformans*, *C. neoformans* var. *grubii* and *C. gattii*), Kwoniella (*B. dendrophila*, *C. bestiolae*, *C. dejecticola*, *K. mangroviensis* and *C. heveanensis*) and Tremella (*T. mesenterica* and *T. globispora*). *C. humicola* is the outgroup. The tree was built using Maximum Likelihood method and the numbers on branches are bootstrap values from 1000 replicates.

of *MAT* alleles from both species are similar, but the order of genes is highly rearranged (Fraser *et al*., 2004). *MAT* alleles are very distinct between each other, but also between the same mating type of different species within this complex (Figure 1.5).

Each *MAT* locus possesses only one gene encoding for homeodomain transcription factors, Sxi1 or Sxi2. Sxi1 is encoded by *SXI1α* gene present only in *MATα* strains and Sxi2 is encoded by *SXI2**a*** gene present only in *MAT**a*** strains (Figure 1.5). The two compatible HD transcription factors, form a heterodimeric complex, which, like HD1-HD2, regulates the fusion of **a** and α cells, the dikaryotic hyphal formation and the establishment of the **a**/α cell identity (Lengeler *et al*., 2002). In all other basidiomycetes studied both divergently oriented *HD* transcription factor genes are present in the same *MAT* allele (Kües *et al*., 2011; Casselton and Challen, 2006).

In the *Cryptococcus* species complex, the *MAT* locus also contains highly divergent alleles of genes for mating type specific Ste3-like pheromone receptors (*STE3α* and *STE3**a***), three pheromone precursor genes for *MAT***a** (*MF***a***1*, *MF***a***2* and *MF***a***3*) or four pheromone precursor genes for *MATα*

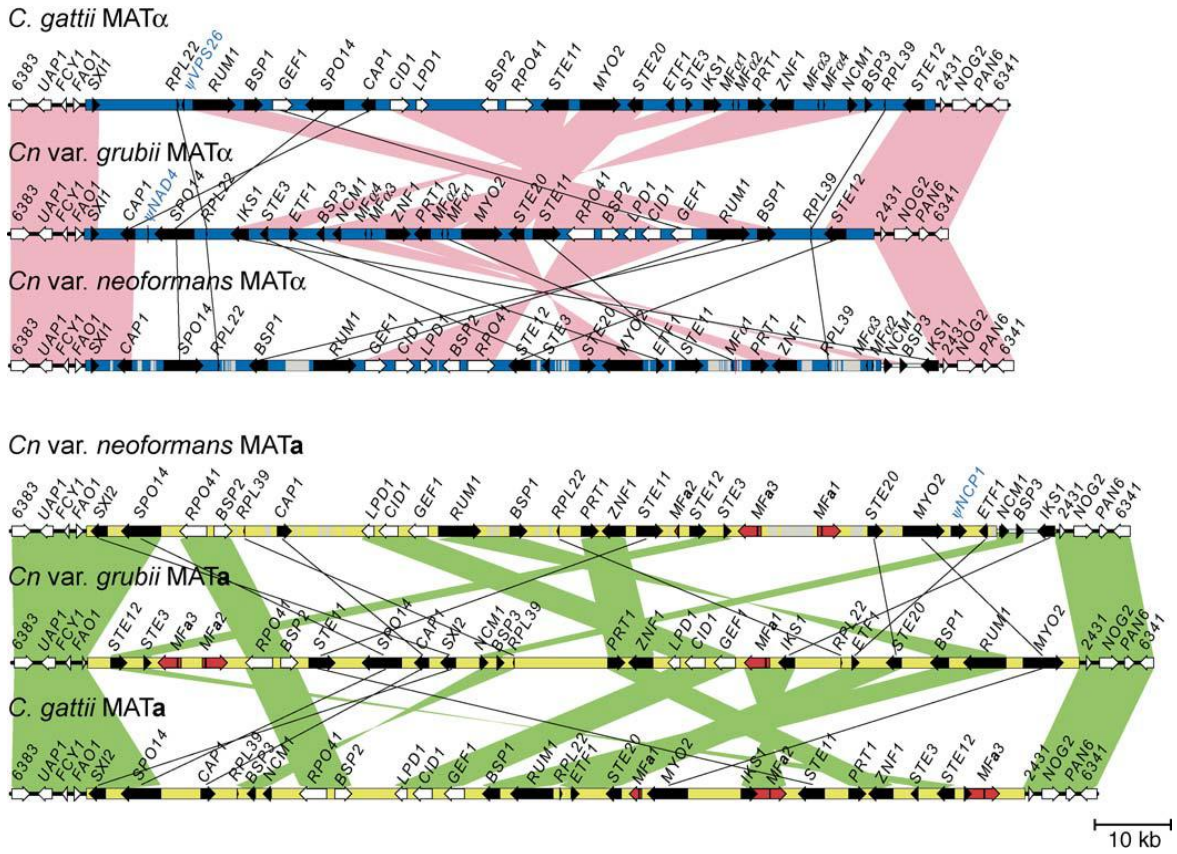


Figure 1.5 - *MAT* structure of *C. neoformans* var. *neoformans*, *C. neoformans* var. *grubii* and *C. gattii* (Fraser *et al*., 2004). Synteny analysis of the *MAT* locus of *C. neoformans* var. *neoformans*, *C. neoformans* var. *grubii* and *C. gattii* shows different gene organization and orientation between the same species and variety. Each species *MATα* is represented in blue while *MAT***a** is represented in yellow. Pink and green colored bars represents regions of synteny, and black lines the relative positions of genes whose position is not conserved. Black arrows represent mating type-specific genes. White arrows represent genes with a species-specific phylogeny. Red arrows represent pheromone precursor genes.

(*MFα1*, *MFα2*, *MFα3* and *MFα4*), and alleles for other genes, predominantly mating type specific, such as: *STE12**a*** and *STE12α*, encoding Ste12-like transcription factors that regulate some aspects of mating; *STE20**a*** and *STE20α*, for p21-activated protein kinases connected to the MAPK pathway; *STE11**a*** and *STE11α*, for kinases of the MAPK cascade; and *SPO14*, *RUM1* and *BSP1* alleles, which may function in meiosis and sporulation (Figure 1.5 and Table 1.1). *LPD1*, *CID1*, *RPO41*, *BSP2*, *GEF1* are five genes within the *Cryptococcus MAT* locus that exhibit species-specific phylogenies (Table 1.1), while highly conserving their synteny. Many other genes present within these extended loci do not have a known function in mating, namely *BSP2*, *BSP3*, *CAP1*, *CID1*, *ETF1*, *FAO1*, *GEF1*, *IKS1*, *LPD1*, *NOG2*, *PRT1* and *ZNF1* (Hsueh *et al*., 2011).

Phylogenetic analyses of the *MAT* genes revealed that the genes within this locus can be classified into mating type-specific, intermediate mating type-specific and species-specific patterns (summarized in Table 1.1) (Fraser *et al.*, 2004). Genes displaying mating type-specific phylogenies include those which are thought to have been most anciently acquired into this recombinationally suppressed region and are highly divergent between *MAT* alleles, namely those encoding for pheromones and pheromone receptors and genes involved in the MAPK pathway (Figure 1.3 and Figure 1.5, red and black arrows). Intermediate mating type-specific pattern genes have likely been contained within this region for a shorter period of time, and include those that cluster into a mating type-specific pattern, but for which different alleles share a higher percentage of similarity than the "ancient class" of genes (also represented as black arrows in Figure 1.5). Genes with a species-specific phylogenetic pattern are presumably the most recently acquired genes, and include genes that share a highly conversed sequence and nucleotide identity between alleles, similar to genes outside *MAT* (Figure 1.5, white arrows). The distinct phylogenetic patterns reflect the evolutionary origin of the different genes that compose the *MAT* locus, their relative acquisition into this genomic region and lastly, provide some clues about how this genomic structure evolved.

Based on these results, an evolutionary model was proposed by Fraser *et al*. (2004) (Figure 1.6). This evolutionary model suggests that the ancestral form of *MAT* was tetrapolar, one locus encoding pheromones and the pheromone receptor genes (P/R) and the other encoding one homeodomain transcription factor (HD). These ancient tetrapolar loci expanded to incorporate additional *MAT*-specific genes, beginning with the acquisition of components of the pheromone signaling MAPK cascade (*STE20*, *STE11* and *STE12*) into the P/R locus. A second round of acquisition of genes occurred, incorporating genes with unknown role in mating (intermediate class I) into the P/R locus, followed by acquisition of some genes involved in the dikaryon or meiosis (intermediate class II), into the HD locus. In one mating type, a chromosome fusion through translocation events and fused both loci, incorporating some species-specific genes (recent class) and creating an intermediate tripolar mating stage. Chromosomes from the opposite mating type were eventually fused as well by dual recombination events, turning this intermediate tripolar mating stage

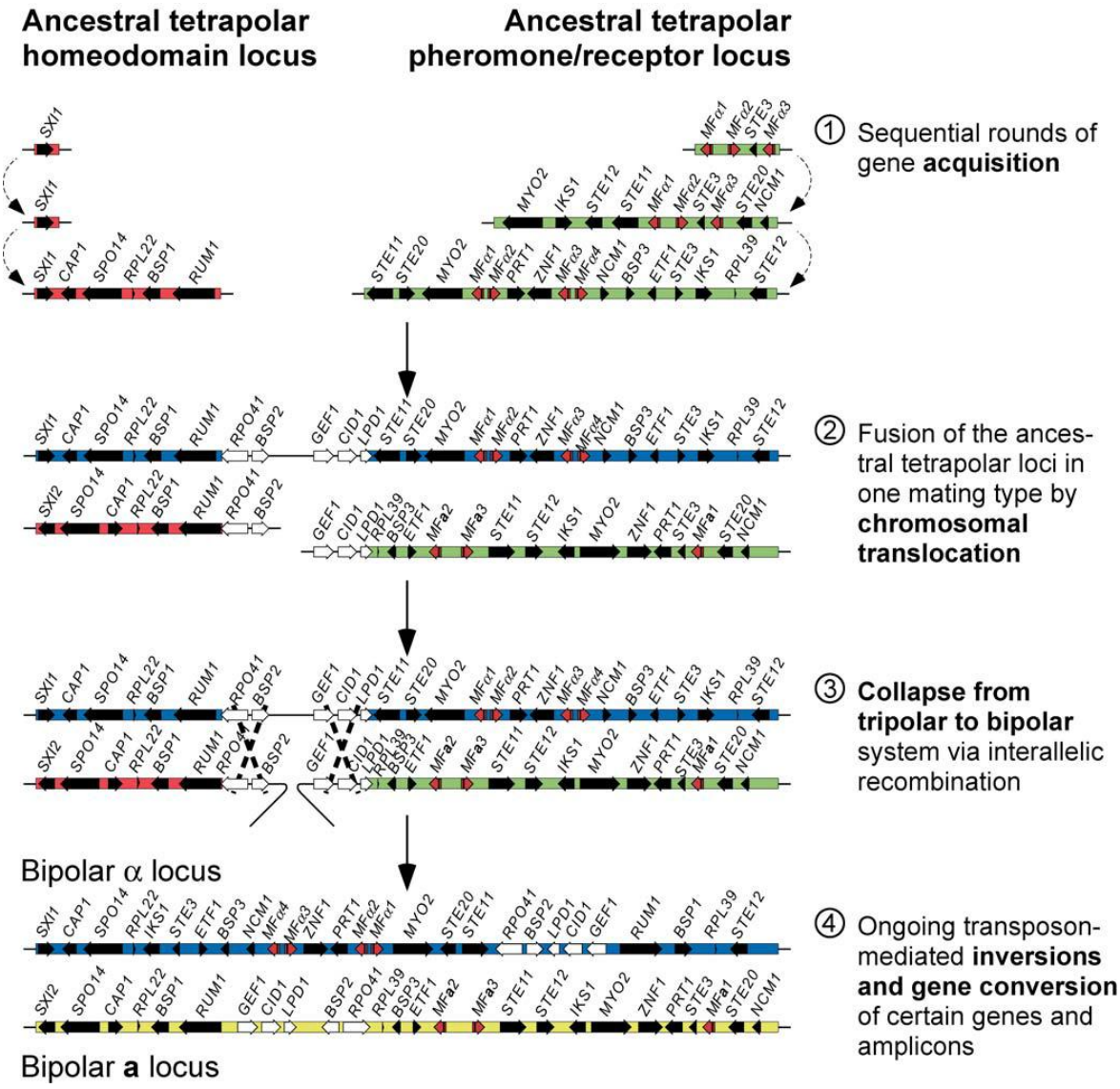


Figure 1.6 - Original model of *MAT* evolution (Fraser *et al.*, 2004). This evolutionary model proposes that the ancient P/R locus and the ancient HD locus were unlinked. These tetrapolar loci expanded twice and incorporated additional *MAT*-specific genes. In one mating type, chromosomal translocation events fused both loci, creating an intermediate tripolar mating state. Chromosomes from the opposite mating type were eventually fused by dual recombination events, turning this intermediate tripolar mating state into a bipolar mating system. Inversions produced homogeneous bipolar structures and suppressed recombination between *MAT* alleles.

into a bipolar mating system. Finally, inversions produced more homogeneous bipolar structures and suppressed recombination between *MAT* alleles.

A recent multilocus sequence typing study (Findley *et al.*, 2009) helped to clarify the phylogeny of taxa related to the *Cryptococcus* species complex. This study places this pathogenic complex phylogenetically closely related to saprobic basidiomycetous yeasts, namely, in the Filobasidiella, Kwoniella and Tremella clades (Figure 1.4). Comparative genomic approaches are useful to study and understand how the *Cryptococcus MAT* gene cluster evolved and also to test and

improve the proposed evolutionary model. Due to their phylogenetic relatedness, the analysis of *MAT* structure was therefore extended to the several species included in those clades.

1.7 - Evolution of *MAT* in the Filobasidiella clade

1.7.1 - *Cryptococcus amylolentus* and *Tsuchiyaea wingfieldii*

Cryptococcus amylolentus and *Tsuchiyaea wingfieldii* are saprobic basidiomycetous yeasts, phylogenetically closely related to the pathogenic *Cryptococcus* species complex (Figure 1.4). There are only two known strains of *C. amylolentus* and one of *T. wingfieldii* worldwide. The two available strains of *C. amylolentus* were found to be sexually compatible but *T. wingfieldii* lacks a known sexual cycle, maybe due to the lack of a compatible strain or sterility (Findley *et al*., 2012). Nevertheless, the *MAT* locus of both species was sequenced and analyzed (Figures 1.7 and 1.8) (Findley *et al*., 2012).

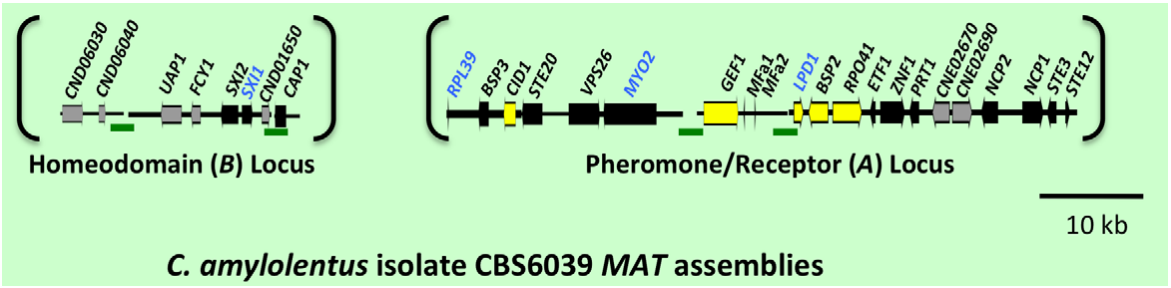


Figure 1.7 - *MAT* loci of *Cryptococcus amylolentus* CBS 6039 (Findley *et al*., 2012). Grey arrows represent genes that either flank *MAT* or are hypothetical genes, black arrows represent *MAT*-specific genes in pathogenic *Cryptococcus* species, and yellow indicates the genes more recently acquired into the pathogenic *Cryptococcus* species *MAT* locus. Green bars represent gaps in the sequence.

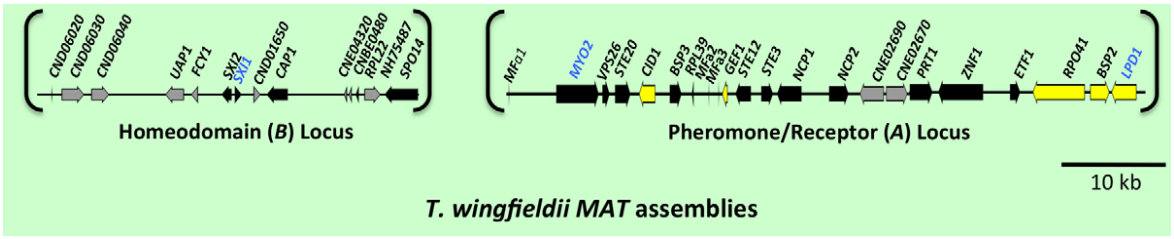


Figure 1.8 - *MAT* loci of *Tsuchiyaea wingfieldii* CBS 7118 (Findley *et al*., 2012). Grey arrows represent genes that either flank *MAT* or are hypothetical genes, black arrows represent *MAT*-specific genes in pathogenic *Cryptococcus* species, and yellow indicates the genes more recently acquired into the pathogenic *Cryptococcus* species *MAT* locus.

The P/R and HD loci were found to be physically unlinked and located on different chromosomes in both species, suggesting a tetrapolar mating configuration for *C. amylolentus* and *T. wingfieldii*. In *C. amylolentus*, the sequenced fragment of the P/R locus region spans about 60 kb, comprising 21 genes and the HD locus region spans about 20 kb, comprising eight genes (Figure 1.7), while in *T. wingfieldii*, the P/R locus region spans about 70 kb, comprising 22 genes and the HD locus

region spans about 40 kb, comprising 14 genes (Figure 1.8). The gene content seems to be highly conserved between the two species and similar to the *C. neoformans* and *C. gattii MAT* alleles (Figure 1.5) (Hsueh *et al*., 2011).

The P/R locus of both species contained the two pheromone-signalling MAPK pathway genes (*STE12* and *STE20*), the mating pheromone (*MF*) genes (two in *C. amylolentus* and three in *T. wingfieldii*), the pheromone receptor gene (*STE3*), and the five genes (*LPD1*, *RPO41*, *BSP2*, *CID1*, and *GEF1*) hypothesized to be the most recently acquired genes in *C. neoformans MAT* locus (Figure 1.6). In the pathogenic *Cryptococcus* species complex, *STE11* is present within the P/R locus (Figure 1.5), however, in *C. amylolentus* and *T. wingfieldii*, *STE11* gene is missing from the P/R locus region and located elsewhere in the genome. Additionally, *NCP1* and *NCP2* are duplicated genes in *C. amylolentus* and *T. wingfieldii* while there is only one of these genes in the *C. neoformans* var. *neoformans MAT***a** allele, suggesting that the other copy has been lost or translocated to another region in the genome, while in the other mating-type or in *C. neoformans* var. *grubii* and *C. gattii NCP1* and *NCP2* are absent. In both *C. amylolentus* and *T. wingfieldii* HD locus region, two divergently transcribed homeodomain transcription factors are present (*SXI1* and *SXI2*).

Phylogenetic analyses were performed on several genes (*CID1*, *ETF1*, *GEF1*, *LPD1*, *STE3*, *STE12*, *STE20*, *SXI1*, and *SXI2*) located within the *MAT* locus of *C. amylolentus* and *T. wingfieldii* that were also in *C. neoformans MAT* (Findley *et al*., 2012). The results showed that *CID1*, *GEF1*, *LPD1* and *ETF1* exhibit a species-specific pattern, while *STE3*, *STE12*, *STE20* exhibit a mating-type specific pattern and *SXI1* and *SXI2* are mating-type unique genes (see also Table 1.1, page 19). *ETF1* is species-specific in these two sibling species and an intermediate class I gene in *C. neoformans*. Comparing sequences of both *C. amylolentus* strains led the authors to conclude that the P/R locus, in this species, is defined by the mating pheromone genes (*MF*), *STE3*, *STE12* and *STE20*, among other genes, and that *SXI1* and *SXI2* span about 2 kb and define the HD locus. Due to the lack of additional *T. wingfieldii* strains, it was not possible to determine which genes define the P/R and HD loci in this species.

From synteny analysis between the two sibling species, *C. amylolentus* and *T. wingfieldii*, two major inversions were detected between the P/R regions of each species and a syntenic conservation of the HD regions (Findley *et al*., 2012). Synteny comparison of both species with *C. neoformans* and *C. gattii* revealed extensive gene rearrangements and inversions and that the *C. amylolentus* and *T. wingfieldii MAT* arrangement and gene content may correspond to an evolutionary intermediate in *MAT* evolution, in which the *MAT* loci have expanded but not yet fused. Homologues of several hypothetical genes contained in *C. amylolentus* and *T. wingfieldii MAT* were found in chromosome 4 (chromosome containing *MAT*) and 5 in *C. neoformans*, indicating that intra and interchromosomal

translocations may have occurred between these two chromosomes during the *MAT* evolution of the *Cryptococcus* species complex.

1.7.2 - *Filobasidiella depauperata*

Filobasidiella depauperata is a strictly filamentous species that forms sexual structures regardless of environmental conditions (Kwon-Chung, 2011). This species has aseptate basidia, monokaryotic hyphae without clamp connections and monokaryotic basidiospores. These morphological characteristics are similar to the modified form of sexual cycle of *C. neoformans* which involves monokaryotic fruiting (same-sex mating). A fragment of the putative *MAT* locus of *F. depauperata* was sequenced and analyzed (Figure 1.9) (Rodriguez-Carres *et al*., 2010).

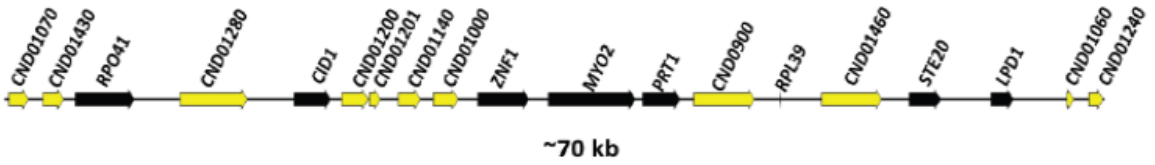


Figure 1.9 - Fragment of *MAT* locus of *Filobasidiella depauperata* CBS 7855 (Hsueh *et al*., 2011). Black arrows represent genes with homologues in the *MAT* locus of the pathogenic *Cryptococcus* species, and yellow represents additional genes identified in the other closely related species.

A contiguous sequence of about 70 kb was obtained, comprising 19 genes, corresponding to regions of the chromosome 4 of *C. neoformans* where *MAT* is located. Another sequence was obtained, comprising the *STE11* gene, corresponding to regions of chromosome 4 and 5 of *C. neoformans*. Nine (*LPD1*, *CID1*, *RPO41*, *STE20*, *MYO2*, *PRT1*, *ZNF1*, *RPL39* and *STE11*) of the 20 genes present in the *C. neoformans MAT* locus were present in the sequenced regions of *F. depauperata* (Figure 1.9, black arrows); however, pheromone, *STE3*, *STE12* and *SXI* genes were missing. None of the nine genes showed conserved synteny with *C. neoformans MAT* region. Phylogenetic analysis showed that *STE11*, *CID1*, *LPD1*, *RPO41*, *RPL22* and *ZNF1* had a species-specific pattern (Table 1.1). *STE20* and *MYO2* genes formed a monophyletic cluster between *F. depauperata* and one of the *C. neoformans* alleles, suggesting that, at least, *STE20* and *MYO2 MAT*-genes, may have been acquired into *MAT* prior to the *F. depauperata* divergence (Rodriguez-Carres *et al*., 2010). Three of the species-specific genes (*LPD1*, *CID1* and *RPO41*), and five of the ancestral *MAT* associated genes (*STE20*, *MYO2*, *PRT1*, *ZNF1* and *RPL39*) were found in a contiguous cluster, suggesting that at least three of the five recently acquired genes were closely located in the common ancestor of *C. neoformans* and *F. depauperata*.

Synteny analysis showed a higher number of translocations in *F. depauperata* corresponding to chromosomes 4 and 5 of *C. neoformans* (Rodriguez-Carres *et al*., 2010). One of the translocated

regions included *STE11*, which was translocated out of the *MAT* cluster, whereas it is part of the cluster associated with the pheromone and pheromone receptor genes in *C. neoformans* and *C. gattii*. From the evolutionary point of view, the results from synteny analysis between *F. depauperata* and *C. neoformans*, showed that some genes linked to the *MAT* locus in *C. neoformans* are present in different chromosomes in *F. depauperata*, suggesting that a higher number of chromosomal rearrangements occurred in the genes associated with *MAT* in *C. neoformans*, while the synteny in the other genomic regions was conserved. Chromosomal rearrangements appear, therefore, to be the most relevant event in speciation and sexual divergence in these closely related pathogenic and saprobic species (Hsueh *et al*., 2011; Rodriguez-Carres *et al*., 2010).

1.8 - Evolution of *MAT* in the Kwoniella clade

1.8.1 - *Cryptococcus heveanensis*

Cryptococcus heveanensis is a saprobic basidiomycetous yeast, located in the Kwoniella clade, a sister clade to the Filobasidiella clade (Figure 1.4). A sexual cycle was discovered for this species - teleomorph name *Kwoniella heveanensis* (Sheng et al., 2011) - and the *MAT* locus was sequenced and analyzed (Figure 1.10) (Metin *et al*., 2010). Two apparently unlinked genomic loci were found, P/R and HD loci, indicating a tetrapolar *MAT* organization for *C. heveanensis*. The region containing the P/R locus spans about 180 kb, comprising 47 genes, and the region containing the HD locus spans about 80 kb, comprising 22 genes.

The P/R locus and flanking regions contained a pheromone gene, the pheromone receptor gene *STE3*, the pheromone-signalling MAPK pathway genes *STE11*, *STE12* and *STE20*, and many other genes that are present in the *MAT* locus of *C. neoformans* and *C. gattii* (*LPD1*, *GEF1*, *CID1*, *BSP1*,

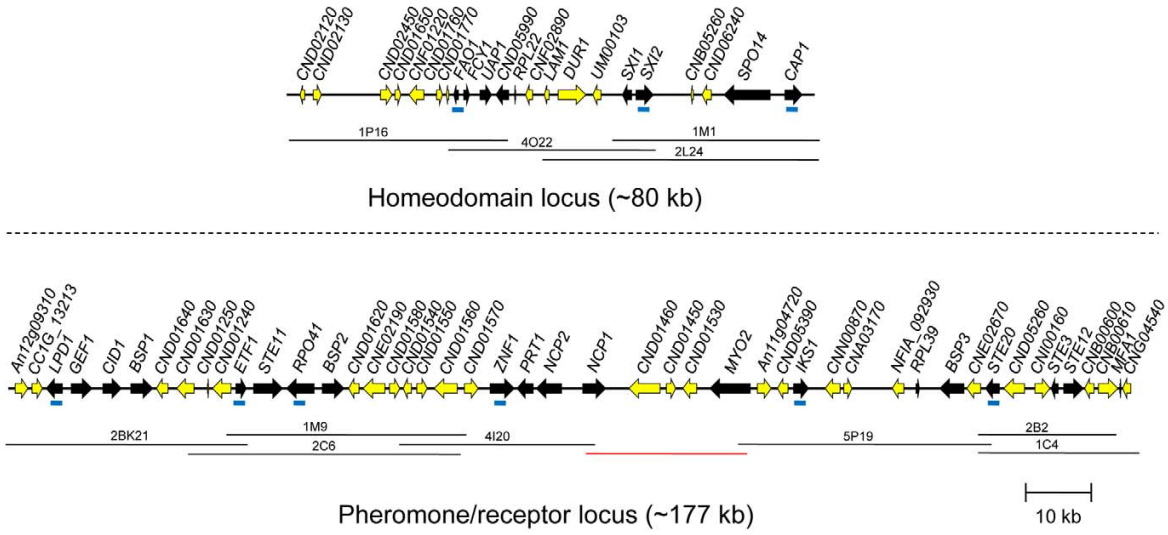


Figure 1.10 - *MAT* loci of *Cryptococcus heveanensis* CBS 569 (Metin *et al.*, 2010). Black arrows represent genes that are present in the *MAT* locus of the pathogenic *Cryptococcus* species and yellow represent genes without homologues in the *MAT* locus of the pathogenic *Cryptococcus* species.

ETF1, *RPO41*, *BSP2*, *ZNF1*, *PRT1*, *NCP1*, *MYO2*, *IKS1*, *RPL39*, and *BSP3*) (Figure 1.10, black arrows). There were additional genes that are not located within the *MAT* locus of *C. neoformans* and *C. gattii*, but located elsewhere in the genome (Table 1.1) and also genes that do not have homologues in these pathogenic species (*An11g04720*, *UM00103*, *UM02602*). This region contained also the five most recently acquired genes (*LPD1*, *GEF1*, *CID1*, *RPO41*, and *BSP2*) as hypothesized by the evolutionary model (Figure 1.6).

The HD locus region (Figure 1.10) contained two divergently transcribed HD genes (*SXI1* and *SXI2*), as well as *FAO1*, *FCY1*, *UAP1*, *RPL22*, *SPO14* and *CAP1* genes. *FAO1*, *FCY1* and *UAP1* are located in the left boundary of *C. neoformans* and *C. gattii MAT* locus, while *RPL22*, *SPO14*, *CAP1* are contained within the *MAT* locus of these pathogenic *Cryptococcus* species (Figure 1.5).

Phylogenetic analyses of *MAT* genes from several *C. heveanensis* strains (Metin *et al*., 2010) showed that *STE3* and *STE12* were the only genes exhibiting a mating-type specific pattern, along with *SXI1* and *SXI2*, in the sequenced regions (Table 1.1). Although, *BSP1*, *SPO14*, *ETF1*, *STE20* and *STE11* exhibit a mating type-specific pattern in *C. neoformans* and *C. gattii*, these genes are species-specific in *C. heveanensis*. *CID1*, *GEF1* and *LPD1* also exhibit a species-specific pattern. The P/R locus appeared to be at least biallelic and to contain at least the pheromone, *STE3*, *STE12*, *CNB00600* and *CNG04540* genes. The HD locus appeared to be multiallelic and defined by *SXI1* and *SXI2*. Specificity is defined by a 2 kb spanning region, including the intergenic region between the two genes, and the 5'-end of both genes, where the sequence is highly divergent (Figure 1.11).

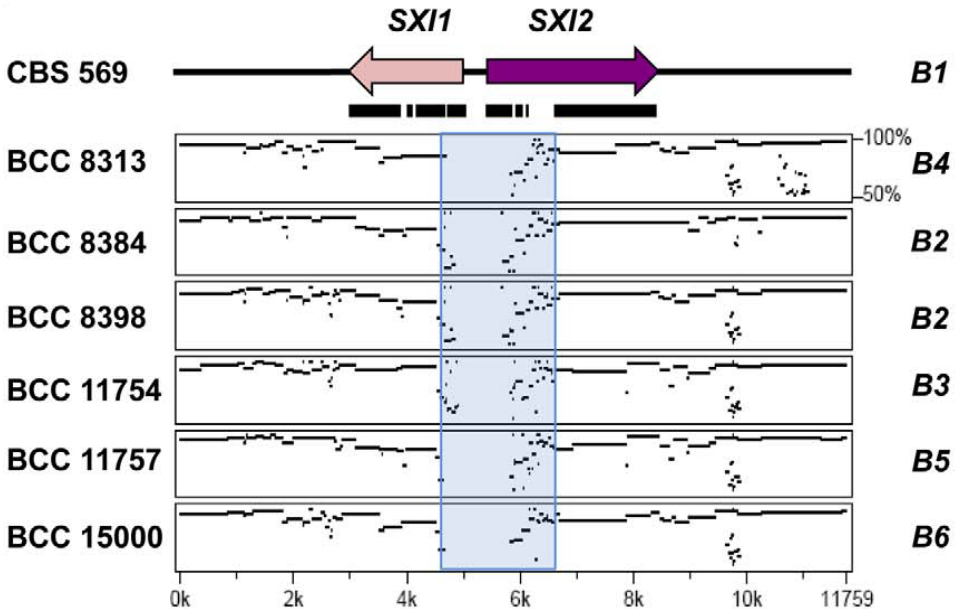


Figure 1.11 - Percent sequence identity plots of *SXI1* and *SXI2* (Metin *et al.*, 2010). Comparison of the same HD region, between several strains (left bold numbers) and alleles (B1-B6), shows a high variability of the sequence in the intergenic region and in the 5'-end of both genes (blue box). The blacks under the genes represent the exons.

The P/R and HD gene clusters of *C. heveanensis* have several linked genes that are not present in the *MAT* locus of the pathogenic *Cryptococcus* species complex (Figure 1.10, yellow arrows). Furthermore, some of these genes do not have an apparent homolog in these pathogenic species (*An11g04720, UM02602*), suggesting that these genes might have been lost during the evolution of the *Cryptococcus MAT* locus. The genes that have homologues in *C. neoformans* are located in chromosome 4, which comprises the *MAT* locus, suggesting intrachromosomal rearrangements during *MAT* evolution.

1.9 - Evolution of *MAT* in the Tremella clade

1.9.1 - *Tremella mesenterica*

Tremella mesenterica is a heterothallic basidiomycetous saprobic and dimorphic fungus, more distantly related to *C. neoformans* (Figure 1.4), that produces fruiting body structures (Bandoni and Boekhout, 2011). The whole genome of one strain of this species was sequenced by the Department of Energy's Joint Genome Initiative (JGI) (http://www.jgi.doe.gov/) and two unlinked *MAT* loci were determined and analyzed (Figure 1.12) (Findley, 2010; Metin *et al*., 2010).

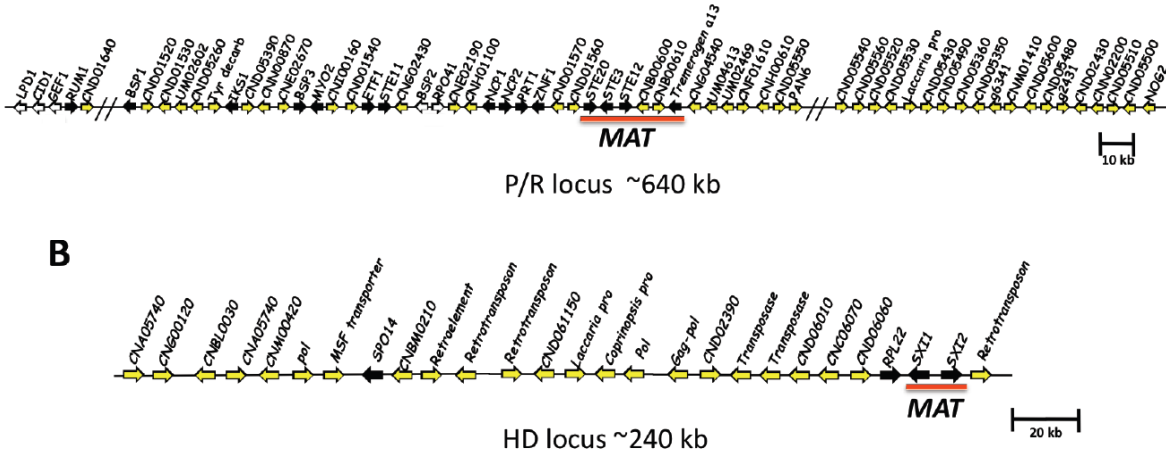


Figure 1.12 - *MAT* loci of *Tremella mesenterica* ATCC 24925 (adapted from Findley, 2010). Black arrows represent genes with homologs in *C. neoformans* and yellow indicates hypothetical genes. *MAT*-restricted genes are underlined in red.

Several hypothetical genes were found along the chromosomes, corresponding to homologues located in the chromosome containing *MAT* in *C. neoformans.* Homologues found in unrelated fungi are also located in the P/R locus (for example, *An11g04720, UM02602* and *UM04613*). The authors hypothesize that, in *C. neoformans*, these genes have been lost from *MAT* and relocated to the

telomeric ends of the chromosome containing *MAT* (Metin *et al*., 2010). Two copies of the *NCP1* gene (*NPC1* and *NCP2*) are present also as found in *C. amylolentus* and *T. wingfieldii*.

The HD region contained two homeodomain transcription factors (*SXI1* and *SXI2*) linked and divergently oriented. Additionally, there were two HD linked genes (*RPL22* and *SPO14*) also present, while *SPO14*, a gene closely located to HD genes in *C. neoformans*, is not adjacent to the HD genes in this species. Also, the *C. neoformans* HD flanking genes (*FAO1, UAP1,* and *FCY1*) (Figure 1.5) are not linked to the HD genes and are located elsewhere in *T. mesenterica* genome.

The five most recently acquired genes (*RPO41, BSP2, GEF1, CID1* and *LPD1*) in the proposed evolutionary model (Figure 1.6) are present, with *RPO41* and *BSP2* linked and located on the 5' of the P/R locus region, while *LPD1, CID1,* and *GEF1* are linked but located distantly from the P/R or HD loci (Figure 1.12).

Phylogenetic analyses for unlinked and *MAT*-linked genes (Findley, 2010; Findley *et al*., 2012) showed that *STE3, STE12* and *ETF1* exhibit a mating-type specific pattern, *GEF1, RPO41, CID1,* and *LPD1* a species-specific pattern and *SXI1* and *SXI2* were very divergent from their *C. neoformans* and *C. gattii* homologues (Table 1.1). From these analyses, it was determined that, in *T. mesenterica*, the P/R locus is biallelic and spans about 18 kb, defined by *STE3, STE12*, one pheromone and two hypothetical genes (*CNB00600* and *CNB00610*), while the HD locus is multiallelic and spans about 3 kb, and is defined by the two homeodomain transcription factors (*SXI1* and *SXI2*). In contrast with the pathogenic *Cryptococcus* species, in *T. mesenterica*, *STE11* is located outside the P/R locus, although it is within the same chromosome and relatively close to the *MAT* determining region.

1.10 - *MAT* gene content and organization comparison

From the gene content, phylogenetic pattern and synteny analyses performed on *C. amylolentus*, *T. wingfieldii*, *F. depauperata*, *C. heveanensis* and *T. mesenterica MAT* regions, comparisons were done with *C. neoformans* and *C. gattii*, in order to determine relevant similarities and differences that could provide some insight into this region's evolutionary process (Hsueh *et al*., 2011) (see also Table 1.1).

The presence of a tetrapolar mating system in *C. neoformans* and *C. gattii* closest relatives, *C. amylolentus* and *T. wingfieldii*, but also in more distant species, such as *C. heveanensis* and *T. mesenterica*, allows to infer that the tetrapolar configuration represents the ancestral form of *MAT* and that the bipolar mating system present in the pathogenic *Cryptococcus* species complex was developed more recently than hypothesized by the original evolutionary model (Findley et al., 2012; Hsueh *et al*., 2011; *Metin et al*., 2010; Rodriguez-Carres *et al*., 2010).

Table 1.1 - Summary of the *MAT* region gene content and phylogenetic pattern of *MAT* genes in the studied species.

Species	HD region		P/R region		
	HD genes[a]	Additional genes present (bold) or absent in *C. neoformans MATα*	Pheromone genes[a]	Pheromone Receptor gene[a]	Additional genes present (bold) or absent in *C. neoformans MATα**
C. neof. var. *neoformans* (*MATα*)	*SXI1α*	***BSP1*[c], *CAP1*[c], *FAO1*[b], *FCY1, RPL22*[c], *RUM1*[c], *SPO14*[c], *UAP1***	*MFα1, MFα2, MFα3*	*STE3*	***BSP2*[b], *BSP3*[b], *CID1*[b], *ETF1*[b], *GEF1*[b], *IKS1*[b], *LPD1*[d], *MYO2*[a], *NCM1, NOG2*[b], *PAN6, PRT1*[c], *RPL39*[c], *RPO41*[b], *STE11*[a], *STE12*[a], *STE20*[a], *ZNF1*[c]**
C. neof. var. *neoformans* (*MAT***a**)	*SXI2***a**	***BSP2*[b], *CAP1*[c], *FAO1*[b], *FCY1, RPL39*[c], *RPO41*[b], *SPO14*[c], *UAP1***	*MF***a***1, MF***a***2, MF***a***3*	*STE3*	***BSP1*[c], *BSP3*[b], *CID1*[b], *ETF1*[b], *GEF1*[b], *IKS1*[b], *LPD1*[d], *MYO2*[a], *NCM1, NOG2*[b], *PAN6, PRT1*[c], *RPL22*[c], *RUM1*[c], *STE11*[a], *STE12*[a], *STE20*[a], *ZNF1*[c], *NCP1***
C. neof. var. *grubii* (*MATα*)	*SXI1α*	***CAP1*[c], *FAO1*[b], *FCY1, RPL22*[c], *SPO14*[c], *UAP1,* *NADA4***	*MFα1, MFα2, MFα3, MFα4*	*STE3*	***BSP1*[c], *BSP2*[b], *BSP3*[c], *CID1*[b], *ETF1*[c], *GEF1*[b], *IKS1*[a], *LPD1*[d], *MYO2*[a], *NCM1, NOG2*[b], *PAN6, PRT1*[c], *RPL39*[c], *RPO41*[b], *RUM1*[c], *STE11*[a], *STE12*[a], *STE20*[a], *ZNF1*[c]**
C. neof. var. *grubii* (*MAT***a**)	*SXI2***a**	***BSP3*[c], *CAP1*[c], *FAO1*[b], *FCY1, NCM1, RPL39*[c], *SPO14*[c], *UAP1***	*MF***a***1, MF***a***2, MF***a***3*	*STE3*	***BSP1*[c], *BSP2*[b], *CID1*[b], *ETF1*[c], *GEF1*[b], *IKS1*[a], *LPD1*[d], *MYO2*[a], *NOG2*[b], *PAN6, PRT1*[c], *RPL22*[c], *RPO41*[b], *RUM1*[c], *STE11*[a], *STE12*[a], *STE20*[a], *ZNF1*[c]**
C. gattii (*MATα*)	*SXI1α*	***BSP1*[c], *CAP1*[c], *CID1*[b], *FAO1*[b], *FCY1, GEF1*[b], *LPD1*[b], *RPL22*[c], *RUM1*[c], *SPO14*[c], *UAP1,* *VPS26***	*MFα1, MFα2, MFα3, MFα4*	*STE3*	***BSP2*[b], *BSP3*[c], *ETF1*[c], *IKS1*[a], *MYO2*[a], *NCM1, NOG2*[b], *PAN6, PRT1*[c], *RPL39*[c], *RPO41*[b], *STE11*[a], *STE12*[a], *STE20*[a], *ZNF1*[c]**
C. gattii (*MAT***a**)	*SXI2***a**	***BSP3*[c], *CAP1*[c], *FAO1*[b], *FCY1, NCM1, RPL39*[c], *SPO14*[c], *UAP1***	*MF***a***1, MF***a***2, MF***a***3*	*STE3*	***BSP1*[c], *BSP2*[b], *CID1*[b], *ETF1*[c], *GEF1*[b], *IKS1*[a], *LPD1*[b], *MYO2*[a], *NOG2*[b], *PAN6, PRT1*[c], *RPL22*[c], *RPO41*[b], *RUM1*[c], *STE11*[a], *STE12*[a], *STE20*[a], *ZNF1*[c]**

…/…

Table 1.1 (cont.) - Summary of the *MAT* region gene content and phylogenetic pattern of *MAT* genes in the studied species.

Species	HD region		P/R region		
	HD genes[a]	Additional genes present (bold) or absent in *C. neoformans MATα*	Pheromone genes[a]	Pheromone Receptor gene[a]	Additional genes present (bold) or absent in *C. neoformans MATα**
C. amylolentus	*SXI1α*, *SXI2***a**	***CAP1, FCY1, UAP1*** *CND01650, CND06030, CND06040*	*MF***a***1*, *MF***a***2*	*STE3*	***BSP2, BSP3, CID1[b], ETF1[b], GEF1[b], LPD1[b], MYO2, PRT1, RPL39, RPO41, STE12[a], STE20[a], ZNF1,*** *CNE02670, CNE02690, NCP1, NCP2, VPS26*
T. wingfieldii	*SXI1α*, *SXI2***a**	***CAP1, FCY1, RPL22, SPO14, UAP1,****CNBE0480, CND01650, CND06020, CND06030, CND06040, CNE04320, NH75487*	*MF***a***1*, *MF***a***2*, *MF***a***3*	*STE3*	***BSP2, BSP3, CID1[b], ETF1[b], GEF1[b], LPD1[b], MYO2, PRT1, RPL39, RPO41, STE12[a], STE20[a], ZNF1,*** *CNE02670, CNE02690, NCP1, NCP2, VSP26*
F. depauperata	-	-	-	-	***CID1[b], LPD1[b], MYO2[a], PRT1, RPL39[b], RPO41[b], STE20[a], ZNF1[b],*** *CND01240, CND01430, CND01460*
C. heveanensis	*SXI1α*, *SXI2***a**	***CAP1, FAO1, FCY1, RPL22, SPO14[b], UAP1,*** *UM00103*	*MFA1* or *MFA2*	*STE3*	***BSP1[b], BSP2, BSP3, CID1[b], ETF1[b], GEF1[b], IKS1, LPD1[b], MYO2, PRT1, RPL39, RPO41, STE11[b], STE12[a], STE20[b], ZNF1,*** *An11g04720, CNA03170, CNB00600, CNB00610, CND01240, CND01250, CND01450, CND01460, CND01530, CND01540, CND01550, CND01560, CND01570, CND01580, CND01630, CND01640, CND05260, CND05390, CNE02190, CNE02670, CNG04540, CNI00160, CNN00870, NCP1, NCP2, UM02602*
T. mesenterica	*SXI1α*, *SXI2***a**	***RPL22[b], SPO14,*** *UM00103*	*Tremerogen-**a**13*	*STE3*	***BSP1, BSP2, BSP3, CID1[b], ETF1[a], GEF1[b], IKS1, LPD1[b], MYO2, PAN6, PRT1, RPL39, RPO41[b], RUM1, STE11, STE12[a], STE20, ZNF1,*** *An11g04720, CNA03170, CNB00600[a], CNB00610[a], CND01240, CND01250, CND01430, CND01450, CND01460, CND01530, CND01540, CND01550, CND01560, CND01570[b], CND01580, CND01630, CND01640, CND05260, CND05390, CNE02190, CNE02670, CNG04540[b], CNI00160, CNN00870, NCP1, NCP2, UM02602*

a – Mating-type specific genes; b – Species-specific genes; c – Intermediate I/II genes; d – 3' Mating-type specific and 5' species-specific gene; Underlined– The hypothesized 5 most recently acquired genes; Bold – genes present in *C. neoformans MATα*; * – Only genes with homologs within *MAT* region in other species are listed.

In the pathogenic *Cryptococcus* species complex, *STE11* is present within the P/R locus. In *C. amylolentus*, *T. wingfieldii* and *F. depauperata*, this gene is missing from the P/R locus and located elsewhere in the genome; however, in *T. mesenterica* and *C. heveanensis*, *STE11* is linked but distant from the P/R locus (Table 1.1). Synteny analysis showed a higher number of translocations in *F. depauperata* corresponding to chromosomes 4 and 5 of *C. neoformans*. One of the translocated regions included *STE11*, which was translocated out of the *MAT* cluster, whereas it was found as part of the cluster associated with the pheromone and pheromone receptor genes in closely related species, such as *C. neoformans*, *C. gattii*, *C. heveanensis* and *T. mesenterica*, but not in *C. amylolentus* and *T. wingfieldii*. These results suggest that the presence of *STE11* gene in the P/R gene cluster may represent the ancestral configuration with retention in *C. neoformans* and *C. gattii* and translocation out of the *MAT* in *C. amylolentus*, *T. wingfieldii* and *F. depauperata* (Findley *et al*., 2012).

Pheromone and pheromone receptor genes in the genomes of the heterothallic species *C. neoformans*, *C. gattii* (Figure 1.5), *C. heveanensis* (Figure 1.10) and *T. mesenterica* (Figure 1.12), are arranged in a cluster that includes STE12, *STE20*, *PRT1*, *ZNF1*, *RPL39* and *STE11* (Table 1.1), suggesting that this gene cluster is ancestral to species divergence and that heterothallism is the ancestral mode of reproduction. *STE20*, *PRT1*, *ZNF1* and *RPL39* are also clustered with pheromone and pheromone receptor genes in *C. amylolentus* (Figure 1.7) and *T. wingfieldii* (Figure 1.8), supporting the idea of antiquity of this gene cluster and the translocation of *STE11*out of the *MAT*, in the latter two species.

The P/R region of *C. amylolentus*, *T. wingfieldii*, *C. heveanensis* and *T. mesenterica* contains the five most recently acquired genes (*LPD1*, *GEF1*, *CID1*, *RPO41*, and *BSP2*) (Table 1.1) hypothesized by the evolutionary model (Figure 1.6). According to the hypothesis (Fraser *et al*., 2004), *RPO41* and *BSP2* were located at the boundary of the ancient HD locus and *LPD1*, *GEF1* and *CID1* at the boundary of the ancient P/R locus of the ancient *MAT*, before these regions fused. The presence of these five genes in those phylogenetically distant species implies that these genes were likely to be linked to the ancestral P/R locus and may have been acquired by the ancestor of the pathogenic *Cryptococcus* cluster during expansion of this locus before the *MAT* fusion event. Additionally, *NCP1* and *NCP2* are duplicated genes in *C. amylolentus*, *T. wingfieldii*, *C. heveanensis* and *T. mesenterica*, while there is only one of these genes and only in *C. neoformans* var. *neoformans MAT***a** (Figure 1.5), suggesting that the presence of both genes is the ancestral configuration. In the HD locus of *C. amylolentus* (Figure 1.7), *T. wingfieldii* (Figure 1.8), *C. heveanensis* (Figure 1.10) and *T. mesenterica* (Figure 1.12), two divergently transcribed homeodomain transcription factors are present (*SXI1* and *SXI2*), suggesting that this configuration was present in the ancestral species to the *C. neoformans* and *C. gattii* clade ancestral species. The loss of one of the HD genes might have occurred during the evolution of *MAT* in the pathogenic *Cryptococcus* species.

FAO1, *FCY1* and *UAP1* are located in the left boundary of *C. neoformans* and *C. gattii MAT* locus, while *RPL22*, *SPO14*, *CAP1* are contained within the *MAT* locus of these pathogenic *Cryptococcus* species (Figure 1.5), being linked to the HD genes. The HD locus region, of *C. heveanensis*, contains all these six genes, while in *T. mesenterica*, *SPO14* and *RPL22* are linked to the HD genes. *UAP1*, *FCY1*, *RPL22* and *SPO14* are found in the HD region of *T. wingfieldii* and *UAP1*, *FCY1* and *CAP1* in *C. amylolentus*. This suggests that the five most recently acquired genes may have been acquired due to the expansion of the P/R locus rather than entrapped by the *MAT* fusion event.

In *C. neoformans*, *C. gattii* and *C. amylolentus STE20* shows a mating-type specific pattern, but in *C. heveanensis*, this is a species-specific gene. The mating-type specificity of *STE20* may have been acquired only in the former *Cryptococcus* species or lost in *C. heveanensis*. The species-specific phylogeny pattern of *CID1*, *GEF1* and *LPD1*, in *C. neoformans, C. gattii, C. amylolentus, C. heveanensis* and *T. mesenterica*, is consistent with the hypothesis that these genes had been recruited into the *MAT* locus of *C. neoformans* during the transition from a tetrapolar to a bipolar mating system. Although a mating-type specific gene in *T. mesenterica, ETF1* is species-specific in *C. amylolentus*, *T. wingfieldii* and *C. heveanensis*, but an intermediate class I gene in *C. neoformans*. This gene may have acquired its mating-type specificity in the pathogenic *Cryptococcus* species complex, after the common ancestor of *C. neoformans* and *C. gattii* diverged from *C. amylolentus* and *T. wingfieldii* common ancestor.

The P/R and HD gene clusters of *C. heveanensis* (Figure 1.10) have several linked genes that are either present outside of the *MAT* locus (*CND01630*, *CND01640*, *CND05260*, *CND05390*, *CNN00870*, *CNE02670*, *CNI00160*, *CND01540*, *CNG02430*, *CNE02190*, *CND01240*, *CND01560*, and *CND01570*) or are completely absent (*An11g04720* and *UM02602*) in the genome of the pathogenic *Cryptococcus* species complex (Figure 1.5). However, most of these genes are also present in the P/R locus of *T. mesenterica* (Figure 1.12), suggesting that these genes were lost recently and/or relocated to other genomic regions in the Filobasidiella clade.

1.11 - Evolutionary model revision

From *C. amylolentus*, *T. wingfieldii*, *F. depauperata*, *C. heveanensis* and *T. mesenterica MAT* analyses, several amendments were suggested to the original evolutionary model (Figure 1.6) (Findley, 2010; Findley *et al*., 2012; Hsueh *et al*., 2011; Metin *et al*., 2010; Rodriguez-Carres *et al*., 2010). The revised evolutionary model suggests that, the ancient HD locus contained both HD genes (*SXI1* and *SXI2*) in a divergent orientation. Both the HD and P/R loci underwent gene acquisition and further expansion, collecting additional *MAT*-specific genes, and the five most recently acquired genes in *C. neoformans* and *C. gattii* were linked at the boundary of the ancestral P/R locus. The loss of one of the HD genes occurred either simultaneously or prior to a chromosomal translocation event. Intrachromosomal rearrangements and a translocation event occurred in and between chromosomes 4

and 5 of *C. neoformans*, generating a transient tripolar intermediate. Some of the genes entrapped in *MAT* were translocated to somewhere else in the genome, or eventually lost.

1.12 - *Kwoniella mangroviensis*

Kwoniella mangroviensis is a saprobic basidiomycetous yeast, located in the Kwoniella clade (Figure 1.4). In the original study describing this species (Statzell-Tallman *et al*., 2008), it was characterized as possessing a heterothallic life cycle and a bipolar mating-system. Being phylogenetically close to a tetrapolar species, *C. heveanensis*, in a sister clade to Filobasidiella, which comprises bipolar (*Cryptococcus* species complex) and tetrapolar species (*C. amylolentus* and *T. wingfieldii*), the *MAT* analysis of *K. mangroviensis* became mandatory in order to provide insight into *Cryptococcus MAT* evolution.

A preliminary study (Á. Fonseca, unpublished) was done to clarify the *MAT* structure and gene content of *Kwoniella mangroviensis*. The study began with the construction of fosmid libraries of two sexually compatible strains (K1 and K2) (Table 1.2). Degenerate primers for *MAT* genes (*SXI1/SXI2, RPL22, CAP1, STE20, CID1, LPD1* and *RPO41)* were then designed based on available sequences of *C. neoformans* and related species. The designed degenerate primers were then used for PCR-amplification, cloning and sequencing, and sequences from *SXI1/SXI2* (including the 5'-end from both genes and the intergenic spacer), *STE20, LPD1* and *RPO41* genes were obtained for both strains. The sequences were analyzed and compared, revealing a mating-type specific pattern for *SXI1*/*SXI2* and *STE20* and a species-specific pattern for *LPD1* and *RPO41*. Based on the sequences obtained, specific probes were designed for *STE20* and *SXI* genes for K1, *STE20* for K2 and *RPO41*, *LPD1* for both strains. The fosmid libraries were then probed and the positive fosmids were confirmed by PCR. Positive results for *STE20* were obtained in K1 libraries and *LPD1* and *RPO41* in K2 libraries. To ensure that the fosmid included *MAT* genes, the positive clone (fosmid 2K2) for *STE20* was selected for sequencing in the present study.

Subsequently, another study (J.A. Rodrigues and Á. Fonseca, unpublished) was conducted to clarify the mating system of this species. The study included seven strains from the work of Statzell-Tallman *et al*. (2008), with three *MAT*α strains (K1, K3 and K6), one *MAT***a** strain (K2) and three strains with unknown *MAT* (K4, K5 and K7). Crossing experiments were performed (Table 1.2) and *STE20* and *SXI1, SXI2* and their intergenic region, were partially sequenced. The results showed the presence of two alleles for the P/R locus (A1 and A2) and three alleles were obtained for the HD locus (B1, B2 and B3). The molecular mating-type for each strain was determined and it revealed the presence of four different mating-types, with different allele combinations among strains (Table 1.2). Mating experiment results showed correlation to molecular mating-types since positive matings were only observed for strains having different alleles at both loci (Table 1.2). The observed allele

recombination and the mating ability between compatible strains were the first evidence that *K. mangroviensis* is, in fact, a heterothallic tetrapolar species and not bipolar as initially described.

Table 1.2 - Summary of previous mating experiments.

Molecular *MAT*	Strain (*MAT*[1])	**K1**	**K3**	**K6**	**K7**	**K4**	**K5**	**K2**
A1B1	**K1** (α) CBS 8507(T)	No						
	K3 (α) CBS 10793	No	No					
A1B3	**K6** (α) CBS 10792	No	No	No				
	K7 (?) ML 4078	No	No	No	No			
A2B1	**K4** (?) * CBS 8886	No	No	Yes	Yes	No		
	K5 (?) * ML 3895	No	No	Yes	Yes	No	No	
A2B2	**K2** (A) CBS 10435	Yes	Yes	Yes	Yes	No	No	No

Yes – Presence of dikaryotic mycelium. No – Absence of dikaryotic mycelium. *Self-filamentous. 1 – *MAT* designation from Statzell-Tallman *et al.* (2008); ? – no mating reaction.

1.13 - Unveiling the mating system and genetic variability in *Kwoniella mangroviensis* using molecular approaches

The study described in this thesis aimed to expand the preliminary studies on the mating system and the *MAT* structure in the Kwoniella clade, particularly in *Kwoniella mangroviensis*, and to determine genetic variability and clarify species boundaries using molecular markers. The *K. mangroviensis* isolates studied included strains from the original work of Statzell-Tallman *et al*. (2008) sampled from Bahamas and Spain, but also additional isolates sampled from the British Virgin Islands, Botswana, Switzerland, Slovenia and a clinical isolate from the USA. Genetic variability was determined through molecular typing methods, namely DNA-fingerprinting and MLST analyses. *MAT* structure and organization was further explored based on *STE20* and *SXI1*/*SXI2* analyses and on the sequencing of a fragment of the P/R locus of the type strain of *K. mangroviensis*.

2 - MATERIALS AND METHODS

2.1 - Yeast cultures

The yeast cultures used in this study are listed in Table 2.1. Strains were obtained from the following culture collections: Portuguese Yeast Culture Collection, Portugal (PYCC); Agricultural Research Service Culture Collection, USA (NRRL); The Spanish Type Culture Collection, Spain (CECT); Centraalbureau voor Schimmelcultures, The Netherlands (CBS); Culture Collection of Industrial Microorganisms, Slovenia (ZIM); (Agro)Industrial Fungi & Yeasts Collection, Belgium (MUCL); Mopane tree isolates and a clinical isolate (DS) were provided by Dr. Deborah Springer (Department of Molecular Genetics and Microbiology, Duke University, USA). All isolates were grown and maintained on MYP medium at 25ºC and 4ºC, respectively. MYP medium contained 0.7% (w/v) Malt extract, 0.25% (w/v) Soytone 0.05% (w/v) Yeast Extract and 1.5% Agar.

2.2 - Physiological Tests

Physiological tests were performed on sterile 96 well microplates (Nunclon™Δ Surface, Denmark) according to Kurtzman *et al*. (2011b) and read with *StatFax 2100* microplate reader (Awareness Technology Inc., USA) using absorbance measures at 630 nm. Media were prepared according to the information available on the CBS website (http://www.cbs.knaw.nl/collections/DefaultInfo.aspx?Page=YeastMethods, version of March 16th, 2011). The inoculum was grown overnight on liquid MYP medium at 25ºC, in Certomat® U orbital shaker (Sartorius, Germany) at 48 rpm, and diluted on sterile distilled water for the plate inoculation. Each well contained 200µl of medium and 5µl of diluted inoculum ($O.D._{600nm}$ = 0.04). After inoculation, the microplates were sealed with a sealing pellicle (Nunclon™Δ Surface, Denmark) and were incubated at 25ºC in a Denley Wellwarm 1 (Denley) microplate incubator with shaking, up to 3 weeks.

2.3 - Mating Assays

To test for sexual compatibility, isolates were grown on MYP agar plates at 25ºC for 3 days. A dense suspension was made from the cultures, for each isolate, in sterile distilled water, and 5 µl of each pair of cell suspensions were inoculated and mixed together on a plate. Mating assays were performed on CMA, YCB and MEA media at 25ºC and examined with a phase contrast microscope (Zeiss, Germany), after 2-4 weeks incubation, for the presence of sexual structures. CMA medium contained 1.5% (w/v) Corn meal agar and 0.5% (w/v) Agar. YCB medium contained 1.17% (w/v) Yeast carbon base and 2% Agar. MEA medium contained 2.5% (w/v) Malt extract and 2% (w/v) Agar.

Table 2.1 - Yeast cultures used in this study.

Species	Code	Strain Numbers[a]	Substrate	Geographic Origin	*MAT* Assignment
Kwoniella mangroviensis	K1	CBS 8507(T); PYCC 5950; ML 3810	Seawater/Mangrove	Bahamas	α^{b} (A1B1)[c]
	K2	CBS 10435; PYCC 5951; ML 4135	Seawater/Mangrove	Bahamas	A^{b} (A2B2)[c]
	K3	CBS 10793; PYCC 5959; ML 4671	Seawater/Mangrove	Bahamas	α^{b} (A1B1)[c]
	K4	CBS 8886; PYCC 5958; ML 3904	Seawater/Mangrove	Bahamas	$?^{b}$ (A2B1)[c]
	K5	ML 3895; PYCC 5961	Seawater/Mangrove	Bahamas	$?^{b}$ (A2B1)[c]
	K6	CBS 10792; PYCC 5960; ML 4136	Seawater/Mangrove	Bahamas	α^{b} (A1B3)[c]
	K7	ML 4078; PYCC 5962	Seawater/Mangrove	Bahamas	$?^{b}$ (A1B3)[c]
	K8	ML 4631; PYCC 6328; NRRL Y-48772	Seawater/Mangrove	Bahamas	α^{b}
	K9	DS 873; B9012	Human pleural peel tissue	USA	
	K10	DS 258	Mopane tree	Botswana	
	K11	DS 418	Mopane tree	Botswana	
	K14	DS 282	Mopane tree	Botswana	
	K16	DS 226	Mopane tree	Botswana	
	K17	DS 729	Mopane tree	Botswana	
	K20	DS 267	Mopane tree	Botswana	
	K21	DS 269	Mopane tree	Botswana	
	K22	DS 333	Mopane tree	Botswana	
	K23	DS 731	Mopane tree	Botswana	
	K24	DS 732	Mopane tree	Botswana	
	K25	CECT 11955; PYCC 6162; ML 5063	Cork	Spain	$?^{b}$
	K26	CECT 11979; PYCC 6204	Cork	Spain	$?^{b}$
	K27	CBS 11279; PYCC 6203; ML 2772	Seawater/Reef	British Virgin Islands	
	K28	CBS 7868; PYCC 6329	Grape juice	Switzerland	
	K29	ZIM 605; PYCC 6207	Grape berries	Slovenia	
Bullera dendrophila	Bd	CBS 6074(T); PYCC 6298; ATCC 24608	Insect frass	South Africa	
Cryptococcus amylolentus	Ca	CBS 6039(T); PYCC 4486; ATCC 56469	Insect larvae frass	South Africa	A1B1[d]
	Ca2	CBS 6273; PYCC 6299	Insect frass	South Africa	A2B2[d]
Cryptococcus bestiolae	Cb	CBS 10118(T); PYCC 5964	Insect frass	Vietnam	
Cryptococcus dejecticola	Cd	CBS 10117(T); PYCC 5965	Insect frass	Vietnam	
Cryptococcus heveanensis	Ch	CBS 569(T); PYCC 3877; ATCC 32065	Sheet rubber	Indonesia	A1B1[e]
	Ch2	BCC 8398	Insect frass	Thailand	A2B2[e]

…/…

Table 2.1 (cont.) - Yeast cultures used in this study.

Species	Code	Strain Numbers[a]	Substrate	Geographic Origin	*MAT* Assignment
Cryptococcus pinus	Cp1	CBS 10737(T); PYCC 6206; VKM Y-2958	Dead needles of pine tree	Russia	
	Cp2	MUCL 53261; PYCC 6327	Insect larvae	Belgium	
	Cp3	MUCL 53265	Insect larvae	Belgium	
	Cp4	MUCL 53268	Insect larvae	Belgium	
Cryptococcus shivajii	Cs	CBS 11374(T); PYCC 6300	Biogas reactor	United Kingdom	
Tsuchiyaea wingfieldii	Tw	CBS 7118(T); PYCC 5373; NRRL Y-17143	Insect frass	South Africa	

a – **PYCC**: Portuguese Yeast Culture Collection, Portugal; **CBS**: Centraalbureau voor Schimmelcultures, The Netherlands; **ML**: University of Miami, USA; **NRRL**: Agricultural Research Service Culture Collection, USA; **DS**: Isolates provided by Dr. Deborah Springer (Department of Molecular Genetics and Microbiology, Duke University, USA); **CECT**: The Spanish Type Culture Collection, Spain; **ZIM**: Culture Collection of Industrial Microorganisms, Slovenia; **ATCC**: American Type Culture Collection, USA; **VKM**: All-Russian Collection of Microrganisms, Russia; **MUCL**: (Agro)Industrial Fungi & Yeasts Collection, Belgium. b – *Kwoniella mangroviensis MAT* designation from Statzell-Tallman *et al*., 2008; **?** – No mating reaction observed. c – *Kwoniella mangroviensis MAT* designation (between parentheses) according to J.A. Rodrigues and Á. Fonseca (unpublished). d – *Cryptococcus amylolentus MAT* designation from Findley *et al*., 2012. e – *Cryptococcus heveanensis MAT* designation from Metin *et al*., 2010. (T) – Type-strain.

2.4 - DNA Extraction

Genomic DNA was extracted using a phenol/chloroform method. Cells were harvested from liquid MYP or MYP plates, together with 0.2 ml of glass beads (⌀ 230-320 µm) at 13000 rpm for 3 min. The supernatant was discarded and the pellet was stored at -20ºC for 1h. 250 µl of TE-Phenol (phenol saturated with 10 mM Tris-HCl pH 8.0; 1 mM EDTA), 250 µl of Chloroform and 500 µl of Lysis Buffer (100 mM NaCl; 10 mM Tris-HCl pH 8.0; 1 mM EDTA; 2% Trinton X-100; 1% SDS) were added in the respective order. Cells were disrupted on *VRX basic Vibrax* orbital shaker (IKA®, Germany) at 2000 rpm for 20 min, followed by a 25 min centrifugation at 13000 rpm at 4ºC performed on a *Sigma 3-16K* centrifuge (Sigma, Germany). 400 µl of the supernatant were added to 1 ml of Ethanol (100%) and stored at -20ºC for 1h. The sample was then centrifuged for 10 min at 13000 rpm at 4ºC. The supernatant was discarded and 1 ml of Ethanol (70%) was added to the pellet and then centrifuged for 10 min at 13000 rpm at 4ºC. The supernatant was discarded and the pellet was left to dry at room temperature. The pellet was dissolved with 40 µl of TE+RNase (1X TE + 50 µg/mL RNase) and incubated at 55ºC for 15 min. The DNA was then stored at -20ºC.

2.5 - Fosmid Extraction

A fosmid library from K1 (CBS 8507) genome, constructed using a *CopyControl™ Fosmid Library Production Kit* (Epicentre) with pCC1FOS vector hosted in *E. coli* EPI300-T1, was provided by Prof. Álvaro Fonseca. The library was previously probed for *STE20* gene, and a positive clone was identified: 2K2.

Bacterial cells of the 2K2 were grown overnight in *CERTOMAT® IS* (Sartorius, Germany) incubator at 37ºC, and 180 rpm, in 30 ml of LB liquid medium containing 12.5 µg/ml of Chloramphenicol, 50 µg/ml of X-gal and 25 µg/ml of IPTG. LB medium contained 1% (w/v) of Bacto-Tryptone, 0.5% (w/v) of Bacto-yeast extract, 1% (w/v) of NaCl. 30 ml of cells were harvested and the fosmid DNA was extracted with *JETstar The Novel Plasmid Purification System* (Genomed) according to the manufacturer instructions. The fosmid was sequenced and assembled by STABVida (Portugal), using Next Generation sequencing technology. The obtained fosmid sequence was deposited in EMBL (HE997060).

2.6 - PCR Fingerprinting

An oligonucleotide of the minisatellite-specific core sequence of the wild/type phage M13 (5' GAGGGTGGCGGTTCT 3') (Vassart *et al*., 1987) was used as a single primer in the PCR. Amplification reactions were performed in volumes of 25 µl containing 1x DreamTaq™ Buffer (Fermentas), 3 mM of $MgCl_2$, 0.2 mM dNTPs (Fermentas), 0.4 µM M13-primer, 0.05 U/µl of DreamTaq™ DNA Polymerase (Fermentas) and 6 ng of DNA in the final solution. PCR amplification was performed in *T1-Thermocycler* (Biometra, Germany). PCR program was as follows: 5 min of initial denaturation at 95ºC, followed by 35 cycles of 30 seconds at 95ºC, 1 min at 50ºC and 1 min at 72ºC and terminating with 7 min of final extension at 72ºC.

Amplification products were separated by electrophoresis in 1.4% agarose gels in 0.5X Tris-borate-EDTA (TBE) buffer for 5h at 70V, in *Wide Mini-Sub Cell GT Cell* (Bio-Rad, USA). The gel was stained, with 1X GelRed (Biotium) diluted in distilled water for 5-7 min followed by destaining in distilled water for 30 min and visualized and photographed using *Molecular Imager Gel Doc XR* (Bio-Rad, USA) and Quantity One v.4.5.0 software (Bio-Rad).

2.7 - Primers and Primer design

Specific and degenerated primers were ordered or designed as needed, and used to amplify and sequence the following loci: ITS (*ITS1, 5.8S, ITS2*), nLSU (D1/D2 fragment), *RPB1*, *RPB2*, *TEF1α*, *MCM7*, *STE20* and *SXI* (*SXI1*+intergenic region+*SXI2)* fragments. Primer information is available in Table 2.2.

Primers were designed or redesigned (as described below) from multiple sequence alignments performed with BioEdit 7.0.9.0 software (Hall, 1990), using the built-in ClustalW version 1.4 (Thompson *et. al*., 1990) with the default parameters. Newly designed primers were analyzed using OligoAnalyzer 3.1 (http://eu.idtdna.com/analyzer/applications/oligoanalyzer/). All the primers used in this study were synthesized by STABVida (Portugal), with a Cartridge purification method.

Table 2.2 - List of primers used in this study.

Primer	Sequence (5' → 3')	Tm (ºC)	Reference	Expected amplicon length (bp)
ITS5	GGA AGT AAA AGT CGT AAC AAG G	51,3	White *et al*., 1990	~ 1500
LR6	CGC CAG TTC TGC TTA CC	53,4	Vilgalys and Hester, 1990	
RPB1_Af	GAR TGY CCD GGD CAY TTY GG	57,8	Stiller and Hall, 1997	~ 860
RPB1_Cr	CCN GCD ATN TCR TTR TCC ATR TA	54,2	Matheny *et al*. 2002	
RPB2_F1	GAY GAY MGW GAT CAY TTY GG	51,9	This study[a]	~ 2100
RPB2_R1	GCR TGR ATC TTR TCR TCY ACC	54,4	This study[b]	
MCM7F	CCA ASG GYC AAY TKC AYA TGC	57,2	L. Sharma (unpublished)	~ 560
MCM7R	CAT AAC MGC VGC TGT CAA ACC	57,1	L. Sharma (unpublished)	
EF1α_F1	TAY AAG TGC GGT GGT ATY GAC A	55,9	This study[c]	~ 850-950
EF1α_F2	AGC YTC STT CAA GTA CGC	53,7	This study	
EF1α_R1	ACG GAC TTG ACT TCR GTR GT	55,8	This study[d]	
SXI2_F1	GCC TCG RTT YCK YTT GTT CTG	56,4	Á. Fonseca (unpublished)	~ 1400-1900
SXI2_F5	TTA CCC AGG TTC GTA CCT ACG	59,8	J.A. Rodrigues (unpublished)	
SXI1_R5	GTG CTC KTT GAC GYT TWG AGG	59,8	J.A. Rodrigues (unpublished)	
STE20_F1	AAC TTY GTY CAT CAR GTI CAT G	54,7	Á. Fonseca (unpublished)	~ 1000
STE20_R1	TCC ATR WAY TCC ATK ACI ACC C	55,1	Á. Fonseca (unpublished)	
STE20_F3	AAY TTY GTB CAY CAR GTI CAY G	56,5	This study[e]	~ 1000
STE20_R3	TCC ATR WAY TCC ATK ACI ACC C	55,1	This study[f]	
STE12aF1	GYG TMT TCT CCG ATY TKC G	53,3	This study	Unknown
STE12aR1	CGM ARA TCG GAG AAK ACR C	53,3	This study	
STE12αF1	TKT TRT TYC GCA ATG GCT G	52,8	This study	
STE12αR1	CAG CCA TTG CGR AAY AAM A	52,8	This study	
STE3aF1	GYC CMG TTT GGT GYG ATA TC	54,9	This study	
STE3aR1	GAT ATC RCA CCA AAC KGG RC	54,9	This study	
STE3αF1	CGT CGC CTA GTG ATG ATY GC	57,0	This study	
STE3αR1	GCR ATC ATC ACT AGG CGA CG	57,0	This study	

a – Redesigned based on Liu *et al*., 1999 (RPB2-5f); b – Redesigned based on Liu *et al*., 1999 (RBP2-11aR); c – Redesigned based on AFTOL (EF1α1F); d – Redesigned based on AFTOL (EF1α1R); e – Redesigned based on STE20_F1 primer; f – Redesigned based on STE20_R1 primer. F – Forward; R – Reverse.

RPB2_F1 and RPB2_R2 primers were redesigned based on RPB2-5f and RBP2-11aR primers (Liu *et al*., 1999), respectively, and on *RPB2* sequences from the following strains: *Cryptococcus neoformans* var. *neoformans* JEC21 (AE017344), *Cryptococcus neoformans* var. *neoformans* B-3501A (XM_770778), *Cryptococcus gattii* WM276 (CP000289), *Cryptococcus neoformans* var. *grubii* H99 (JGI) and *Tremella mesenterica* ATCC 24925 (JGI).

EF1α_F1 and EF1α_R1 primers were redesigned based on EF1α1F and EF1α1R primers (AFTOL), respectively, and on *TEF1α* sequences from the following strains: *Cryptococcus neoformans* var. *neoformans* JEC21 (AE017353), *Cryptococcus neoformans* var. *neoformans* B-3501A (NC_009189), *Cryptococcus gattii* WM276 (CP000298), *Cryptococcus neoformans* var. *grubii* H99 (JGI) and *Tremella mesenterica* ATCC 24925 (JGI).

EF1α_F2 primer was designed based on the newly obtained *TEF1α* sequences from the following strains: K1, K9, K10, K14, K17, K22, K25, K28, K29, *Cryptococcus dejecticola* CBS 10117 and *Cryptococcus heveanensis* CBS 569.

STE20_F3 and STE20_R3 primers were redesigned based on STE20_F1 and STE20_R1 primers, respectively, and on the *STE20* sequences from the following strains: *Tremella mesenterica* ATCC 24925 (HM440938), *Tsuchiyaea wingfieldii* CBS 7118 (HM368524), *Cryptococcus amylolentus* CBS 6039 (HM640221), *Cryptococcus neoformans* var. *neoformans* JEC20 (AF542530), *Cryptococcus neoformans* var. *neoformans* JEC21 (AF542531), *Cryptococcus neoformans* var. *grubii* 125.91 (AF542528), *Cryptococcus neoformans* var. *grubii* H99 (AF542529), *Cryptococcus gattii* E566 (AY710429), *Cryptococcus gattii* WM276 (AY710430), *Cryptococcus heveanensis* CBS 569 (GU205379), *Filobasidiella depauperata* CBS 7855 (GU131349).

STE3αF1 and STE3αR1 primers are reverse complement of each other and were designed based on *STE3* sequences from the following strains: *Cryptococcus neoformans* var. *neoformans* JEC21 (AF542531), *Cryptococcus neoformans* var. *grubii* H99 (AF542529), *Cryptococcus gattii* WM276 (AY710430), *Cryptococcus heveanensis* CBS 569 (GU205379).

STE3aF1 and STE3aR1 primers are reverse complement of each other and were designed based on *STE3* sequences from the following strains: *Cryptococcus neoformans* var. *neoformans* JEC20 (AF542530), *Cryptococcus neoformans* var. *grubii* 125.91 (AF542528), *Cryptococcus gattii* E566 (AY710429), *Cryptococcus heveanensis* CBS BCC8398 (GU129940).

STE12αF1 and STE12αR1 primers are reverse complement of each other and were designed based on *STE12* sequences from the following strains: *Cryptococcus neoformans* var. *neoformans* JEC21 (AF542531), *Cryptococcus neoformans* var. *grubii* H99 (AF542529), *Cryptococcus gattii* WM276 (AY710430), *Cryptococcus heveanensis* CBS 569 (GU205379).

STE12aF1 and STE12aR1 primers are reverse complement of each other and were designed based on *STE12* sequences from the following strains: *Cryptococcus neoformans* var. *neoformans* JEC20 (AF542530), *Cryptococcus neoformans* var. *grubii* 125.91 (AF542528), *Cryptococcus gattii* E566 (AY710429), *Cryptococcus heveanensis* CBS BCC8398 (GU129940).

2.8 - PCR amplification

Each PCR reaction contained 1x DreamTaq™ Buffer (Fermentas), 0.2 mM dNTPs (Fermentas), 0.4 µM of each primer, 0.04 U/µl of DreamTaq™ DNA Polymerase (Fermentas) and 8 ng of DNA in a total volume of 50 µl. Additionally for ITS and nLSU regions of the rRNA, comprising ITS1, *5.8S* gene ITS2 and D1/D2 fragment from LSU, 0.2 µM of each primer was used and for *RPB1*, 2.0 µM of each primer and 10 ng of DNA was used instead.

PCR amplifications were performed in one of the following thermocyclers: *T1-Thermocycler* (Biometra, Germany); *Uno II* (Biometra, Germany); *TGradient* (Biometra, Germany) or *S1000™ Thermal Cycler* (Bio-Rad, USA). PCR amplification details are given in Table 2.3.

Table 2.3 - PCR amplification conditions for each genomic locus.

Strains	Primers	PCR Program
LSU + ITS		
K3, K14, K20, K22, K23, K24, K25, K26, K27, K28, K29, Cp2, Cp3, Cp4	Fw: ITS5 Rv: LR6	5 min of initial denaturation at 95ºC, followed by 35 cycles of 30 sec at 95ºC, 30 sec at 50ºC and 2 min at 72ºC and terminating with 7 min of final extension at 72ºC
RPB1		
K4, K6, K27, K9, K10, K14, K17, K22, K25, K27, K28, K29, Cp1, Cp2, Cb, Cd, Ch, Cs	Fw: RPB1_Af Rv: RPB1_Cr	2 min of initial denaturation at 96ºC, followed by 35 cycles of 20 seconds at 96ºC, 40 sec at 52ºC and 1 min at 72ºC and terminating with 7 min of final extension at 72ºC
RPB2		
K1, K2, K4, K6, K9, K10, K14, K17, K22, K25, K27, K28, K29, Cp1, Cp2, Ch, Cs	Fw: RPB2_F1 Rv: RPB2_R1	4 min of initial denaturation at 95ºC, followed by 35 cycles of 1 min at 95ºC, 1 min at 50ºC, an increase of 0.3ºC/sec to 72ºC and 1 min at 72ºC and terminating with 10 min of final extension at 72ºC
TEF1α		
K2, K4, K6, K27, Cp1, Cp2	Fw: EF1α_F2 Rv: EF1α_R1	4 min of initial denaturation at 95ºC, followed by the first stage of 20 cycles of 1 min at 95ºC, 1 min at 60ºC with a decreasing of 0.5ºC/cycle and 1 min at 72ºC, followed by the second stage with 35 cycles of 1 min at 95ºC, 1 min at 50ºC and 1 min at 72ºC and terminating with 10 min of final extension at 72ºC
Cb, Cd, Cs, Bd	Fw: EF1α_F1 Rv: EF1α_R1	
K9, K10, K14, K17, K22, K25, K28, K29	Fw: EF1α_F1 Rv: EF1α_R1	4 min of initial denaturation at 95ºC, followed by the first stage of 20 cycles of 1 min at 95ºC, 1 min at 64ºC with a decreasing of 0.5ºC/cycle and 1 min at 72ºC, followed by the second stage with 35 cycles of 1 min at 95ºC, 1 min at 54ºC and 1 min at 72ºC and terminating with 10 min of final extension at 72ºC
Ch	Fw: EF1α_F1 Rv: EF1α_R1	5 min of initial denaturation at 95ºC, followed by 35 cycles of 1 min at 95ºC, 1 min at 55ºC and 1 min at 72ºC and terminating with 10 min of final extension at 72ºC
MCM7		
K1, K2, K4, K6, K9, K10, K14, K17, K22, K25, K27, K28, K29, Cb, Cd, Cp1, Cp2	Fw: MCM7F Rv: MCM7R	7 min of initial denaturation at 94ºC, followed by 40 cycles of 45 sec at 94ºC, 50 sec at 52ºC and 1 min at 72ºC and terminating with 5 min of final extension at 72ºC
Bd		7 min of initial denaturation at 94ºC, followed by 40 cycles of 45 sec at 94ºC, 50 sec at 45ºC and 1 min at 72ºC and terminating with 5 min of final extension at 72ºC
Cs		7 min of initial denaturation at 94ºC, followed by the first stage of 10 cycles of 45 sec at 94ºC, 50 sec at 60ºC with a decreasing of 0.5ºC/cycle and 1 min at 72ºC, followed by the second stage with 40 cycles of 45 sec at 94ºC, 50 sec at 50ºC and 1 min at 72ºC and terminating with 5 min of final extension at 72ºC
Ca		7 min of initial denaturation at 94ºC, followed by the first stage of 30 cycles of 45 sec at 94ºC, 50 sec at 60ºC with a decreasing of 0.4ºC/cycle and 1 min at 72ºC, followed by the second stage with 40 cycles of 45 sec at 94ºC, 50 sec at 48ºC and 1 min at 72ºC and terminating with 5 min of final extension at 72ºC
STE20		
K2, K8, K9, K10, K11, K14, K16, K17, K20, K21, K22, K23, K24, K25, K26, K27, K28, K29, Cb, Cp1, Cp2, Cp3, Cp4, Ca2	Fw: STE20_F1 Rv: STE20_R1	5 min of initial denaturation at 95ºC, followed by 35 cycles of 30 sec at 95ºC, 30 sec at 50ºC and 2 min at 72ºC and terminating with 7 min of final extension at 72ºC

…/…

Table 2.3 (cont.) - PCR amplification conditions for each genomic locus.

Strains	Primers	PCR Program
SXI1 + SXI2		
K3, K5, K8, K9, K10, K11, K14, K16, K17, K20, K21, K22, K23, K24, K25, K26, K28	Fw: SXI_F5 Rv: SXI_R5	5 min of initial denaturation at 95ºC, followed by 35 cycles of 30 sec at 95ºC, 30 sec at 55ºC and 2 min at 72ºC and terminating with 7 min of final extension at 72ºC
K27, K29, Cd	Fw: SXI_F1 Rv: SXI_R5	5 min of initial denaturation at 95ºC, followed by 35 cycles of 30 sec at 95ºC, 30 sec at 50ºC and 2 min at 72ºC and terminating with 7 min of final extension at 72ºC
Cb	Fw: SXI_F5 Rv: SXI_R5	
STE3 + STE12*		
K1, K3, K10, K11, K25, K28, K29	Fw: STE3α_F1 Rv: STE12a_R1	5 min of initial denaturation at 95ºC, followed by 35 cycles of 30 sec at 95ºC, 30 sec at 50ºC and 2 min at 72ºC and terminating with 7 min of final extension at 72ºC

*Several primer combinations and PCR programs were attempted; however this region was not possible to amplify for K2, K4, K9, K14 and K16.

2.9 - Electrophoresis

Amplification products were separated by electrophoresis in 1.0% agarose gels in 1X Tris-acetate-EDTA (TAE) buffer with a concentration of 0.25X GelRed (Biotium) for 50min at 80V in *Wide Mini-Sub Cell GT Cell* (Bio-Rad, USA). Gels were visualized and photographed using *Molecular Imager Gel Doc XR* (Bio-Rad, USA) and Quantity One v.4.5.0 software (Bio-Rad)

2.10 - Amplification Product Purification and Sequencing

Amplification products were purified using *illustra™ GFX™ PCR DNA and Gel Band Purification Kit* (GE Healthcare), according to the manufacturer's instructions. When unspecific amplification was observed, the specific band, based on size predictions, was cut from the gel and then purified using the same Kit.

All amplification products were sequenced by STABVida (Portugal). Primers used for PCR amplification were also used for sequencing. Additionally, for ITS and LSU, the following inner primers were used for sequencing, respectively: ITS1: TCCGTAGGTGAACCTGCGG (White et al, 1990), ITS4: TCCTCCGCTTATTGATATGC (White et al, 1990), NL1: GCATATCAATAAGCGGAGGAAAAG (O'Donnell, 1993) and NL4: GGTCCGTGTTTCAAGACGG (O'Donnell, 1993). All the sequences generated in this study were deposited in EMBL. The respective accession numbers are listed in Table 3.1 (chapter 3).

ITS sequences from K9, K10, K11, K16, K17, K20, K21, K22, K23, K24 and LSU sequences from K9, K10, K11, K16, K17, K21, were provided by Dr. Deborah Springer (Department of Molecular Genetics and Microbiology, Duke University, USA).

2.11 - Sequence data analysis

Annotations of the different genomic loci were done manually by comparing the newly obtained sequences with the annotated sequences of *C. neoformans* var. *neoformans* JEC21 and *C. heveanensis* CBS 569. The comparison was done by pairwise alignment (BioEdit 7.0.9.0) and BLASTn, BLASTp BLASTx and tBLASTx (2.2.26 version) (Altschul et al., 1997) searches.

Multiple sequence alignments were performed with the genomic sequences using either MUSCLES (Edgar, 2004) or ClustalW built-in version of MEGA 5.05 (Tamura *et al*., 2011) with default parameters. Independent alignments and phylogenetic analyses were performed for each locus. For MLST, independent alignments were concatenated manually for each strain. Phylogenetic relationships were inferred either by Maximum Likelihood (ML) or Neighbor-Joining (NJ). ML method was based on the General Time Reversible (GTR) model (Tavaré, 1986) from 1000 bootstrap replicates, using Gamma distributed of 5 discrete gamma categories, all sites of the data set, all the codon positions and noncoding sites and Nearest-Neighbor-Interchange (NNI) as the ML Heuristic Method in MEGA 5.05. NJ method was based on Jukes-Cantor model (Jukes and Cantor, 1969), from 1000 bootstrap replicates, using Gamma Distributed parameter of 1 and pairwise deletion in MEGA 5.05. Jukes-Cantor Distances were calculated in MEGA 5.05 by computing Pairwise Distances, based on Jukes-Cantor model, using the same parameters as NJ method analysis.

Sequences from GenBank were used when available. Accession numbers of sequences used for *MAT* and MLST analyses are listed in Table 2.4 and Table 2.5, respectively. Sequences from *Cryptococcus neoformans* var. *neoformans* JEC20 (*MAT***a**) and JEC21 (*MAT*α), *Cryptococcus neoformans* var. *grubii* 125.91 (*MAT***a**) and H99 (*MAT*α), *Cryptococcus gattii* E566 (*MAT***a**) and WM276 (*MAT*α) and *Cryptococcus heveanensis* CBS 569 (A1B1) and BCC 8398 (A2B2) and *Filobasidiella depauperata* (Fd) were all retrieved either from GenBank or DOE Joint Genome Institute (JGI), while all sequences from *Tremella mesenterica* ATCC 24925 were retrieved from JGI, and therefore, no accession number is provided.

Table 2.4 - GenBank accession numbers of sequences of the listed loci used for *MAT* analyses.

Strain	*SXI1*	*SXI2*	*STE20*	*BSP3*	*CNB00600*	*CNB00610*	*CND05260*
Ch	GU129941	GU129941	GU205379	GU205379	GU205379	GU205379	GU205379
Ch2	GU129944	GU129944	–	–	GU129940	–	–
Ca	HM640220	HM640220	HM640221	HM640221	–	–	–
Ca2	HM640224	HM640225	–	–	–	–	–
Tw	HM368525	HM368525	HM368524	HM368524	–	–	–
JEC20	–	AF542530	AF542530	AF542530	–	–	–
JEC21	AF542531	–	AF542531	AF542531	AE017342	AE017342	AE017344
125.91	–	AF542528	AF542528	AF542528	–	–	–
H99	AF542529	–	AF542529	AF542529	–	–	–
E566	–	AY710429	AY710429	AY710429	–	–	–
WM276	AY710430	–	AY710430	AY710430	CP000287	CP000287	CP000294
Fd	–	–	GU131349	–	–	–	–

…/…

Table 2.4 (cont.) - GenBank accession numbers of sequences of the listed loci used for *MAT* analyses.

Strain	*CNE02670*	*CNF01610*	*CNG04540*	*CNI00160*	*MF*	*STE3*	*STE12*
Ch	GU205379	GU205379	GU205379	GU205379	GU205379	GU205379	GU205379
Ch2	–	–	GU129940	–	GU129940	–	GU129940
Ca	HM640222	–	–	–	HM640223	HM640222	HM640222
Ca2	–	–	–	–	–	–	–
Tw	HM368524	–	–	–	HM368524	HM368524	HM368524
JEC20	–	–	–	–	AF542530	AF542530	AF542530
JEC21	AE017345	AE017346	AE017347	NC_006694	AF542531	AF542531	AF542531
125.91	–	–	–	–	AF542528	AF542528	AF542528
H99	–	–	–	–	AF542529	AF542529	AF542529
E566	–	–	–	–	AY710429	AY710429	AY710429
WM276	CP000290	CP000291	NC_014944	CP000293	AY710430	AY710430	AY710430
Fd	–	–	–	–	–	–	–

Table 2.5 - GenBank accession numbers of sequences of listed the loci used for MLST analyses.

Strain	ITS	D1D2	*RPB1*	*RPB2*	*TEF1α*	*MCM7*
K1	FJ534882	FJ534912	FJ534928	–	–	–
K2	EF174040	EF174033	–	–	–	–
K3	–	*	–	–	–	–
K4	EF174038	EF174031	–	–	–	–
K5	EF174039	EF174032	–	–	–	–
K6	EF174041	EF174034	–	–	–	–
K7	**	*	–	–	–	–
K8	***	*	–	–	–	–
K25	–	AY167602	–	–	–	–
K26	–	AY296055	–	–	–	–
K29	–	AM748530	–	–	–	–
Bd	FJ534871	AF189870	FJ534917	FJ534933	FJ534856	–
Ca	FJ534872	FJ534902	EF211452	FJ534934	FJ534857	–
Ca2	JN019831	–	–	–	–	–
Cb	FJ534882	FJ534903	–	FJ534935	–	–
Cd	FJ534874	FJ534904	–	FJ534936	–	–
Ch	GU585748	FJ534905	–	–	–	–
Cp1	EF672246	EF672245	–	–	–	–
Cs	FM212571	FM212443	–	–	–	–
Tw	FJ534886	FJ534916	EF211454	FJ534947	FJ534870	–
JEC20	EF211189	–	EF211397	–	–	–
JEC21	FJ534880	FJ534910	EF211398	AE017344	AE017353	AE017346
125.91	EF211129	–	EF211337	–	–	–
H99	AY217027	JGI	EF211354	JGI	JGI	JGI
E566	EF211206	–	EF211414	–	–	–
WM276	EF211211	CP000287	NC_014942	CP000289	CP000298	CP000291
Fd	FJ534881	FJ534911	EF211456	FJ534942	FJ534865	–

Additional sequence information from Statzell-Tallman *et al.* (2008): * Sequence identical to K1 (FJ534912); ** Sequence identical to K6 (EF174041); *** Sequence identical to K1 (FJ534882).

3 - RESULTS

3.1 - Molecular identification and phylogenetic diversity in *Kwoniella mangroviensis*

The *Kwoniella mangroviensis* yeast cultures used in this study, were firstly analyzed to validate their species assignment on the basis of nLSU rRNA gene (D1/D2 domains) sequences (LSU). Newly obtained LSU sequences are listed in Table 3.1. According to the results of nBLAST searches and phylogenetic analysis, all *Kwoniella mangroviensis* strains were closely related to the *Kwoniella mangroviensis* type-strain (Figure 3.1). The next closest species to *K. mangroviensis* was *Cryptococcus bestiolae*. In the Kwoniella clade, there were also *Cryptococcus dejecticola*, *Cryptococcus pinus*, *Bullera dendrophila* and *Cryptococcus shivajii*. Although *C. pinus* strains were from distinct geographic origins (Russia and Belgium), LSU sequences were identical. *Cryptococcus heveanensis* had an isolated position, in this analysis. Whether or not *Cryptococcus cuniculi*, *Kwoniella shandongensis* and *Cryptococcus tronadorensis*, belong to the Kwoniella or Filobasidiella clades was not clear due to low bootstrap support. Several undescribed species clustered among these three latter species with low bootstrap support. Further studies are required to clarify the phylogenetic position of *C. cuniculi, K. shandongensis, C. tronadorensis* and some of these undescribed species.

In order to improve intraspecific discrimination within *K. mangroviensis*, ITS sequences were determined (Table 3.1). A phylogenetic analysis of concatenated ITS (ITS1 + 5.8S rRNA gene + ITS2) and LSU sequences (Figure 3.2), revealed three different, but closely related phylogenetic clades: one clade (highlighted in blue on Figure 3.2) included the Botswana strains, the second clade (green) included strains from Europe (Spain, Switzerland and Slovenia) and USA, and the third clade (red) included the strains from the Caribbean (British Virgin Islands and Bahamas). When the sequences were compared among all *K. mangroviensis* strains, it was found that ITS sequences were more divergent than LSU, namely between strains from the three different phylogenetic clades (with five to nine nucleotide substitutions) (Table 3.2). Among LSU sequences, there were no major differences, and most of the strains had identical sequences (Table 3.2). When the phylogenetic tree based on ITS and LSU was analyzed for the Kwoniella and Filobasidiella clades, *C. bestiolae* was the closest species to *K. mangroviensis*. *C. pinus* origins were now reflected in the analysis, with a small divergence between the Russian type-strain and the Belgium strains (four nucleotide substitutions). *C. dejecticola* appeared to be the closest species *to C. pinus*, based on this analysis. *C. heveanensis* clustered with the Kwoniella clade, but with no significant bootstrap support. *K. shandongensis, C. cuniculi* and *C. tronadorensis* clustered into a single clade, but once again it was unclear to which main clade they may belong to. The two *C. neoformans* varieties clustered apart with *C. gattii* closer to *C. neof. neoformans*, however with an unreliable bootstrap support. Both varieties differed by six substitutions in LSU+ITS sequences, whereas *C. neoformans* and *C. gattii* differed by 13 substitutions.

Table 3.1 - Accession numbers of nucleotide sequences of Kwoniella mangroviensis and related species.

Strain	ITS	LSU	*RPB1*	*RPB2*	*TEF1α*	*MCM7*	*STE20*	*SXI*
K1	–	–	–	–	–	HE996985	HE997013[a]	HE997032[a]
K2	–	–	HE997000[a]	HE996974	HE997047	HE996986	HE997014[a]	HE997033[a,b]
K3	HE984337	–	–	–	–	–	HE997015[b]	HE997034[b]
K4	–	–	HE997001	HE996975	HE997048	HE996987	HE997016[b]	HE997035[b]
K5	–	–	–	–	–	–	HE997017[b]	HE997036[b]
K6	–	–	HE997002	HE996976	HE997049	HE996988	HE997018[b]	HE997037[b]
K7	–	–	–	–	–	–	HE997019[b]	HE997038[b]
K8	–	–	–	–	–	–	HE997020	HE997039
K9	HF545755[c]	HF545768[c]	HF545792	HF545812	HF545826	HF545818	HF545780	HF545798
K10	HF545756[c]	HF545769[c]	HF545793	HF545813	HF545827	HF545819	HF545781	HF545799
K11	HF545757[c]	HF545770[c]	–	–	–	–	HF545782	HF545800
K14	HF545758	HF545771	HF545794	HF545814	HF545828	HF545820	HF545783	HF545801
K16	HF545759[c]	HF545772[c]	–	–	–	–	HF545784	HF545802
K17	HF545760[c]	HF545773[c]	HF545795	HF545815	HF545829	HF545821	HF545785	HF545803
K20	HF545761[c]	HF545774	–	–	–	–	HF545786	HF545804
K21	HF545762[c]	HF545775[c]	–	–	–	–	HF545787	HF545805
K22	HF545763[c]	HF545776	HF545796	HF545816	HF545830	HF545822	HF545788	HF545806
K23	HF545764[c]	HF545777	–	–	–	–	HF545789	HF545807
K24	HF545765[c]	HF545778	–	–	–	–	HF545790	HF545808
K25	HE984339	–	HE997004	HE996978	HE997051	HE996990	HE997022	HE997041
K26	HE984340	–	–	–	–	–	HE997023	HE997042
K27	HE984338	HE996973	HE997003	HE996977	HE997050	HE996989	HE997021	HE997040
K28	HE984341	HE996972	HE997005	HE996979	HE997052	HE996991	HE997024	HE997043
K29	HE984342	–	HE997006	HE996980	HE997053	HE996992	HE997025	HE997044
Bd	–	–	–	–	–	HE996997	NA	NA
Ca	–	–	–	–	–	HE996998	–	–
Ca2	–	NA	NA	NA	NA	NA	HE997031	–
Cb	–	–	HE997007	–	HE997054	HE996993	HE997026	HE997045
Cd	–	–	HE997008	–	HE997055	HE996994	NA	HE997046
Ch1	–	–	HE997012	HE996983	HE997058	NA	–	–
Cp1	–	–	HE997009	HE996981	HE997056	HE996995	HE997027	NA
Cp2	HE984343	HE984343	HE997010	HE996982	HE997057	HE996996	HE997028	NA
Cp3	HE984344	HE984344	–	–	–	–	HE997029	NA
Cp4	HE984345	HE984345	–	–	–	–	HE997030	NA
Cs	–	–	HE997011	HE996984	HE997059	HE996999	NA	NA

a – Sequence determined by Á. Fonseca; b – Sequence determined by J.A. Rodrigues; c – Sequence determined by D. Springer; NA – Sequences not available. Sequences from previous studies are represented by a dash (see Tables 2.4 and 2.5).

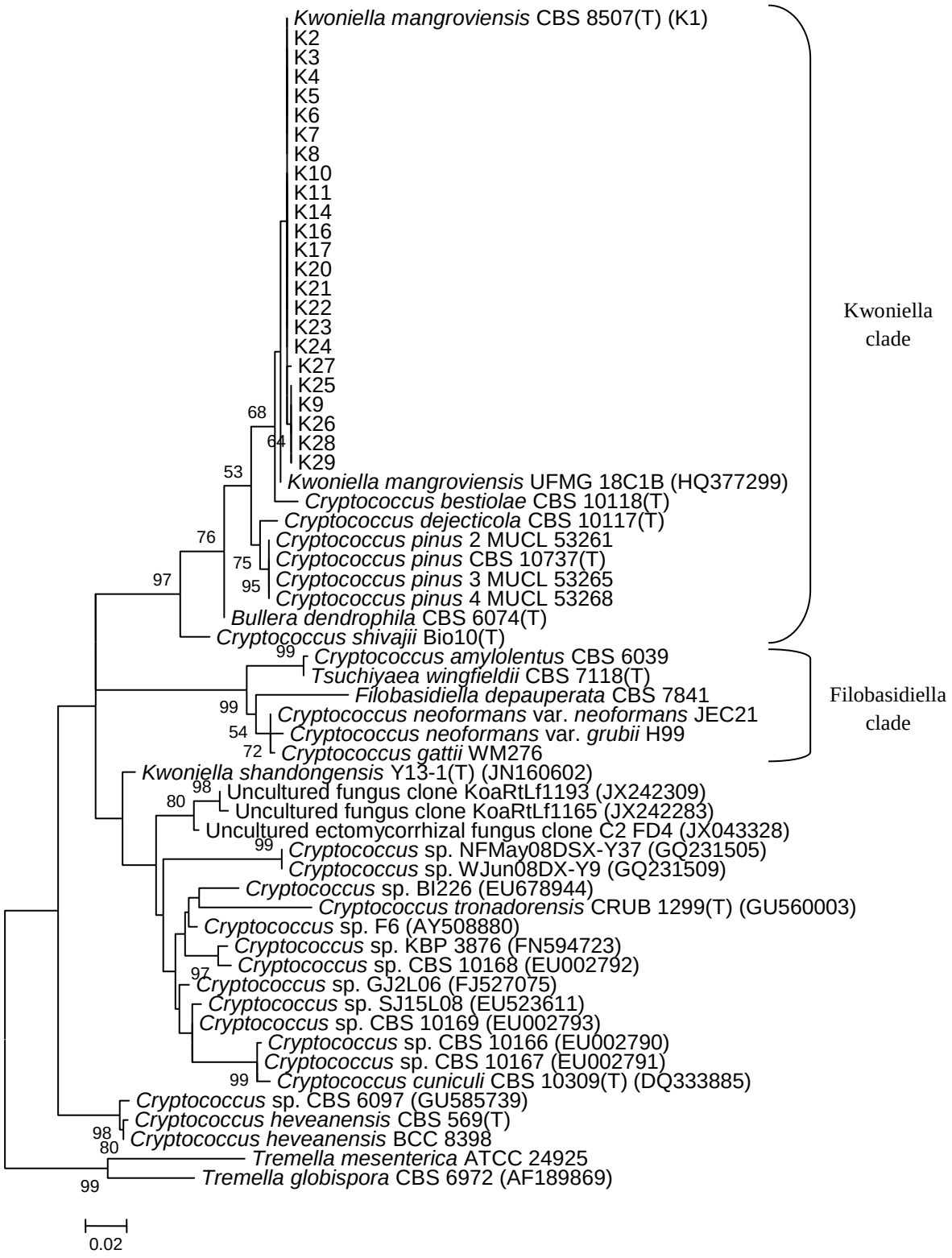


Figure 3.1 - Phylogenetic tree of *K. mangroviensis* strains and related species in the Kwoniella and Filobasidiella clades based on LSU sequences. The tree was constructed using Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) for 1000 replicates. *Tremella mesenterica* and *Tremella globispora* were used as outgroups in the analysis. (T) represents type-strain. Sequences determined in this study are listed in Table 3.1; additional sequences were retrieved from GenBank or from JGI (see also Table 2.5).

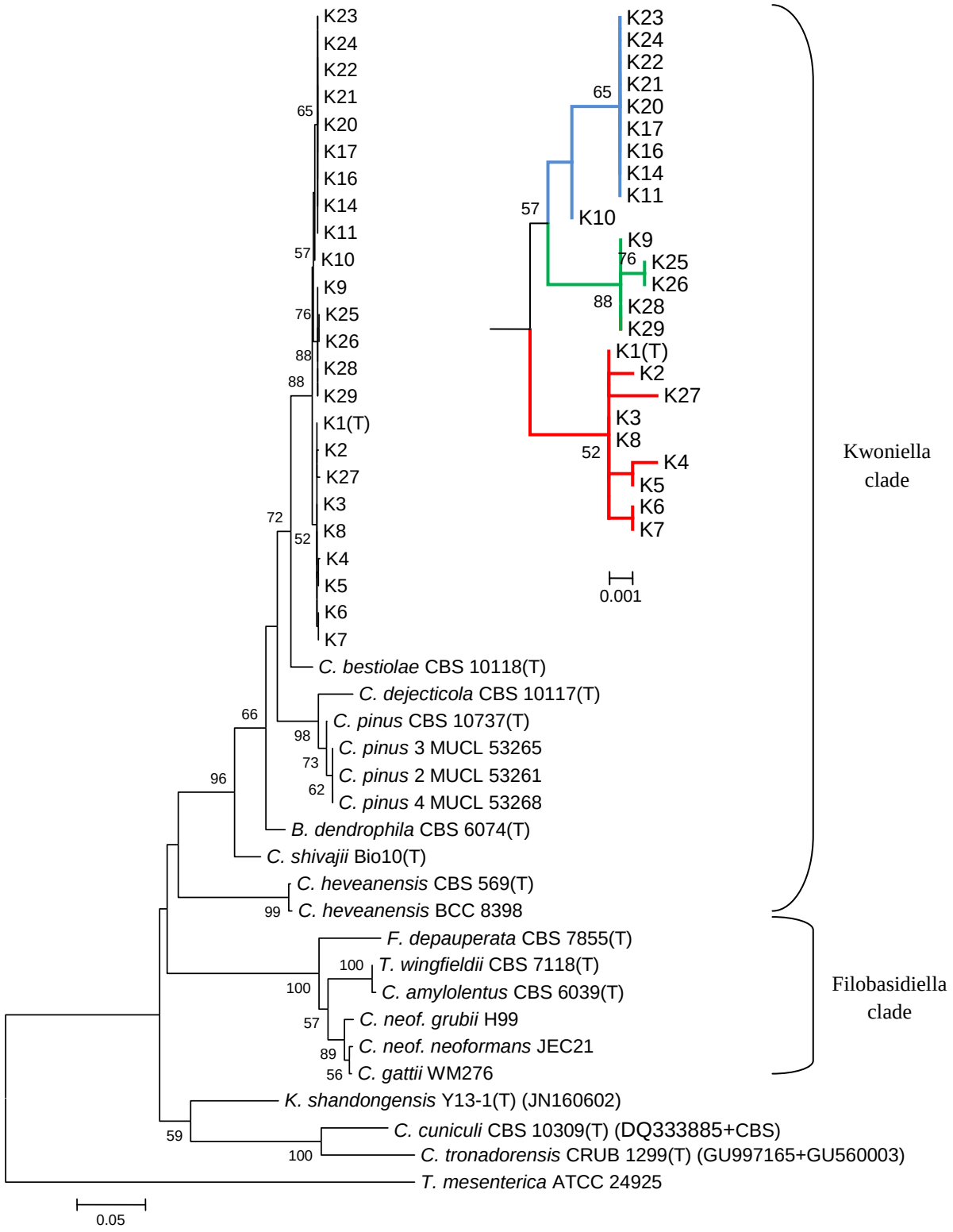


Figure 3.2 - Phylogenetic tree of *K. mangroviensis* strains and related species in the Kwoniella and Filobasidiella clades based on concatenated LSU and ITS sequences. The tree was constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Insert: Detailed view of the *K. mangroviensis* clades drawn to a different scale. Sequences determined in this study are listed in Table 3.1; additional sequences were retrieved from GenBank, CBS database or from JGI (see also Table 2.5).

Table 3.2 - Nucleotide substitutions in ITS (top right) and LSU (bottom left) sequences of *K. mangroviensis* strains.

	K1	K2	K3	K4	K5	K6	K7	K8	K27	K10	K11	K14	K16	K17	K20	K21	K22	K23	K24	K9	K25	K26	K28	K29
K1	ID	2	1	3	2	2	2	0	2	6	8	8	8	8	8	8	8	8	8	7	8	8	7	7
K2	0	ID	1	3	2	2	2	2	2	6	8	8	8	8	8	8	8	8	8	7	8	8	7	7
K3	0	0	ID	2	1	1	1	1	1	5	7	7	7	7	7	7	7	7	7	6	7	7	6	6
K4	0	0	0	ID	1	3	3	3	3	7	9	9	9	9	9	9	9	9	9	8	9	9	8	8
K5	0	0	0	0	ID	2	2	2	2	6	8	8	8	8	8	8	8	8	8	7	8	8	7	7
K6	0	0	0	0	0	ID	0	2	2	6	8	8	8	8	8	8	8	8	8	7	8	8	7	7
K7	0	0	0	0	0	0	ID	2	2	6	8	8	8	8	8	8	8	8	8	7	8	8	7	7
K8	0	0	0	0	0	0	0	ID	2	6	8	8	8	8	8	8	8	8	8	7	8	8	7	7
K27	1	1	1	1	1	1	1	1	ID	6	8	8	8	8	8	8	8	8	8	7	8	8	7	7
K10	0	0	0	0	0	0	0	0	1	ID	2	2	2	2	2	2	2	2	2	3	4	4	3	3
K11	0	0	0	0	0	0	0	0	1	0	ID	0	0	0	0	0	0	0	0	5	6	6	5	5
K14	0	0	0	0	0	0	0	0	1	0	0	ID	0	0	0	0	0	0	0	5	6	6	5	5
K16	1	1	1	1	1	1	1	1	2	1	1	1	ID	0	0	0	0	0	0	5	6	6	5	5
K17	0	0	0	0	0	0	0	0	1	0	0	0	1	ID	0	0	0	0	0	5	6	6	5	5
K20	0	0	0	0	0	0	0	0	1	0	0	0	1	0	ID	0	0	0	0	5	6	6	5	5
K21	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	ID	0	0	0	5	6	6	5	5
K22	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	ID	0	0	5	6	6	5	5
K23	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	ID	0	5	6	6	5	5
K24	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	ID	5	6	6	5	5
K9	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	ID	1	1	0	0
K25	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	0	ID	0	1	1
K26	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	0	0	ID	1	1
K28	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	0	0	0	ID	0
K29	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	0	0	0	0	ID

3.2 - Mating experiments

To test for sexual compatibility, strains were co-cultured and examined morphologically for production of sexual structures. The results of mating experiments are summarized in Table 3.3 in which strains were, grouped according to the clades in the LSU+ITS tree (Figure 3.2). Strains K1 (A1B1), K2 (A2B2) and K6 (A1B3) (J.A. Rodrigues and Á. Fonseca, unpublished) were used as controls and reference for the mating assays. From the assays performed, seven developed sexual structures (K1×K2, K2×K6, K2×K27, K6×K8, K27×K4, K27×K5 and K27×K8) and one strain (K8) was self-filamentous, as had been found for K4 (A2B1) and K5 (A2B1) (J.A. Rodrigues and Á. Fonseca, unpublished). The sexual structures included, as previously described (Statzell-Tallman *et al*., 2008), dikaryotic mycelium with clamp connections and basidia, while the filaments produced in monoculture lacked clamp connections and basidia. The control assays performed (K1×K2, K1×K6 and K2×K6) ensured that the media and conditions were appropriate for the mating experiments. K8 mated with K6 (A1B3) and K27, while K27 mated with K2 (A2B2), K4 (A2B1) and K5 (A2B1). These mating assays were consistent with the hypothesis of a tetrapolar mating-system in this species since K27 only mated with strains with A2 allele, while K8 mated with A1 strains, suggesting that both strains had different A alleles, which was supported by the observed mating reactions between each other.

Table 3.3 - Summary of mating assays performed with *K. mangroviensis* strains.

	Strain	K1	K2	K4	K5	K6	K8	K10	K11	K14	K16	K9	K25	K28
Caribbean clade	K1	N												
	K2	Y	N											
	K6	N	Y			N								
	K8	N	N			Y	N*							
	K27	N	Y	Y	Y	N	Y							
Botswana clade	K10	N	N			N		N				N		
	K11	N	N			N		N	N			N		
	K14	N	N			N		N	N	N		N		
	K16	N	N			N		N	N	N	N	N		
	K17	N	N			N								
	K20	N	N			N								
	K21	N	N			N								
	K22	N	N			N								
	K23	N	N			N								
	K24	N	N			N								
Eurpean + USA clade	K9	N	N			N						N		
	K25	N	N			N		N	N	N	N	N	N	
	K26	N	N			N		N	N	N	N	N	N	
	K28	N	N			N		N		N		N	N	N
	K29	N	N			N		N		N		N	N	N

Y – Presence of dikaryotic mycelium; N – Absence of dikaryotic mycelium; * – Self-filamentous strain; Empty cells – Test not performed

3.3 - M13 PCR fingerprinting

M13 PCR fingerprinting was performed with all *Kwoniella mangroviensis* strains. Four different profiles were detected based on this molecular typing technique (Figure 3.3). One profile corresponded to strains from the Caribbean (K1, K2, K3, K4, K5, K6, K7, K8 and K27), a unique profile was identified for the clinical isolate (K9), another profile corresponded to strains from Europe (K25, K26, 28 and K29) and a fourth profile corresponded to strains from Botswana (K10, K11, K14, K16, K17, K20, K21, K22, K23 and K24).

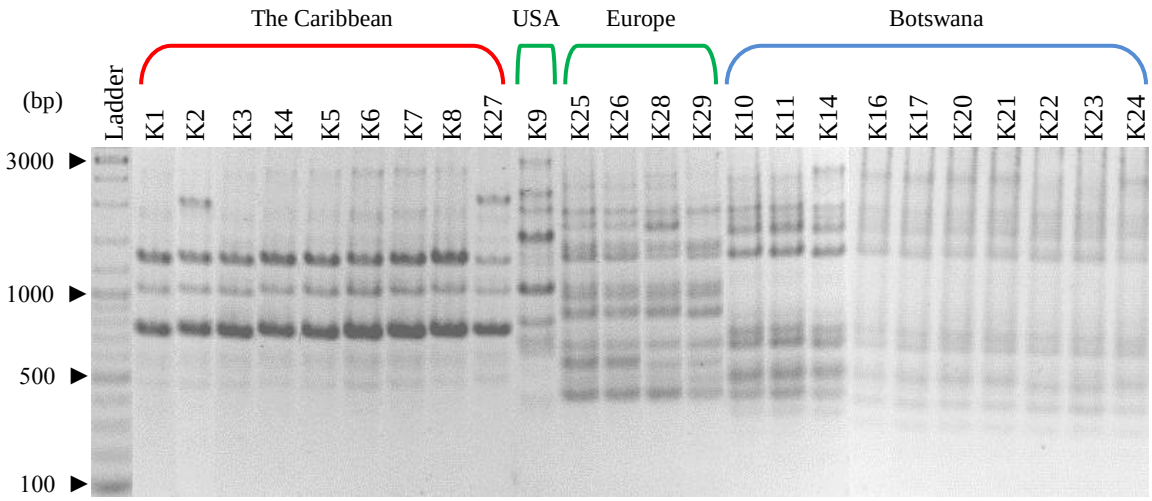


Figure 3.3 - Fingerprinting profiles of the *Kwoniella mangroviensis* isolates used in the study. Fingerprinting of *K. mangroviensis* strains performed with M13 primer. The ladder used was GeneRuler™ DNA Ladder Mix. 1,4% agarose gel in 1X TBE, stained with GelRed.

3.4 - Physiological Tests

Physiological tests were already available for K1 (Statzell-Tallman *et al*., 2008) and K28 (CBS database), therefore, representatives from the other fingerprint profiles (K9 and K10) were selected and tested. K1 and K28 were used as controls and were repeated in two different and independent experiments. Results from the physiological tests are compiled in Table 3.4. No significant differences were observed between the strains, except for the growth temperatures. K9, K10 and K28 grew up to 35ºC, while K1 only grew up to 30ºC.

Table 3.4 - Physiological tests results for *K. mangroviensis* strains.

	K1	K9	K10	K28		K1	K9	K10	K28
Growth at 30ºC	+	+	+	+	Galactitol	+, d, w	d	+	+, d, w
Growth at 35ºC	-	+	+	+	myo-Inositol	-, d, w	+	-	+
Growth at 37ºC	-	-	-	-	D-Glucono-1,5-lactone	v	d	-	v
YM	+	+	+	+	2-Keto-D-Gluconate	+	+	+	+
D-Glucose	+	+	+	+	5-Keto-D-Gluconate	+	+	+	+, d, w
D-Galactose	+	d	+	+	D-Gluconate	+	+	+	+
L-Sorbose	+	d	d	+	D-Glucuronate	+	+	+	+
D-Glucosamine	-	-	-	-	D-Galacturonate	+, d, w	+	+	+
D-Ribose	+	+	+	+	DL-Lactate	-	d, w	d, w	-, d, w
D-Xylose	+	+	+	+	Succinate	+, d, w	+	+	+
L-Arabinose	+	+	+	+, d, w	Citrate	+, d, w	w	+	v
D-Arabinose	+	+	+	+	Methanol	-	-	d, w	-
L-Rhamnose	+	+	+	+	Ethanol	-, d, w	-	-	v
Sucrose	+	+	+	+	L-tartaric acid	-	w	w	v
Maltose	+	+	+	+	D-tartaric acid	-	-	-	-
α,α-Trehalose	+	+	+	+	m-tartaric acid	-	-	-	-
Methyl α-D-Glucoside	-, d, w	d	d	-	Saccharic acid	+, d, w	w	w	v
Cellobiose	+	+	+	+	Mucid acid	d, w	w	+	v
Salicin	d, w	d	d, w	+, d, w	Protocatechuic acid	+	-	-	v
Melibiose	-	-	-	-	Vanillic acid	-, d, w	-	-	-
Lactose	+	d	+	+, d, w	Ferulic acid	-, d, w	-	-	-
Raffinose	+	+	+	+	Veratric acid	-	-	-	-
Melezitose	+	+	+	+	*p*-Hydroxybenzoic acid	v	-	-	v
Inulin	-	-	-	-	m-Hydroxybenzoic acid	v	-	-	-
Starch	+	+	+	+	Gallic acid	v	-	-	v
Glycerol	d, w	d	d	d	Salicylic acid	-, d, w	-	-	-
Erythritol	-, d, w	d, w	-	-	Gentisic acid	-, d, w	-	-	-
Ribitol	+	+	+	+	Catechol	-	-	-	-
Xylitol	d, w	+	+	+	Phenol	-	-	-	-
L-Arabinitol	d, w	+	+	+	L-malic acid	+, d, w	+	+	+, d, w
D-Glucitol	+, d, w	+	+	+	Cycloheximide 0,1%	-	-	-	-, d, w
D-Mannitol	+	+	+	+	Cycloheximide 0,01%	v	-	-	v

+ stands for strong positive reaction, while **w** stands for weak, **-** for negative growth, **d** for delayed growth and **v** for variable (positive/negative) growth from replicates. Results for K1 and K28 are a compilation from two different experiments.

3.5 - Multilocus Phylogenetic Analyses

To further explore the evolutionary relationships within *Kwoniella mangroviensis*, five isolates from the Caribbean (K1, K2, K4, K6 and K27), three isolates from Europe (K25, K28 and K29), one isolate from the USA (K9) and four isolates from Botswana (K10, K14, K17 and K22) were chosen for this analysis, as well as *B. dendrophila*, *C. pinus*, *C. bestiolae*, *C. dejecticola*, *C. shivajii* and *C. heveanensis*, within the Kwoniella clade and *C. neoformans* var. *neoformans*, *C neoformans* var. *grubii*, *C. gattii*, *C. amylolentus*, *T. wingfieldii* and *F. depauperata* within the Filobasidiella clade. Molecular markers, including LSU and ITS (Figure 3.2) and four additionally highly conserved genomic loci (*RPB1, RPB2, TEF1α* and *MCM7*) were sequenced and analyzed. Newly obtained sequences are listed in Table 3.1. Phylogenetic analyses were performed for each locus: *RPB1* (Figure 3.4), *RPB2* (Figure 3.5), *TEF1α* (Figure 3.6) and *MCM7* (Figure 3.7). Nucleotide and amino acid residue substitutions in genomic sequences, coding sequences (CDS) and *in silico* translated protein sequences were then counted and compared between strains, (Tables 3.5 – 3.10). The clades with the *K. mangroviensis* strains were highlighted in the individual phylogenetic trees, according to their previous assignment based on LSU and ITS sequences.

3.5.1 - *RPB1* phylogenetic analysis

Based on the phylogenetic analysis of *RPB1* sequences (Figure 3.4), *K. mangroviensis* grouped into four different clades (visible in the insert subtree displayed next to the main tree), corresponding to the four different fingerprinting profiles (Figure 3.3). The Caribbean clade clustered apart from the others, while K9 clustered closer to the Botswana clade than to the European clade. The *RPB1* fragment analyzed comprises one introns; therefore, the sequences were compared with (genomic) and without the intron (CDS). From the sequence comparisons, most of the nucleotide substitutions were present in the intron, but the differences within the coding sequence still defined the same clades (Figure 3.4); however, only one substitution was not synonymous in K27 and K29, while the other strains encoded the same putative protein (Tables 3.5 and 3.6). All Botswana strains had identical sequences.

Similar to the previous analyses, *C. bestiolae* was the closest species to *K. mangroviensis* and *C. dejecticola* was the closest species to *C. pinus*. *C. heveanensis* was connected to the Kwoniella clade with strong bootstrap support. The species within the Filobasidiella clade clustered with high support, and the observed topology was consistent with the results of Findley *et al*. (2009). The two *C. pinus* strains differed by18 substitutions in the genomic sequence, and by 13 substitutions in the CDS sequence; however, no amino acid residues were substituted. The two *C. neoformans* varieties differed by 47 substitutions in the genomic sequence, by 33 in the CDS sequence and by four amino acid residues; whereas *C. neoformans* and *C. gattii* differed by 56 or 58 substitutions in the genomic sequence, by 35 or 39 in the CDS sequence and by five or six amino acid residues.

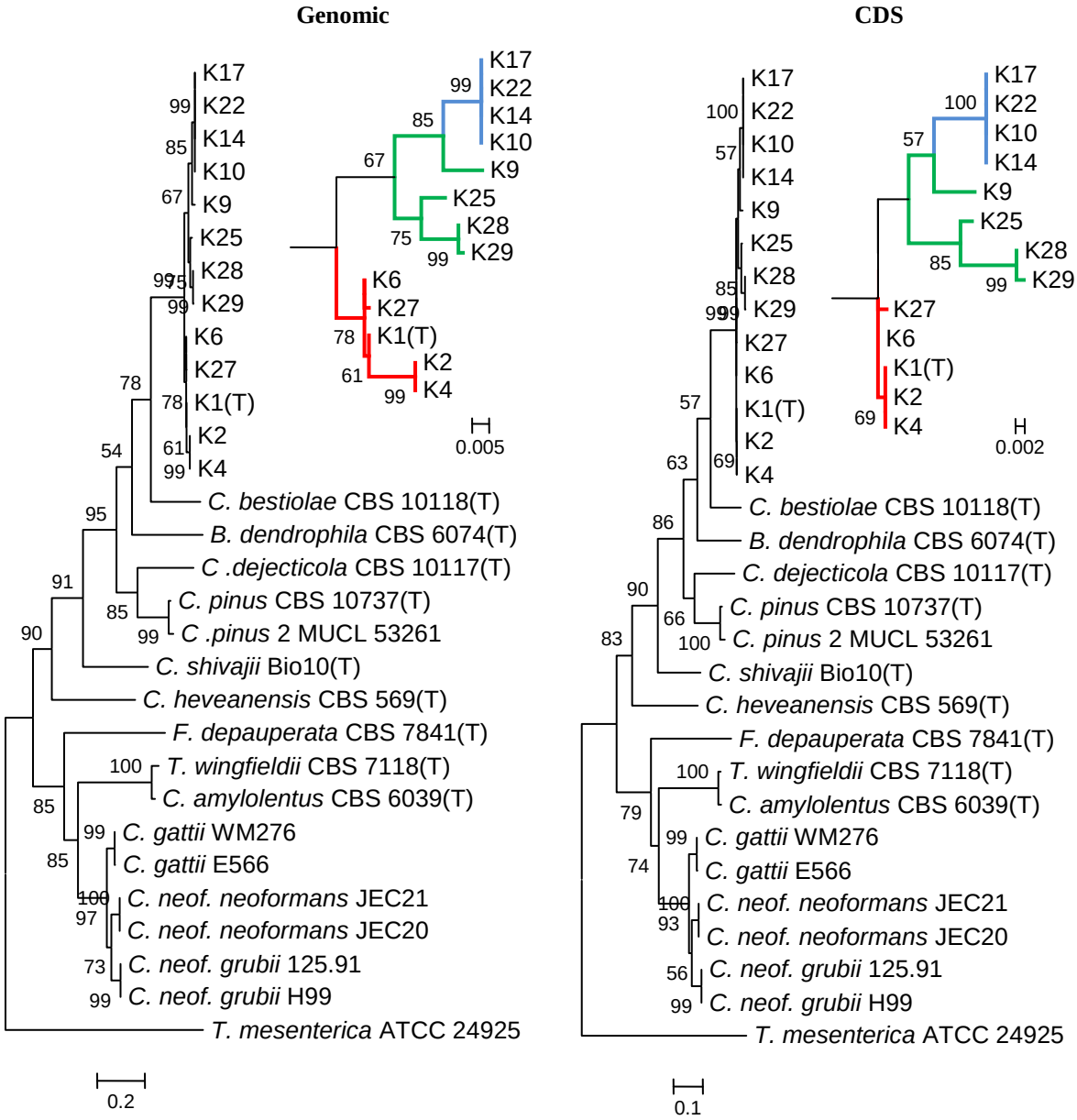


Figure 3.4 - Phylogenetic tree of *K. mangroviensis* strains and related species in the Kwoniella and Filobasidiella clades based on *RPB1* sequences. Phylogenetic relationships of strains in the Kwoniella and Filobasidiella clades based on *RPB1* genomic and coding sequences. Analyses performed with 788 positions (genomic) and 634 positions (CDS) in the final dataset. The tree was constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Insert: Detailed view of the *K. mangroviensis* clades drawn to a different scale. Sequences determined in this study are listed in Table 3.1; additional sequences were retrieved from GenBank or from JGI (see also Table 2.5).

Table 3.5 - Nucleotide substitutions in *RPB1* genomic (top right) and coding sequences (bottom left) of *K. mangroviensis* strains.

	K1(T)	K2	K4	K6	K27	K10	K14	K17	K22	K9	K25	K28	K29
K1(T)	ID	16	16	1	2	38	38	38	38	41	34	33	34
K2	0	ID	0	17	18	51	51	51	51	54	48	49	50
K4	0	0	ID	17	18	51	51	51	51	54	48	49	50
K6	1	1	1	ID	1	39	39	39	39	40	33	34	35
K27	2	2	2	1	ID	40	40	40	40	41	34	35	36
K10	12	12	12	13	14	ID	0	0	0	16	28	27	28
K14	12	12	12	13	14	0	ID	0	0	16	28	27	28
K17	12	12	12	13	14	0	0	ID	0	16	28	27	28
K22	12	12	12	13	14	0	0	0	ID	16	28	27	28
K9	13	13	13	12	13	12	12	12	12	ID	27	26	27
K25	13	13	13	12	13	17	17	17	17	14	ID	14	15
K28	15	15	15	16	17	19	19	19	19	16	9	ID	1
K29	16	16	16	17	18	20	20	20	20	17	10	1	ID

Table 3.6 - Amino acid residue substitutions in *RPB1* translated sequences of *K. mangroviensis* strains.

	K1(T)	K2	K4	K6	K27	K10	K14	K17	K22	K9	K25	K28	K29
K1(T)	ID												
K2	0	ID											
K4	0	0	ID										
K6	0	0	0	ID									
K27	1	1	1	1	ID								
K10	0	0	0	0	1	ID							
K14	0	0	0	0	1	0	ID						
K17	0	0	0	0	1	0	0	ID					
K22	0	0	0	0	1	0	0	0	ID				
K9	0	0	0	0	1	0	0	0	0	ID			
K25	0	0	0	0	1	0	0	0	0	0	ID		
K28	0	0	0	0	1	0	0	0	0	0	0	ID	
K29	1	1	1	1	2	1	1	1	1	1	1	1	ID

3.5.2 - *RPB2* phylogenetic analysis

The results from the phylogenetic analysis of *RPB2* (Figure 3.5) were consistent with the results from *RPB1*. *K. mangroviensis* clustered into the same four different clades, comprising the same strains, but in this analysis, the Botswana and the European clades clustered closer and K9 clustered more divergently. The *RPB2* fragment analyzed does not comprise any intron, therefore, the whole sequence is coding sequence (CDS). From sequence comparisons, there were no nucleotide substitutions among strains from the same clades (except for one substitution within the Caribbean clade); however, many nucleotide substitutions were detected between clades (between 17 and 52 nucleotide substitutions) (Table 3.7). All these nucleotide substitutions were synonymous in all strains, encoding the same putative protein sequence (Table 3.7).

The bootstrap support for the other species was lower than in the *RPB1* analysis. *B. dendrophila* replaced *C. bestiolae* as the closest species to *K. mangroviensis*, but with no bootstrap support. Among the *C. pinus* strains there was no divergence in this locus. The two *C. neoformans* varieties differed by 29 synonymous substitutions in the CDS sequence; whereas *C. neoformans* and *C. gattii* differed by 53 substitutions in the CDS sequence and by one amino acid residue. *C. heveanensis* clustered with the Kwoniella clade with moderate bootstrap support.

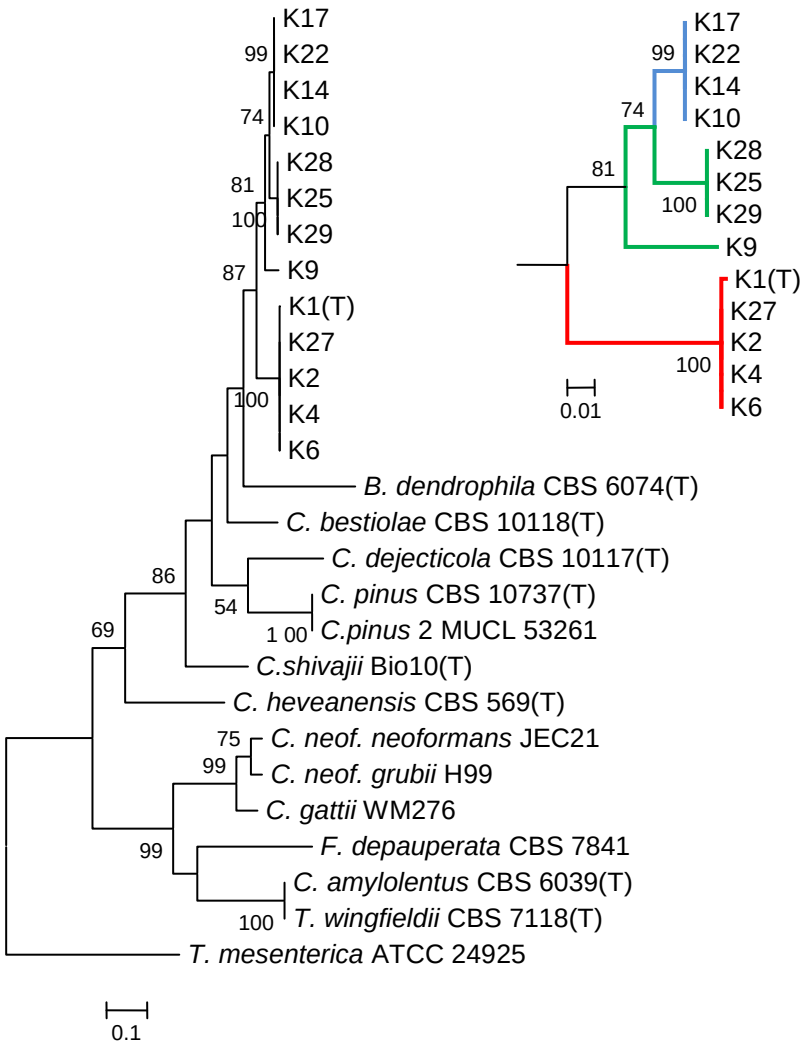


Figure 3.5 - Phylogenetic tree of *K. mangroviensis* strains and related species in the Kwoniella and Filobasidiella clades based on *RPB2* sequences. Phylogenetic relationships of strains in the Kwoniella and Filobasidiella clades based on *RPB2* coding sequences. Analyses performed with 672 positions in the final dataset. The tree was constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Insert: Detailed view of the *K. mangroviensis* clades drawn to a different scale. Sequences determined in this study are listed in Table 3.1; additional sequences were retrieved from GenBank or from JGI (see also Table 2.5).

Table 3.7 - Nucleotide Substitutions in *RPB2* coding sequences (top right) and amino acid residue substitutions in translated sequences (bottom right) of *K. mangroviensis* strains.

	K1(T)	K2	K4	K6	K27	K10	K14	K17	K22	K9	K25	K28	K29
K1(T)	ID	1	1	1	1	46	46	46	46	52	49	49	49
K2	0	ID	0	0	0	45	45	45	45	51	48	48	48
K4	0	0	ID	0	0	45	45	45	45	51	48	48	48
K6	0	0	0	ID	0	45	45	45	45	51	48	48	48
K27	0	0	0	0	ID	45	45	45	45	51	48	48	48
K10	0	0	0	0	0	ID	0	0	0	29	17	17	17
K14	0	0	0	0	0	0	ID	0	0	29	17	17	17
K17	0	0	0	0	0	0	0	ID	0	29	17	17	17
K22	0	0	0	0	0	0	0	0	ID	29	17	17	17
K9	0	0	0	0	0	0	0	0	0	ID	30	30	30
K25	0	0	0	0	0	0	0	0	0	0	ID	0	0
K28	0	0	0	0	0	0	0	0	0	0	0	ID	0
K29	0	0	0	0	0	0	0	0	0	0	0	0	ID

3.5.3 - *TEF1α* phylogenetic analysis

TEF1α sequences were more divergent than those of the previous loci (Figure 3.6). The number of different clades present was not clear due to low bootstrap support. In the *K. mangroviensis* clade, the Botswana strains grouped into two clades in the genomic and in the coding sequence tree, while in the previous analyses, they formed a single clade. The European strains were also grouped divergently in the genomic tree, but grouped into a single clade in the CDS tree. The results for the Caribbean strains were consistent with those of the other loci: all strains grouped into a single clade, with K27 being more divergent. In contrast with the previous results, in the *TEF1α* analysis, K9 grouped closer to the Caribbean clade in the genomic and CDS tree. Most of the nucleotide substitutions were in the intron, but there were also substitutions in the coding sequence, as well in the putative protein sequence (Tables 3.8 and 3.9). Among strains from the same clade, there were no amino acid residues substituted in the translated sequence, except among the European strains (up to one substitution). The two *C. neoformans* varieties differed by 19 substitutions in the genomic sequence, by nine in the CDS sequence and by one amino acid residue; whereas *C. neoformans* and *C. gattii* differed by 26 or 30 substitutions in the genomic sequence, by eight or 11 in the CDS sequence and by one or two amino acid residues.

Regarding the other species analyzed, the results for *TEF1α* were not consistent with the previous ones and neither the Kwoniella clade nor the Filobasidiella clade were well supported. In the genomic sequence and CDS sequence, *C. heveanensis* grouped outside of both clades, while *F. depauperata* appeared to cluster in the Kwoniella clade. The results for *C. pinus* were consistent with *RPB2* results and did not show any divergence. In the CDS tree, the bootstrap values were, in general, very low. *C. pinus*, *C. dejecticola* and *B. dendrophila* grouped into a single clade among the *K. mangroviensis* strains, but there was no support for it. *TEF1α* appears to be a less reliable molecular marker to predict the phylogeny of these species.

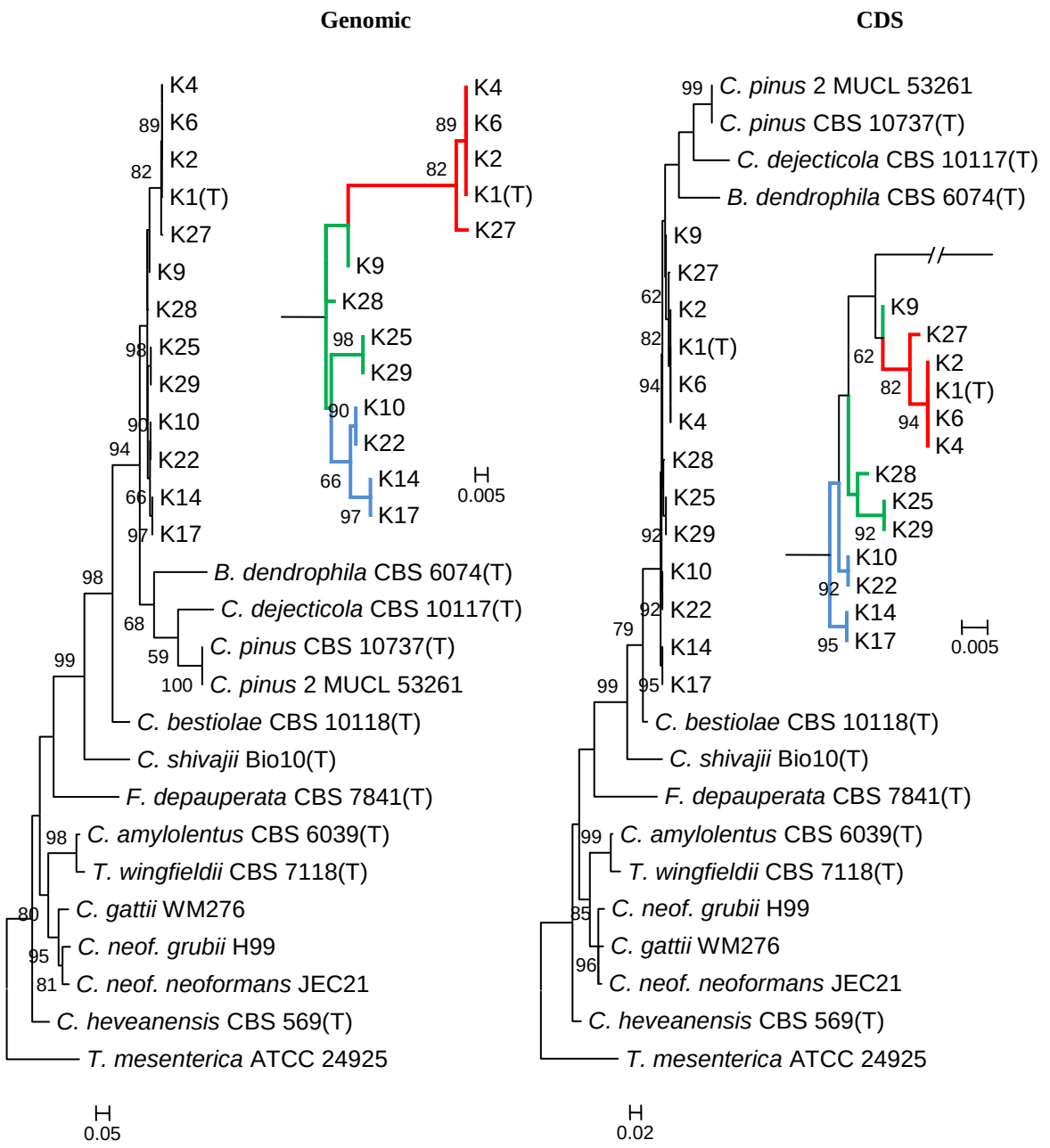


Figure 3.6 - Phylogenetic tree of *K. mangroviensis* strains and related species in the Kwoniella and Filobasidiella clades based on *TEF1α* sequences. Phylogenetic relationships of strains in the Kwoniella and Filobasidiella clades based on *TEF1α* coding sequences. Analyses performed with 854 (genomic) and 544 (CDS) positions in the final dataset. The trees were constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Insert: Detailed view of the *K. mangroviensis* clades drawn to a different scale. Sequences determined in this study are listed in Table 3.1; additional sequences were retrieved from GenBank or from JGI (see also Table 2.5).

Table 3.8 - Nucleotide Substitutions in *TEF1α* genomic (top right) and coding sequences (bottom left) of *K. mangroviensis* strains.

	K1(T)	K2	K4	K6	K27	K10	K14	K17	K22	K9	K25	K28	K29
K1(T)	ID	0	0	0	4	21	22	22	21	17	24	29	24
K2	0	ID	0	0	4	21	22	22	21	17	24	29	24
K4	0	0	ID	0	4	21	22	22	21	17	24	29	24
K6	0	0	0	ID	4	21	22	22	21	17	24	29	24
K27	3	3	3	3	ID	23	24	24	23	17	24	31	24
K10	9	9	9	9	10	ID	5	5	0	10	11	16	11
K14	9	9	9	9	10	4	ID	0	5	11	12	19	12
K17	9	9	9	9	10	4	0	ID	5	11	12	19	12
K22	9	9	9	9	10	0	4	4	ID	10	11	16	11
K9	5	5	5	5	4	6	6	6	6	ID	10	14	10
K25	10	10	10	10	9	6	8	8	6	6	ID	16	0
K28	9	9	9	9	10	4	6	6	4	6	4	ID	16
K29	10	10	10	10	9	6	8	8	6	6	0	4	ID

Table 3.9 - Amino acid residue substitutions in *TEF1α* translated sequences of *K. mangroviensis* strains.

	K1(T)	K2	K4	K6	K27	K10	K14	K17	K22	K9	K25	K28	K29
K1(T)	ID												
K2	0	ID											
K4	0	0	ID										
K6	0	0	0	ID									
K27	0	0	0	0	ID								
K10	1	1	1	1	1	ID							
K14	1	1	1	1	1	0	ID						
K17	1	1	1	1	1	0	0	ID					
K22	1	1	1	1	1	0	0	0	ID				
K9	1	1	1	1	1	0	0	0	0	ID			
K25	1	1	1	1	1	1	1	1	1	1	ID		
K28	1	1	1	1	1	0	0	0	0	0	1	ID	
K29	1	1	1	1	1	1	1	1	1	1	0	1	ID

3.5.4 - *MCM7* phylogenetic analysis

The results from the phylogenetic analysis of *MCM7* (Figure 3.7) were generally consistent with previous analyses: there were three different clades, corresponding to the LSU and ITS clades; K9 was in the same clade as the European strains; the Botswana and the European clades were closer and the Caribbean clade was the most divergent. The *MCM7* fragment analyzed does not comprise introns. From the sequence comparisons, there were no significant nucleotide substitutions among strains from the same clade; however, just like *RPB2*, many nucleotide substitutions were detected between different clades (between 21 and 42 nucleotide substitutions) (Table 3.10). All these nucleotide substitutions were synonymous in all strains, encoding the same putative protein sequence (Table 3.10).

We were not able to determine the *MCM7* sequences from *C. heveanensis* and *T. wingfieldii*. When the remaining species were analyzed, there was no significant bootstrap support, although the Filobasidiella clade was well supported. The two *C. pinus* sequences showed some minor divergence (nine synonymous nucleotide substitutions). The two *C. neoformans* varieties differed by 26 substitutions in the CDS sequence and by one amino acid residue; whereas *C. neoformans* and *C. gattii* differed by 46 or 54 substitutions in the CDS sequence and by three or four amino acid residues.

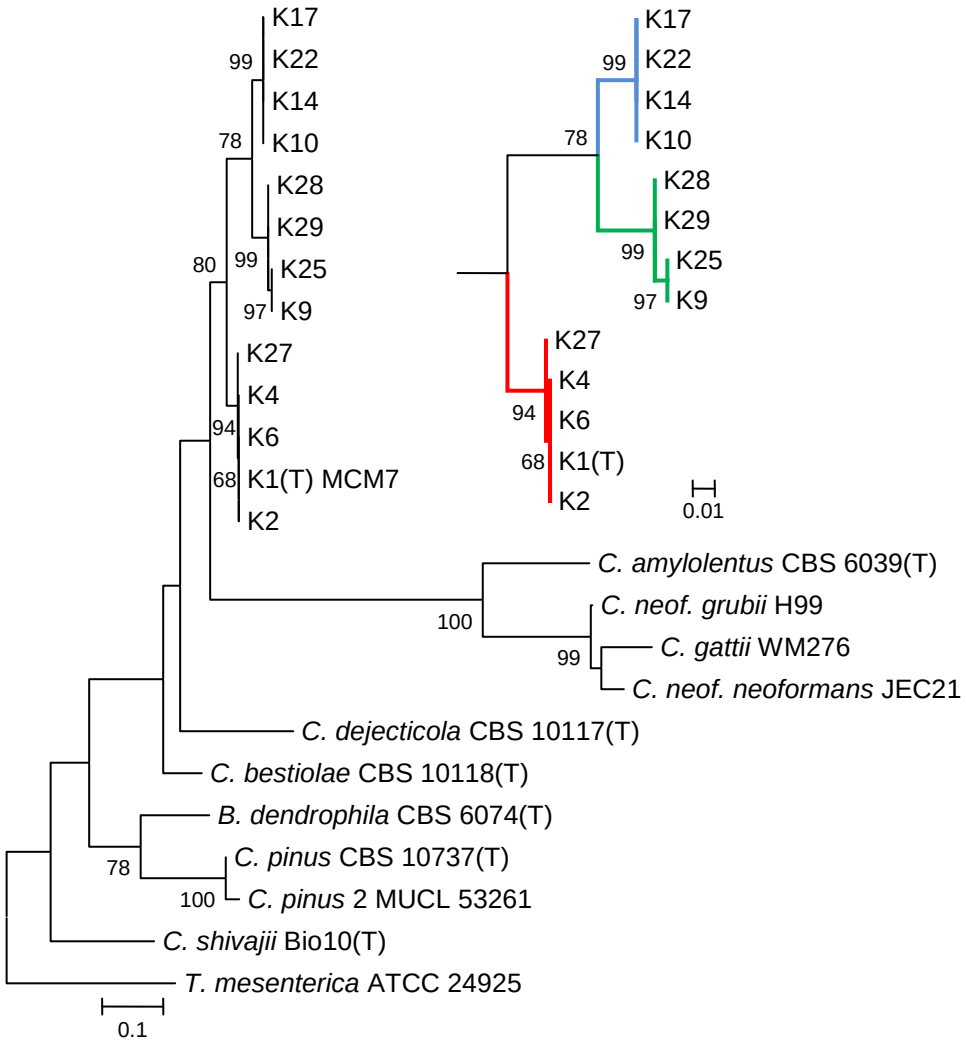


Figure 3.7 - Phylogenetic tree of *K. mangroviensis* strains and related species in the Kwoniella and Filobasidiella clades based on *MCM7* sequences. Phylogenetic relationships of strains in the Kwoniella and Filobasidiella clades based on *MCM7* coding sequences. Analyses performed with 672 positions in the final dataset. The tree was constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Insert: Detailed view of the *K. mangroviensis* clades drawn to a different scale. Sequences determined in this study are listed in Table 3.1; additional sequences were retrieved from GenBank or from JGI (see also Table 2.5).

Table 3.10 - Nucleotide Substitutions in *MCM7* coding sequences (top right) and amino acid residue substitutions in the translated sequences (bottom left) of *K. mangroviensis* strains.

	K1(T)	K2	K4	K6	K27	K10	K14	K17	K22	K9	K25	K28	K29
K1(T)	ID	0	0	0	1	35	35	35	35	42	42	39	39
K2	0	ID	0	0	1	35	35	35	35	42	42	39	39
K4	0	0	ID	0	1	35	35	35	35	42	42	39	39
K6	0	0	0	ID	1	35	35	35	35	42	42	39	39
K27	0	0	0	0	ID	34	34	34	34	41	41	38	38
K10	0	0	0	0	0	ID	0	0	0	24	24	21	21
K14	0	0	0	0	0	0	ID	0	0	24	24	21	21
K17	0	0	0	0	0	0	0	ID	0	24	24	21	21
K22	0	0	0	0	0	0	0	0	ID	24	24	21	21
K9	0	0	0	0	0	0	0	0	0	ID	0	3	3
K25	0	0	0	0	0	0	0	0	0	0	ID	3	3
K28	0	0	0	0	0	0	0	0	0	0	0	ID	0
K29	0	0	0	0	0	0	0	0	0	0	0	0	ID

3.5.5 - Concatenated sequence analyses

To establish the evolutionary relationships among *Kwoniella mangroviensis* strains, but also with other members of the Kwoniella clade, a multilocus phylogenetic analysis was performed, which included five isolates from the Caribbean, three isolates from Europe, one isolate from USA and four isolates from Botswana. Sequences from the six highly conserved genomic loci previously studied (LSU, ITS, *RPB1*, *RPB2*, *TEF1α* and *MCM7*) were concatenated and analyzed. Due to lack of one of the sequences (*MCM7*), *C. heveanensis*, *F. depauperata* and *T. wingfieldii* were not included in the MLST analysis and the resulting phylogenetic tree is presented in Figure 3.8. A phylogenetic analysis of five loci and including the latter three species did not lead to significantly different results (Figure 6.1, Appendix). The phylogenetic clustering of *K. mangroviensis* strains correlated with the respective fingerprinting profiles. There were four distinct clades: one clade (highlighted in red on Figure 3.8) comprised the Caribbean strains, with a slight divergence among them and this clade was the most distant from the other clades; the other clade (green) comprised the European strains, with some divergence among strains; K9 (purple) clustered close to the European clade; and the last clade (blue) comprised the Botswana strains. These phylogenetic analyses suggest the existence of at least three different species among *Kwoniella mangroviensis* strains, corresponding to the Caribbean, the European and the Botswana clades, respectively. Whether K9 belongs to the same species as the European strains or to a fourth species, is not entirely clear. Due to their phylogenetic divergence, the isolates from The Caribbean (K1, K2, K3, K4, K5, K6, K7, K8 and K27), which include the type-strain, should be considered as true *Kwoniella mangroviensis* (Km), and the other isolates should be accommodated in two novel species: *Kwoniella* sp. A (Ka), comprising isolates from Europe and *Kwoniella* sp. B (Kb), comprising isolates from Botswana. K9 was not included in any of these species for lack of sufficient supporting evidence. The original study (Statzell-Tallman *et al.*, 2008) classified K25 and K26 as *K. mangroviensis*, but from this MLST analysis, these two strains belong to a distinct species (Ka).

The Kwoniella and Filobasidiella clades were defined with strong bootstrap support (100), and the topology was similar to that described by Findley *et al.* (2009). However, from this analysis, *C. bestiolae* was the closest species to *K. mangroviensis*, and the phylogenetic position of *B. dendrophila* was not determined precisely, due to low bootstrap support. *C. pinus* was found to be the closest species to *C. dejecticola*, and the two *C. pinus* strains did not show significant divergence among them. The phylogenetic position of *C. shivajii* was defined by a five loci MLST analysis comprising *C. heveanensis* (Figure 6.1), which confirmed the former as a member of the Kwoniella clade, with *C. heveanensis* at a basal position in the clade. The phylogenetic positions of *F. depauperata* and *T. wingfieldii* were also determined, which corresponded to the described phylogeny for these species (Findley *et al.*, 2009): *F. depauperata* basal to the Filobasidiella clade, and *T. wingfieldii* grouping with *C. amylolentus*.

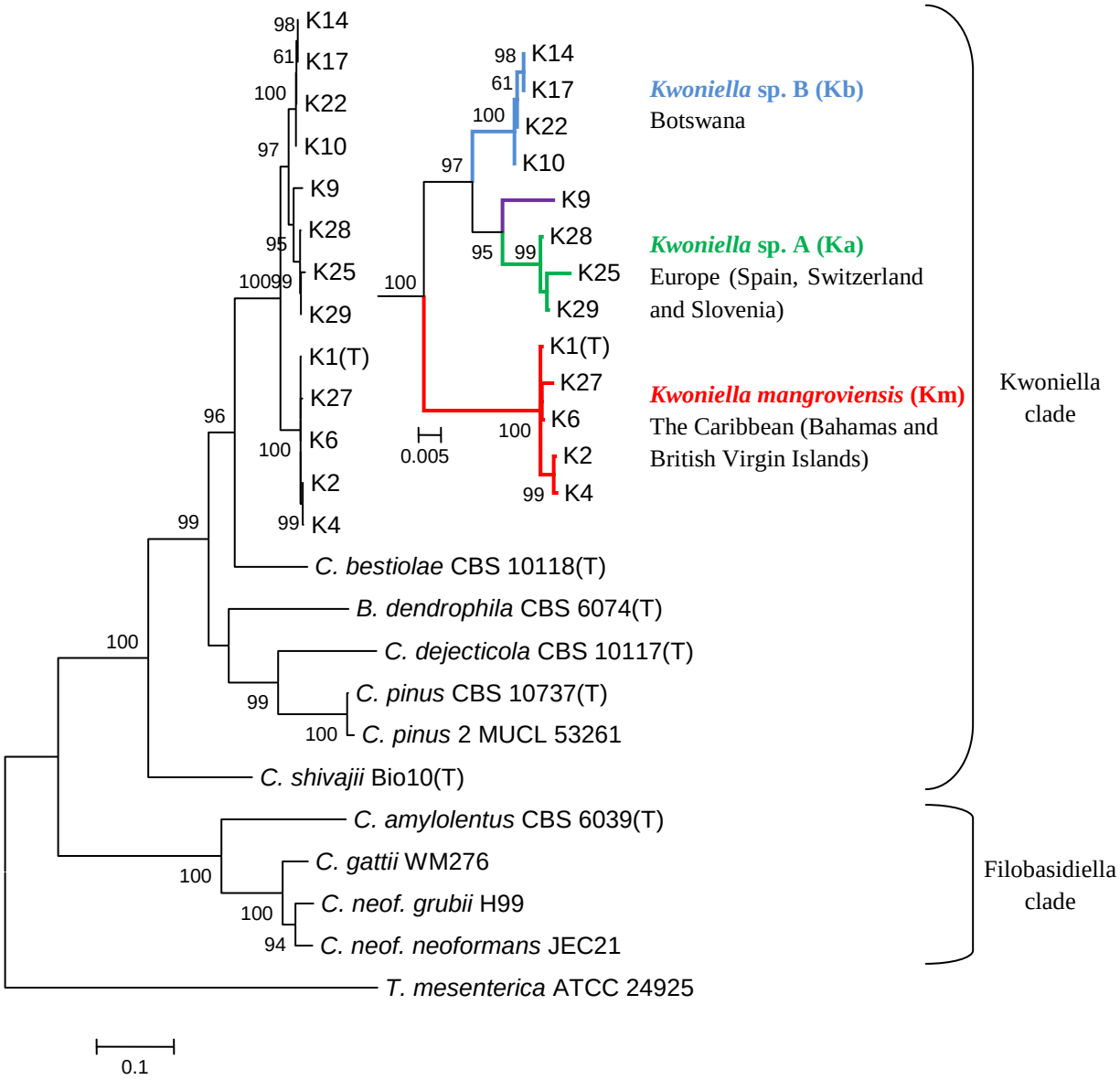


Figure 3.8 - Phylogenetic relationships among members of the Kwoniella and Filobasidiella clades based on a combined data set of concatenated sequences of six genomic loci (LSU, ITS, *RPB1*, *RPB2*, *TEF1α* and *MCM7*). Analyses performed with 3948 positions in the final dataset. The tree was constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Insert: Detailed view of the *K. mangroviensis* clades drawn to a different scale. Sequences determined in this study are listed in Table 3.1; additional sequences were retrieved from GenBank or from JGI (see also Table 2.5).

3.6 - Distance Analysis

To determine evolutionary distances between the species, we calculated Jukes-Cantor distances based on the concatenated sequences of five loci: LSU, ITS, *RPB1*, *RPB2*, and *TEF1α* (Figure 3.9 and Table 3.11). Comparisons were made with *C. amylolentus* and *T. wingfieldii*, which were considered two different species (Findley *et al.*, 2012) and with the pathogenic *Cryptococcus* complex, comprising two varieties (*C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii*) and two different species (*C. neoformans* and *C. gattii*) (Campbell and Carter, 2006). The results showed equivalent distances between Km and Ka, Km and Kb, K9 and Km, which were slightly higher than the difference between the two *C. neoformans* varieties; however, these distances were lower than the distance between *C. neoformans* and *C. gattii*. These distances can be correlated to the respective degree of divergence. When compared, Kb and Ka, K9 and Ka, K9 and Kb showed equivalent distances between them, but higher than the distance between the *C. amylolentus* and *T. wingfieldii*. Much lower values were observed among strains of the same determined species (Km, Ka and Kb) and *C. pinus* strains exhibited a slightly higher distance value when compared to those.

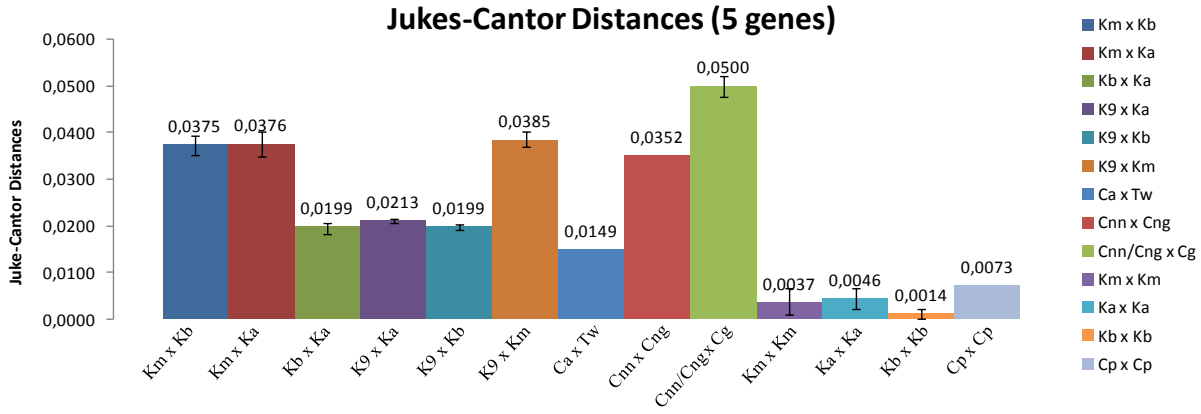


Figure 3.9 - Jukes-Cantor distances comparisons based on five concatenated gene sequences. Distances between *K. mangroviensis* (Km), *Kwoniella sp*. A, *Kwoniella sp*. B, K9, *C. amylolentus* (Ca), *T. wingfieldii* (Tw), *C. neoformans* var. *neoformans* (Cnn), *C neoformans* var. *grubii* (Cng), *C. neoformans* (Cnn/Cng), *C. gattii* (Cg) and *C. pinus* strains (Cp). Values indicated above each bar represent the distances average.

Table 3.11 - Jukes-Cantor distances.

	Km x Kb	Km x Ka	Kb x Ka	K9 x Ka	K9 x Kb	K9 x Km	Ca x Tw	Cnn x Cng	Cnn/Cng x Cg	Km x Km	Ka x Ka	Kb x Kb	Cp x Cp
Maximum	0,0392	0,0403	0,0206	0,0216	0,0203	0,0403			0,0522	0,0065	0,0065	0,0023	
Average	0,0375	0,0376	0,0199	0,0213	0,0199	0,0385	0,0149	0,0352	0,0500	0,0037	0,0046	0,0014	0,0073
Minimum	0,0353	0,0349	0,0182	0,0206	0,0192	0,0370			0,0478	0,0010	0,0023	0,0000	

3.7 - *MAT* gene analyses

Previous studies (Á. Fonseca, unpublished; J.A. Rodrigues and Á. Fonseca, unpublished) determined the sequences of *STE20* and *SXI1*/*SXI2* (*SXI* region) from K1, K2, K3, K4, K5, K6 and K7, which confirmed the mating-type specificity of these genes and constituted the first evidence for a tetrapolar mating system on *K. mangroviensis*. Based on these studies, *STE20* and *SXI* sequences were determined and analyzed for the remaining strains (Table 3.1). Molecular mating-types were then determined based on the obtained sequences of these genes, where the P/R locus (A) was defined by *STE20* alleles and the HD locus (B) defined by the *SXI* region.

3.7.1 - *STE20* gene (P/R locus)

The analysis of *STE20* sequences revealed mating-type specific polymorphisms with at least two alleles for each of the recognized species: *K. mangroviensis*, *Kwoniella* sp. A and *Kwoniella* sp. B (Figure 3.10). In *K. mangroviensis*, KmA1 allele was present in K1, K3, K6, K7 and K27 strains, while KmA2 allele was present in K2, K4, K5 and K8. In *Kwoniella* sp. A, the KaA1 allele was assigned to K25, K26 and K29, while KaA2 was assigned to K28. In *Kwoniella* sp. B, the KbA1 allele was assigned to K10, K11 and K22, while KbA2 was assigned to K14, K16, K17, K20, K21, K23 and K24. A unique allele was identified for K9 (K9A1). This gene was sequenced for the available strains of *C. pinus*, which showed a slight deviation of the type-strain from the remaining sequences that were identical. *C. bestiolae* was sequenced, as well, and clustered close to *C. heveanensis* sequence. It was not possible to determine the phylogenetic pattern of *STE20* in *C. bestiolae*, because only one strain was available. *C. amylolentus* CBS 6273 (A2B2) (Findley *et al*., 2012) was also sequenced and displayed a mating-type specific pattern, when compared with the strain previously sequenced. New primers for *STE20* gene were designed (STE20_F3 and STE20_R3) but we were not able to determine this sequence for *C. dejecticola*, *C. shivajii* and *B. dendrophila*.

A set of degenerate primers for *STE3* and *STE12* genes was also designed, due to their mating-type specificity in related species. The primers were specific for each mating-type, designed based on *C. neoformans MAT*α or *MAT***a** sequences, and also considered every possible orientation for both genes. All primer combinations were tested for all strains. From these experiments, only one primer combination was able to successfully amplify these genes for *K. mangroviensis* strains with KmA1 alleles, *Kwoniella* sp. A strains with KaA1 and KaA2 alleles and *Kwoniella* sp. B strains with KbA1 alleles (data not shown). The obtained amplicons from K1 and K10 strains were sequenced, which confirmed the presence of both genes, that were divergently oriented. This finding was consistent with the divergent orientation of both genes in closely related species (*T. mesenterica*, *C. heveanensis*, *C. amylolentus*, *T. wingfieldii*, *C. neoformans* and *C. gattii*). Due to low quality of sequencing results, the

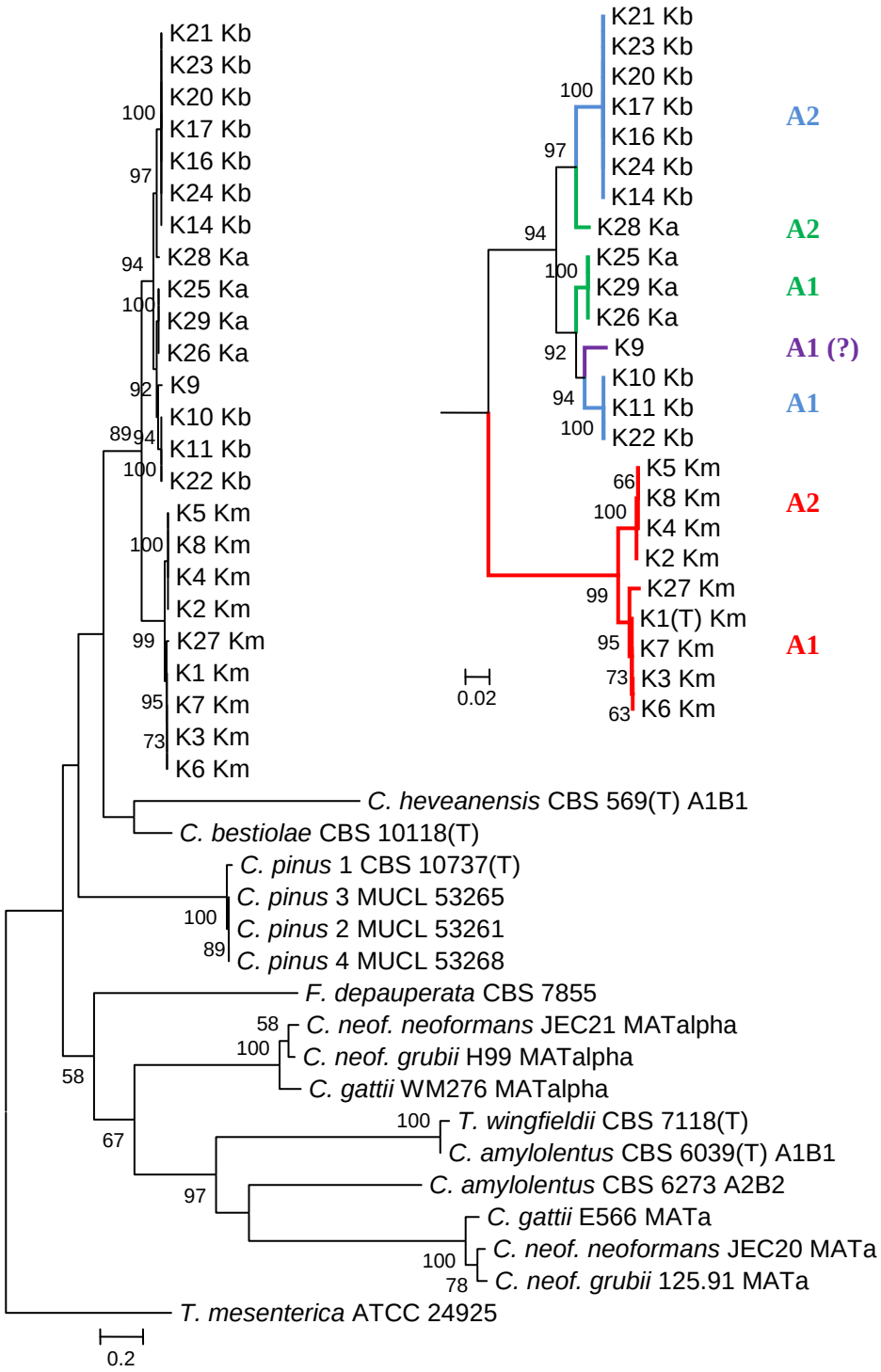


Figure 3.10 - Phylogenetic tree of K. mangroviensis strains and related species in the Kwoniella and Filobasidiella clades based on *STE20* partial sequence. Phylogenetic relationships of strains in the Kwoniella and Filobasidiella clades based on *STE20* sequences. Analyses performed with 1121 positions in the final dataset. The tree was constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Insert: Detailed view of the *K. mangroviensis* clades drawn to a different scale. Sequences determined in this study are listed in Table 3.1 additional sequences were retrieved from GenBank or from JGI (see also Table 2.4).

remaining strains were not sequenced and await primer improvement. The impossibility of amplifying the complementary allele in *K. mangroviensis* indicates the existence of considerable polymorphism between the two alleles, consistent with the mating-type specific pattern observed within the *C. neoformans* species complex and *C. heveanensis*; however, both alleles appeared to be amplified for Ka, data which still needs to be confirmed by sequencing. The amplification of these genes, for Km, Ka and Kb, suggests a possible common origin previous to species divergence.

3.7.2 - *SXI* genes (HD locus)

The analysis of *SXI* region revealed mating-type specific polymorphisms for this region that appears to be multiallelic (Figure 3.11). *K. mangroviensis* exhibited four alleles: KmB1 was present in K1, K3, K4, K5 and K8, KmB2 was present only in K2, KmB3 was present in K6 and K7 and a new allele, KmB4, was assigned to K27. Three alleles were identified for *Kwoniella* sp. A: KaB1 was assigned to K25 and K26, KaB2 was assigned to K28 and KaB3 was assigned to K29. Only one allele was identified for *Kwoniella* sp. B (Figure 3.11). Surprisingly, K9 shared the same allele of K25 and K26 (KaB1). We were not able to determine sequences for this region from *C. pinus*, *B. dendrophila* and *C. shivajii*.

3.7.3 - Molecular mating-types

From the combination of the results for the *STE20* and *SXI* genes we identified five different mating-types for *K. mangroviensis*, three for *Kwoniella* sp. A, two for *Kwoniella* sp. B and K9 sharing one of the alleles with Ka for locus B (Table 3.12). K8, which is self-filamentous, shared the same mating-type as the self-filamentous strains K4 and K5 (J.A. Rodrigues and Á. Fonseca, unpublished). P/R and HD alleles appear to segregate independently, as seen in Table 3.12, since different alleles are shuffled between strains.

Table 3.12 - Molecular mating-types determined for each strain.

Species	Molecular mating-type	Strains
Kwoniella mangroviensis	KmA1B1	K1, K3
	KmA1B3	K6, K7
	KmA1B4	K27
	KmA2B1	K4, K5, K8
	KmA2B2	K2
Kwoniella sp. A	KaA1B1	K25, K26
	KaA2B2	K28
	KaA2B3	K29
Kwoniella sp. B	KbA1B1	K10, K11, K22
	KbA2B1	K14, K16, K17, K20, K21, K23, K24
?	K9A1KaB1	K9

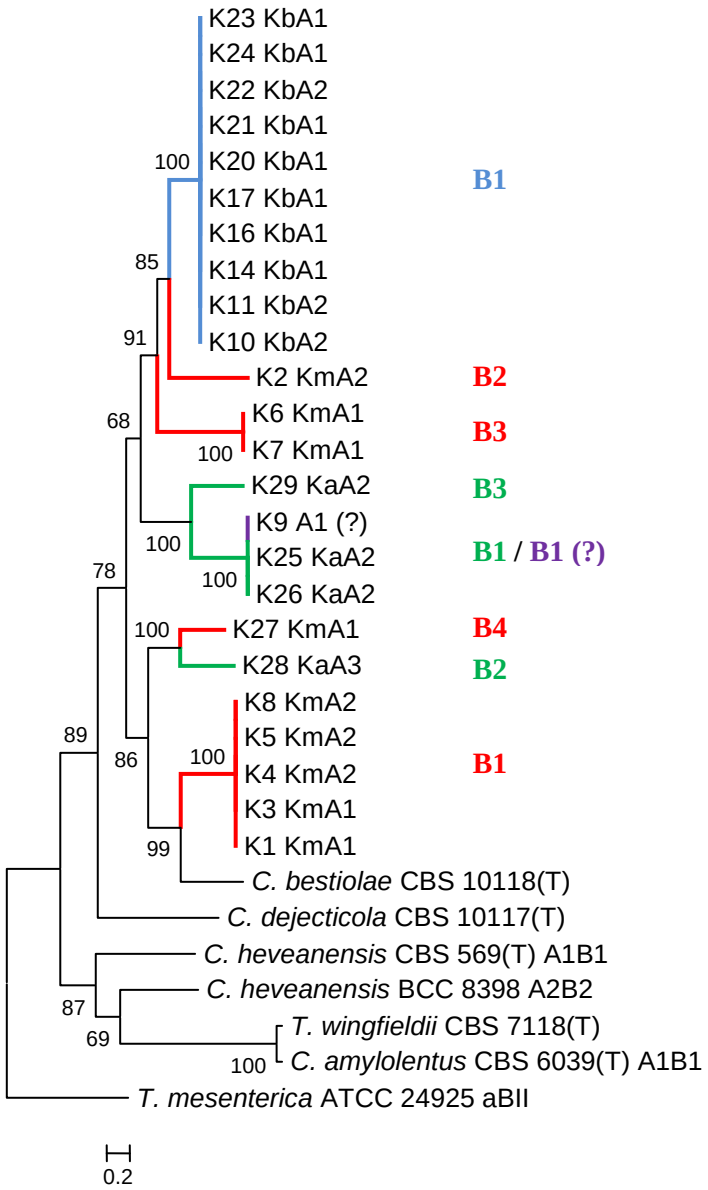


Figure 3.11 - Phylogenetic tree of K. mangroviensis strains and related species in the Kwoniella and Filobasidiella clades based on *SXI1*, *SXI2* partial sequences and their intergenic region. Phylogenetic relationships of tetrapolar strains in the Kwoniella and Filobasidiella clades based on *SXI* region. Analyses performed with 1644 positions in the final dataset. The tree was constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Sequences determined in this study are listed in Table 3.1 additional sequences were retrieved from GenBank or from JGI (see also Table 2.4).

3.8 - Gene content and organization of the P/R region of *K. mangroviensis*

3.8.1 - K1 fosmid sequencing

To determine the structure of *MAT* in *K. mangroviensis*, libraries were constructed from K1 and K2 genomes, and probed for several homolog genes (*STE20, RPO41, LPD1, GDA1* and *SXI***)** that were present in *C. neoformans MAT* locus (Á. Fonseca, unpublished). Positive results for *STE20* were obtained in K1 libraries and *LPD1* and *RPO41* in K2 libraries. The latter two genes were found not to be mating-type specific. *STE20* was the only mating-type gene to be identified within the probed libraries. To ensure that the fosmid included *MAT* genes, the clone 2K2 from K1 library, positive for *STE20* was selected for sequencing. The fosmid sequence spanned about 43 kb and appeared to contain a fragment of the P/R locus (Figure 3.12). Upon annotation, 12 *C. neoformans* gene homologs were identified. The sequenced region included five *MAT* genes previously identified in the *MAT* loci of other studied species, namely a pheromone precursor gene (*MF*), *STE3* and *STE12*, which were confirmed to be divergently oriented, *BSP3*, and the predicted and already studied *STE20* gene. Some homologs of *C. neoformans* genes not present in the *MAT* locus were also found (*CNE02670, CND05260, CNI00160, CNB00600, CNB00610, CNG04540* and *CNF01610).* The linear order of the fragment within *K. mangroviensis* genome was not determined.

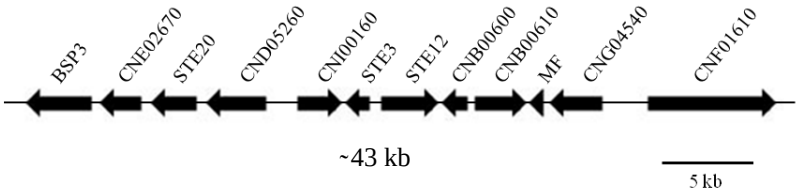


Figure 3.12 - *Kwoniella mangroviensis* P/R locus fragment. A single fosmid was sequenced and analyzed to generate the sequence for *K. mangroviensis* P/R region. The P/R locus is contained in a region that spans 43 kb. The linear order of the fragment within the genome was not determined. Scale bar = 5 kb.

3.8.2 - Synteny analysis

A synteny analysis of the P/R locus fragment identified in *K. mangroviensis* was performed by comparing with homologous regions of *C. neoformans*, *C. heveanensis* and *T. mesenterica* (Figure 3.13). The analysis showed an almost perfect synteny between *K. mangroviensis* and *C. heveanensis*, with only an inversion in the pheromone precursor gene (Figure 3.13). *K. mangroviensis* had an additional gene (*CNF01610*), that was absent in *C. heveanensis* but present in *T. mesenterica* (represented by green arrows in Figure 3.13). When compared with *C. neoformans*, five genes were common to both species (black arrows in Figure 3.13), but no synteny block was observed. Seven genes were not present in *C. neoformans MAT* locus, but five of them were present in *C. heveanensis* P/R region and all of them, were present in *T. mesenterica* P/R region.

3.8.3 - Pheromone precursor analysis

One pheromone precursor gene was found in the *K. mangroviensis* P/R region. Similar to the other *MF* genes found in related species, this gene did not have any introns. When translated *in silico*, the putative pheromone precursor had 39 amino acid residues in length. A comparison performed with related species, showed common motifs at the N-terminal (MDAFT) and at the C-terminal (CVIA) (Figure 3.14). When the *K. mangroviensis* pheromone precursor was compared with its homologs in related species and identities determined, the highest identity was shared with *C. heveanensis* A1B1 (64%), while *C. neoformans* var. *neoformans*, *C. neoformans* var. *grubii* and *C. gattii MAT*α shared 63%, 63% and 55% identities, respectively. *K. mangroviensis* pheromone precursor had 48% identity with the related homolgs in *C. heveanensis* A2B2, *C. amylolentus* A1B1, *T. wingfieldii*, and *C. neoformans* var. *neoformans*, *C. neoformans* var. *grubii* and *C. gattii MAT***a**. *T. mesenterica* tremerogen-a13 shared 33% identity with its homolog in *K. mangroviensis*. From this analysis, the pheromone precursor from *K. mangroviensis* appeared to be more closely related to the *MAT*α than to the *MAT***a** homologs of the *Cryptococcus* species complex.

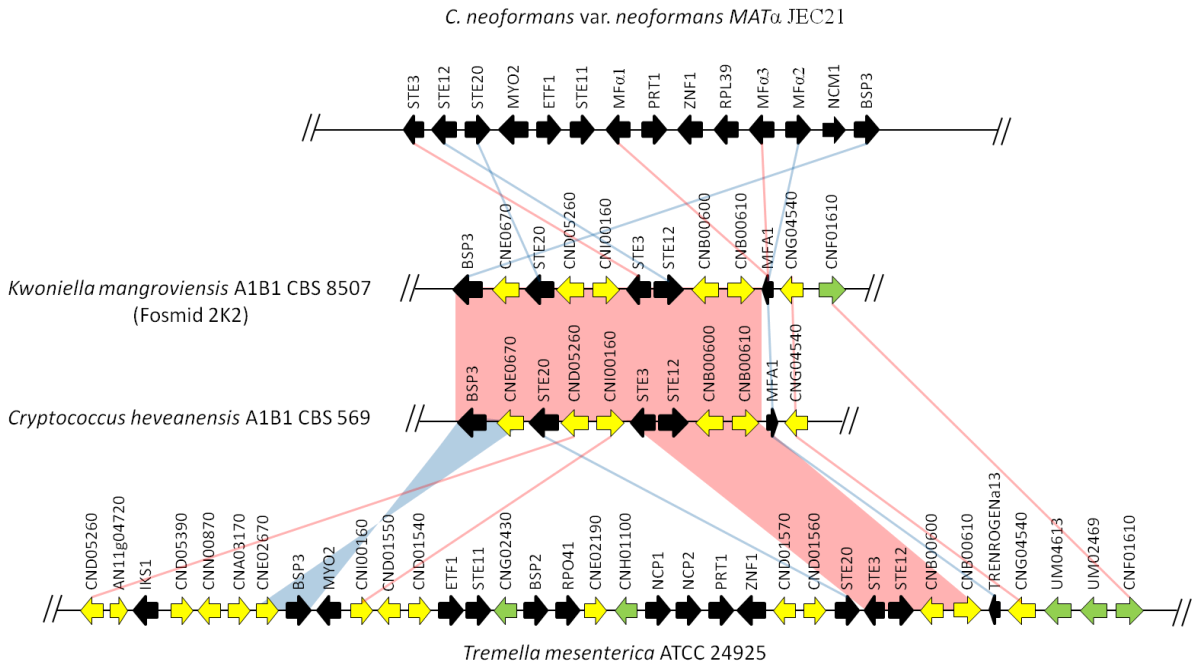


Figure 3.13 - Synteny analysis. Comparison of *MAT* gene content and organization between *C. neoformans*, *K mangroviensis*, *C. heveanensis* and *T. mesenterica*. Black arrows stand for genes present in *C. neoformans*, yellow arrows stands for genes absent in *C. neoformans MAT* but present in *C. heveanensis* and green arrows stand for genes absent in both *C. neoformans* and *C. heveanensis MAT*. Red bars and red lines denote conserved gene orientation and order; while blue bars and blue lines denote inverted gene orientation and order.

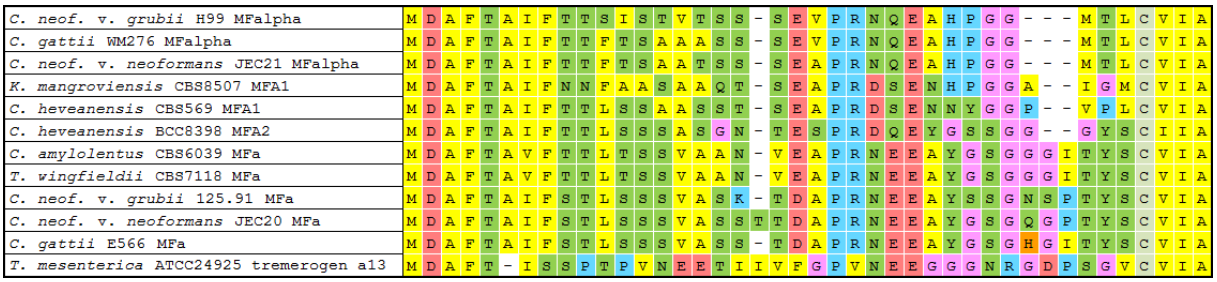
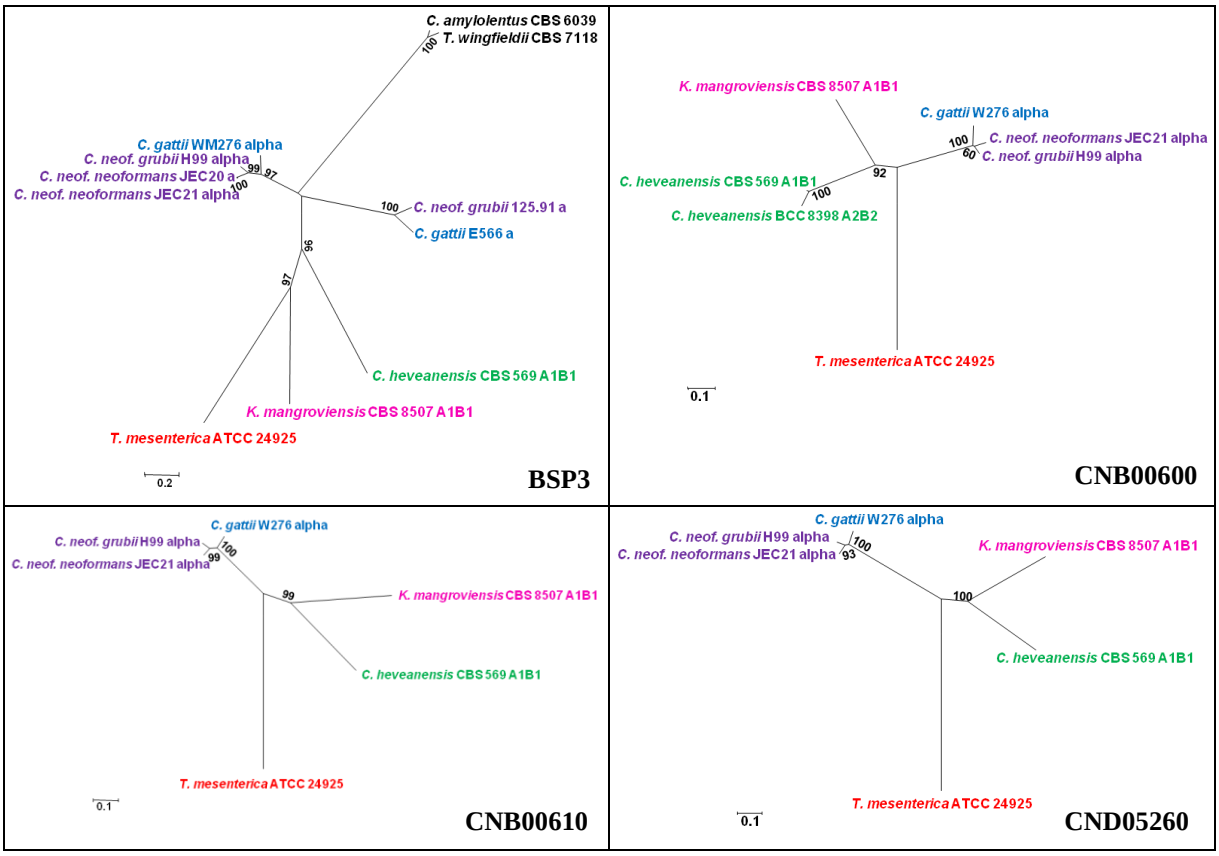


Figure 3.14 - Sequence analysis of the pheromone precursor protein. Multiple protein sequence alignment of the pheromone precursor genes performed in MEGA 5.05 by direct translation of the nucleotide sequence.

3.8.4 - Phylogenetic patterns of genes within the P/R region

Phylogenetic analyses were performed for every gene found within the sequenced P/R region, along with homologous sequences available for related species (Figure 3.15). From the analyses, *MF*, *STE3*, and *STE12* of *K. mangroviensis* KmA1, exhibited a mating-type specific pattern, clustering closer to *C. neoformans MAT*α homologs. From previous analysis *STE20*, also exhibited a mating-type specific pattern (see section 3.8.1), but it was not clear in this analysis, since KmA1 *STE20* clustered between opposite mating-type sequences of *C. neoformans*. Complete *STE20* gene sequences were not obtained for KmA2 strains. To determine accurately the phylogenetic pattern from the remaining genes, additional sequences from KmA2 are required.



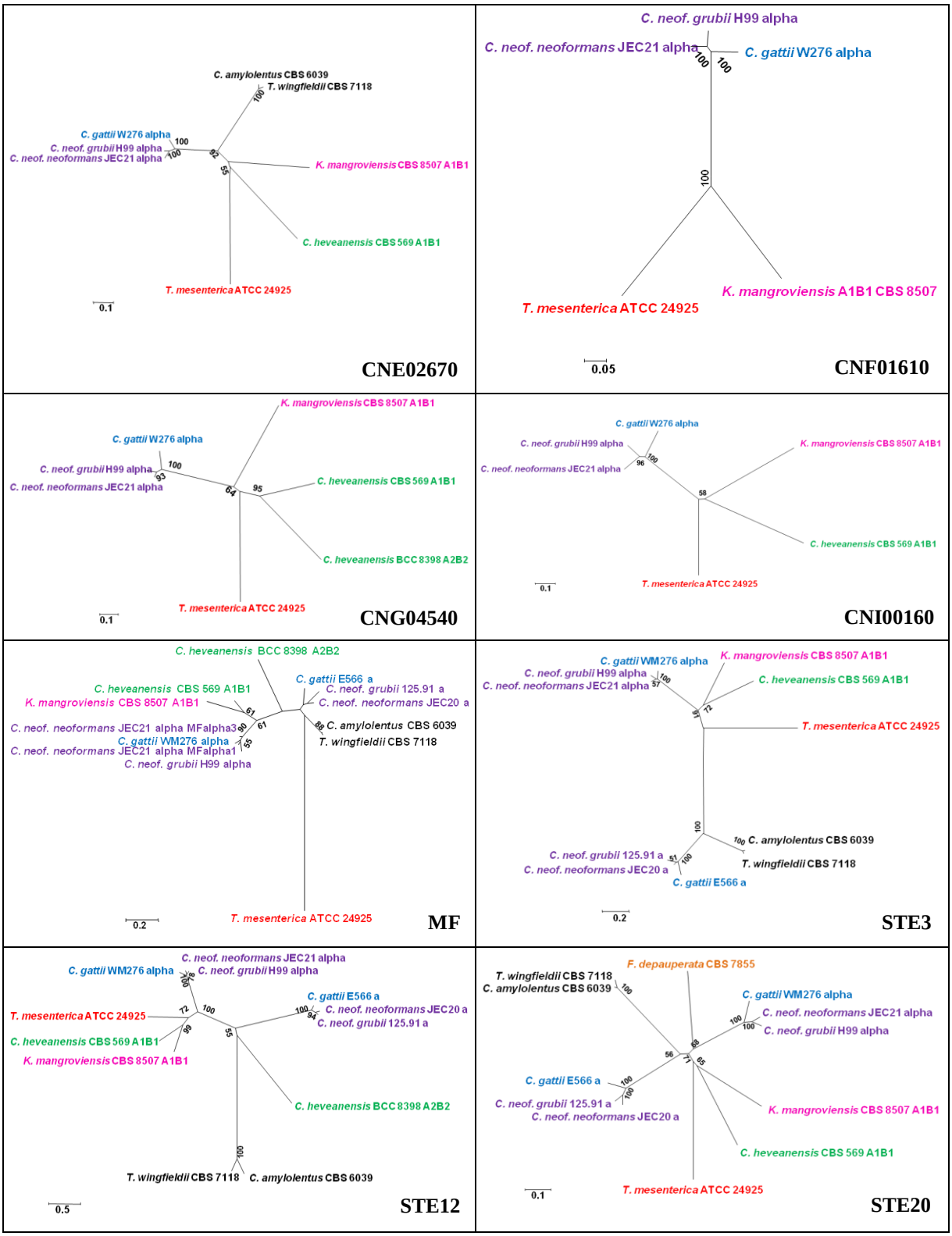


Figure 3.15 - Phylogenetic analyses of *MAT* genes. Phylogenetic analyses of genes comprised within P/R region of *K. mangroviensis* based on complete genomic sequences. The trees are unrooted and were constructed using the Neighbor-Joining method implemented in the software MEGA 5.05. *K. mangroviensis* sequences determined were deposited in EMBL (HE997060); additional sequences were retrieved from GenBank or from JGI (see also Table 2.4).

4 - DISCUSSION

4.1 - Reassessment of species boundaries of *Kwoniella mangroviensis*

In the original study describing *Kwoniella mangroviensis* (Statzell-Tallman *et al*., 2008), species assignment for the studied strains was performed based only on LSU and ITS sequences. Although two strains (K25 and K26) exhibited some divergence in ITS, when compared to the other strains, they were classified in *K. mangroviensis*. In the study described in this thesis, ten strains from the original study and 14 additional strains were analyzed (Table 2.1).

To determine the genetic diversity and evolutionary relationships within *Kwoniella mangroviensis*, two complementary molecular approaches were employed: PCR fingerprinting and multilocus sequence typing (MLST) based on six loci LSU, ITS, *RPB1*, *RPB2*, *TEF1α* and *MCM7*. PCR fingerprinting included all the 24 *K. mangroviensis* strains, while the MLST analysis included 13 strains, five from the original study and eight additional strains, as well as additional species from the Kwoniella and Filobasidiella clades. When the MLST results (Figure 3.8) for *K. mangroviensis* were compared with the fingerprinting profiles (Figure 3.3), the results were consistent. Both techniques showed four genetically distinct groups. The first group comprised strains from the Caribbean, the second group strains from Europe, including K25 and K26, the third group strains from Botswana, and the last group had a single strain, K9. For simplicity, the first group is designated as Km, the second as Ka and the third as Kb. The existence of three consistent and supported clades indicates divergence between these groups. Based on the MLST analysis (Figure 3.8), the Km group clustered separately from the other two groups. When this analysis was compared to the *Cryptococcus* species complex, there was a clear parallel between the phylogeny of the latter and that observed for the *Kwoniella* strains: the two *C. neoformans* varieties clustered more closely and *C. gattii* clustered more divergently. When nucleotide substitutions from protein-coding genes were compared, there were always more substitutions between *C. neoformans* and *C. gattii* than between the two *C. neoformans* varieties. Those substitutions were not all synonymous among the two species; however, completely synonymous substitutions between the two varieties were present in *RPB2*. When the substitutions among *Kwoniella* groups were compared, only the differences between Km and Ka or Km and Kb correlated with the differences observed between the two *Cryptococcus* species. Differences between Ka and Kb were always lower than the differences between the two *Cryptococcus* varieties. The substitutions in K9 were not consistent among the loci, when compared to Ka and Kb. To compare more accurately the *Kwoniella* strains with the *Cryptococcus* complex, and to clarify the phylogenetic relationship between *Kwoniella* groups, genetic distances were calculated.

An analysis based on Jukes-Cantor distances was performed to determine the genetic distance based on LSU, ITS, *RPB1*, *RPB2* and *TEF1α* between strains of each group (Figure 3.9 and Table 3.11). The distances between the pathogenic *Cryptococcus* varieties and species were used as

reference, as well as for the pair *C. amylolentus* and *T. wingfieldii*. From the comparison, three different distance levels were detected among *K. mangroviensis* groups. In the first level, the distances were similar to the distances between the two *C. neoformans* varieties and higher than the distance between *C. amylolentus* and *T. wingfieldii*; in the second level, the distances were about half of *C. neoformans* varieties distance and slightly higher than *C. amylolentus* and *T. wingfieldii*; and in the third level, the distances were minimal (Figure 3.9). The first level corresponded to the distances between Km and Ka, and Kb and K9, respectively. The second level corresponded to the distances between Ka, Kb and K9. The third level corresponded to the distances between strains within Ka, Kb and Km. From these comparisons, Km appeared to be the most distant group of all, whereas Ka and Kb were the closer groups, while K9 was at the same distance from the latter.

These analyses demonstrated significant genetic divergence between the four groups mentioned and, for that reason, the existence of at least three different species is proposed. The Caribbean group included the *K. mangroviensis* type-strain (K1) and, therefore, the strains within this group (K2, K3, K4, K5, K6, K7, K8 and K27) were classified in *K. mangroviensis* (Km). According to the observed phylogenetic clustering (Figure 3.8) and genetic distances (Figure 3.9 and Table 3.11), the other groups were classified as two novel species: *Kwoniella* sp. A (Ka), comprising the European strains (K25, K26, K28 and K29), and *Kwoniella* sp. B (Kb), comprising the Botswana strains (K10, K11, K14, K16, K17, K20, K21, K22, K23 and K24). It was not entirely clear whether or not K9 belonged to species Ka, and if these two newly described species are in fact two species, or just varieties of a single *Kwoniella* species. The genetic distances between Ka and Kb are similar to the distance between the two *C. neoformans* varieties (Figure 3.9 and Table 3.11), although, Ka and Kb appears to be unable to mate (Table 3.3). Both varieties of *C. neoformans* are able to produce progeny but fail when mated with *C. gattii* (Hull and Heitman, 2002). Mating experiments do not help defining the taxonomical level of Ka and Kb; however, *MAT* genes support the existence of two species (see below).

Physiological tests performed for Km, Ka, Kb and K9 (Table 3.4) did not exhibit significant differences that could help to differentiate the species. The only difference observed was in the growth temperature between Km and the other groups: Km strains grew up to 30ºC, while the remaining strains grew up to 35ºC.

4.2 - Phylogeny and taxonomic status of species in the Kwoniella and Filobasidiella clades

From the MLST analysis (Figure 3.8) and calculated genetic distances (Figure 3.9 and Table 3.11), although *C. pinus* strains were from distinct geographical origins, they did not show major divergence among them. In fact, those distances were similar to the distances among strains of the same species in the *K. mangroviensis* clade (Table 3.11); therefore, the studied *C. pinus* strains were

considered to belong to a single species. The MLST analysis also confirmed that *B. dendrophila*, *C. bestiolae*, *C. dejecticola*, *C. heveanensis*, *C. pinus*, *C. shivajii*, *Kwoniella* sp. A, *Kwoniella* sp. B and *Kwoniella mangoviensis* are members of the Kwoniella clade and *F. depauperata*, *C. amylolentus*, *T. wingfieldii*, *C. neoformans* var. *neoformans*, *C. neoformans* var. *grubii* and *C. gattii* are members of the Filobasidiella clade (Figures 3.8 and 6.1). Findley *et al.* (2009), unveiled the phylogeny of the Kwoniella clade, by performing a MLST analysis based on LSU, ITS, *RPB1*, *RPB2*, *TEF1α* and *mitSSU* (Figure 1.4). In this study, in addition to the species studied by Findley *et al*. (2009), the phylogenies of two additional species were determined: *C. pinus* and *C. shivajii*. The results obtained in this study were consistent with the results from the previous study; however, in this analysis *C. heveanensis* position is questionable due to moderate bootstrap support (Figure 6.1); *C. bestiolae* replaced *B. dendrophila* as the closest species to *K. mangroviensis*; and *B. dendrophila* position is uncertain due to low bootstrap support (Figure 3.8). *C. dejecticola* was also found to be the closest species to *C. pinus*, and *C. shivajii* was found to be the second most ancestral species to the Kwoniella clade, after *C. heveanensis*.

From the distance analysis (Figure 3.9 and Table 3.11), a proposal, in the Filobasidiella clade, concerning the *T. wingfieldii* and *C. amylolentus* classification is suggested. *T. wingfieldii* and *C. amylolentus* were classified as different species by Findley *et al*. (2009) based on a MLST and physiological tests and by Findley *et al*. (2012) based on *MAT* genes. From our analysis, the genetic distance between *T. wingfieldii* and *C. amylolentus* is less than half of the distance between the two *C. neoformans* varieties. Based on this data, the taxonomical classification of these taxa should be revised. Due to their genetic distance and by comparison, *T. wingfieldii* and *C. amylolentus* should be classified as members of the same species or varieties.

4.3 - Mating experiments and molecular mating-types

From the mating experiments performed in this study, only two strains (K8 and K27) besides the positive controls, developed sexual structures when crossed with compatible strains (Table 3.3). The molecular mating-types determined for these two strains (Table 3.12), correlate with the mating experiments, since these two strains are only able to mate with strains possessing compatible molecular mating-types. K8 revealed to be self-filamentous, like K4 and K5. There seems to be a correlation between self-filamentous ability and the mating-type, since K4, K5 and K8 shared the same unique mating-type (KmA2B1) and were the only self-filamentous strains observed. Several other strains had one of the alleles (KmA2 or KmB1) and none of them were self-filamentous (Table 3.3). This seems to be a unique feature of this allele combination. The triggering mechanisms, in *K. mangroviensis*, for the self-filamentous ability influenced by the mating-type were not clarified.

The mating (Table 3.3) and the molecular results (Table 3.12) provide additional support to the hypothesis of different species. The strains that were able to mate in this and in the previous

studies (Tables 1.2 and 3.3), belonged to the same assigned species (*K. mangroviensis*), and failed to mate with *Kwoniella* sp. A or *Kwoniella* sp. B strains. Molecular mating-types were determined for all the strains (Figures 3.10 and 3.11 and Table 3.12), and two additional compatible combinations of molecular mating-types (KaA1B1 x KaA2B2 and KaA1B1 x KaA2B3) were determined for *Kwoniella* sp. A (Table 3.12). However, and in contrast with *K. mangroviensis*, these strains were not able to develop any kind of sexual structures or filaments (Table 3.3), even though molecular data suggested sexual compatibility. If these strains are in fact sexually compatible, the required mating conditions might need to be determined for this species.

Genetic isolation or genetic divergence could explain the unsuccessful mating between strains of different species, since the molecular mating-types suggested the existence of different alleles in both loci, and therefore, sexual compatibility was expected.

4.4 - Tetrapolar mating-system in *K. mangroviensis* and sibling species

Kwoniella mangroviensis was firstly described as possessing a bipolar mating-system (Statzell-Tallman *et al*., 2008). Preliminary studies (section 1.12) showed that this organism was tetrapolar and four different mating-types were described (Table 1.2) (Á. Fonseca, unpublished; J.A. Rodrigues and Á. Fonseca, unpublished). From the study described in this thesis, one additional mating-type was identified for *K. mangroviensis*, three for *Kwoniella* sp. A and two for *Kwoniella* sp. B (Table 3.12). This study thus provides additional evidence for a tetrapolar mating-system in *K. mangroviensis* and the identification of potential tetrapolar mating-systems in *Kwoniella* sp. A and *Kwoniella* sp. B.

Four main lines of evidence support the tetrapolar mating-system of these three species: 1) the presence of two HD genes divergently transcribed; 2) a multiallelic HD locus; 3) different allele combinations; and 4) mating experiments. 1) In the bipolar species of the pathogenic *Cryptococcus* complex, only one of the HD genes is present in each *MAT* locus (Figure 1.5) (Fraser *et al*., 2004), while in the tetrapolar species studied to date (*T. mesenterica*, *C. heveanensis*, *C. amylolentus* and *T. wingfieldii*), both genes are present and divergently transcribed (sections 1.7, 1.8 and 1.9). In *K. mangroviensis*, *Kwoniella* sp. A and *Kwoniella* sp. B, both HD genes were identified adjacent and divergently transcribed (Figure 3.11). 2) In tetrapolar basidiomycetes, the P/R locus is usually biallelic and the HD is multiallelic (Kües *et al*., 2011). In the three *Kwoniella* species, a biallelic P/R locus was found (Figure 3.10), with two alleles present in each species. In *K. mangroviensis* and *Kwoniella* sp. A, a multiallelic HD locus was found (Figure 3.11), with the identification of four alleles in Km and three in Ka. In Kb only one allele was identified (Figure 3.11). 3) When alleles of both loci were combined and compared, five different mating-types were identified in total for *K. mangroviensis*, three for *Kwoniella* sp. A and two for *Kwoniella* sp. B (Table 3.12). The most important feature in these mating-types is that alleles were found in different combinations among strains, suggesting

independent segregation (for example, A1B1, A2B1 and A2B2). In bipolar mating-systems only the same allele combination is detected among strains (**a**/α or A1B1/A2B2) due to genetic linkage of both alleles in a single contiguous region, therefore preventing meiotic recombination. In *Kwoniella* sp. B, only two mating-types were described, but since both mating-types were combining two different P/R alleles (Figures 3.10 and 3.11) (KbA1B1 and KbA2B1), the tetrapolar mating-system hypothesis is still supported by allele combination. 4) In the tetrapolar mating systems, individuals must have compatible mating-types, i.e. different alleles at both loci in order to mate (Casselton and Challen, 2006). Direct evidence for tetrapolarity was only totally supported in *K. mangroviensis*. Molecularly determined mating-types correlated with the mating-experiments, in this species (Table 3.3 and 3.12). Strains with different alleles at both loci were able to mate and develop sexual structures, demonstrating a typical tetrapolar sexual compatibility. This was also true for *Kwoniella* sp. B, due to the inability of the strains to mate, consistent with the sexual incompatibility of both mating-types, since all strains share the same HD allele (Figure 3.11). Additional strains with compatible mating-types and additional mating experiments are required to demonstrate sexual compatibility among strains of *Kwoniella* sp. A and *Kwoniella* sp. B.

From all the results previously described, the mating-system of *Kwoniella mangroviensis* is confirmed to be tetrapolar and not bipolar as proposed by Statzell-Tallman *et al*. (2008). From the molecular data, is it suggested that *Cryptococcus bestiolae* and *Cryptococcus dejecticola* might also be tetrapolar due to the presence of both HD genes divergently transcribed (Figure 3.11). *STE20* was also detected in *C. pinus* (Figure 3.10), but we were not able to determine if this was a mating-type specific or a species-specific gene. Further studies are required to provide insight into *C. bestiolae*, *C. dejecticola* and *C. pinus MAT* loci.

4.5 - Strain K9

From MLST (Figure 3.8) and distance analyses (Figure 3.9 and Table 3.11), it was not clear whether or not K9 belonged to any of the newly identified species. From the determined molecular mating-types (Figures 3.10 and 3.11 and Table 3.12), K9 shared the same HD allele as K25 and K26, and a unique P/R allele, suggesting sexual compatibility with K28 and K29, based on this analysis; however, the mating results between K9 and *Kwoniella* sp. A were negative. Molecular mating-type data suggested that K9 may belong to *Kwoniella* sp. A, but on the other hand, MLST and distance analyses suggested that K9, *Kwoniella* sp. A and *Kwoniella* sp. B are equidistant.

Mating-type genes are highly polymorphic even in the same species, with the homeodomain genes and the pheromone receptor gene being the most polymorphic genes. It is unlikely that different species could share some allele of these genes, suggesting that K9 may belong to *Kwoniella* sp. A. In contrast, the genetic divergence of K9 and Ka is comparable to interspecific distances determined for other species (Figure 3.9 and Table 3.11). In basidiomycetes, the P/R locus is usually biallelic. Two

different alleles were detected for this locus in K25/K26 (KaA1) and K28/K29 (KaA2) (Figure 3.10). If K9 belonged to this species, this locus would be multiallelic, a feature never found in related species of the Kwoniella and Filobasidiella clades. From the individual trees of the MLST analyses, K9 was more closely related to Ka in ITS/LSU (Figure 3.2) and *MCM7* (Figure 3.7), but closer to Kb in *RPB1* (Figure 3.4), and it was clustering out of both clades in *RPB2* (Figure 3.5) and *TEF1α* (Figure 3.6). If K9 was a hybrid, these results could be justified. In order to clarify this, additional genes would have to be analyzed and studied for K9, Ka and Kb.

4.6 - *MAT* evolution in the Tremellales

Mating-type loci are highly variable in size, gene content and organization. In the pathogenic *Cryptococcus* species complex, the gene content is similar but the *MAT* locus underwent several gene rearrangements and inversions among species, varieties and mating-types (Figure 1.5) (Fraser *et al*., 2004). Some *MAT* loci were sequenced and analyzed for several related species (sections 1.7, 1.8 and 1.9), such as *C. amylolentus* (Figure 1.7), *T. wingfieldii* (Figure 1.8), *C. heveanensis* (Figure 1.10) and *T. mesenterica* (Figure 1.12). When compared, the gene content and loci arrangement of these different species were highly variable (section 1.10 and Table 1.1), except only between *C. amylolentus* and *T. wingfieldii* HD locus, where synteny was completely conserved (Findley *et al.*, 2012).

In this study, a fragment of the P/R locus of *Kwoniella mangroviensis* type-strain was sequenced and analyzed (section 3.8). Upon annotation, several *MAT* genes were identified within this fragment (Figure 3.12) and comparisons were made with *MAT* loci of related species. Phylogenetic analyses of the genes within this fragment (Figure 3.15) only enabled determination of the corresponding phylogenetic pattern for *MAT* genes whose sequences were available for both mating-types of the pathogenic *Cryptococcus* species. Since we were not able to sequence other genes either than *STE20* (Figure 3.10) in the KmA2 allele of the P/R locus of *K. mangroviensis*, the species-specific genes and the P/R boundaries of this species were not determined. However, from the analyses performed, we found that *STE20*, *STE3*, *STE12* and *MF* are mating-type specific (Figure 3.15); therefore, we can predicted that this locus may comprise at least eight genes: *STE20*, *CND05260*, *CNI00160*, *STE3*, *STE12*, *CNB00600*, *CNB00610* and *MF* genes. From those phylogenetic analyses we also determined that the A1 strain of *K. mangroviensis* was more similar to *C. neoformans MAT*α than to *MAT***a** homologs. To accurately define the P/R boundaries, comparisons with the complementary allele are required.

Synteny analysis (Figure 3.13) revealed that five genes in the sequenced region, had homologs in *C. neoformans MAT* locus; however, no synteny between *K. mangroviensis* and *C. neoformans* was observed. When the genetic structure of the P/R locus was compared with that of the *C. heveanensis* A1 strain, an almost complete synteny was observed, with a single exception in the pheromone

precursor gene orientation (Figure 3.13). In related species, the *MAT* loci undergo several gene rearrangement and inversions between mating-types and across varieties and species (Figure 1.5). The high degree of synteny between *K. mangroviensis* and *C. heveanensis*, was thus not expected . When compared to the *T. mesenterica* P/R locus, only a cluster comprising *STE12*, the pheromone precursor gene, *STE3* and some three additional genes (*CNB00600, CNB0060* and *CNG04540*) was conserved in the same order and orientation between *K. mangroviensis* and *T. mesenterica* (Figure 3.13). This suggests that the origin of this cluster was prior to the last common ancestor of *T. mesenterica*, *C. heveanensis* and *K. mangroviensis*, and that gene order and orientation were conserved, with possibly minor inversions in the pheromone precursor gene. The almost perfect synteny between *C. heveanensis* mat A1 and *K. mangroviensis* mat A1, suggests that: (i) this locus arrangement occurred before the speciation of both species, (ii) *MAT* evolution and divergence was slower in the Kwoniella clade than in the Filobasidiella clade, and (iii) synteny might be conserved between the other Kwoniella clade species. This also suggests that the KmA2 allele of the P/R locus of *K. mangroviensis* may also share a high degree of synteny with the corresponding A2 allele of *C. heveanensis*. Further studies need to be performed in order to determine the *MAT* structure of the KmA2 allele of *K. mangroviensis*, and also from the other species in the Kwoniella clade. Currently, three complete sequences (*CNG04540*, *MFA2* and *STE12*) and two partial sequences (*CNB00600* and *STE3*) are available for genes within the A2 P/R region of *C. heveanensis* (Metin *et al*., 2010). To unveil the P/R locus structure of the *K. mangroviensis* KmA2 allele, degenerate primers should be designed based on KmA1 sequences and the available homologous sequences of *C. heveanensis* A1 and A2 alleles. Then, a long range PCR approach can be used to amplify and sequence fragments of the P/R region, and/or new *MAT* gene probes can be designed in order to re-probe K2 genomic libraries.

An almost perfect synteny among species of the Kwoniella clade is a remarkable discovery and will enable to determine accurately the steps and mechanisms behind *MAT* loci evolution in the Tremellales.

4.7 - Conclusions and future perspectives

The study described in this thesis, which used an innovative combination of MLST and *MAT* gene analyses, led to the identification of two novel species in the Kwoniella clade, *Kwoniella* sp. A and *Kwoniella* sp. B. These two novel species appeared to be closer to each other than to *K. mangroviensis*. A clinical isolate was also analyzed. A common *MAT* allele was identified between *Kwoniella* sp. A and the clinical isolate. However, the ambiguous phylogenetic position and genetic distance of the clinical isolate to the newly identified species, did not allow us to determine with certainty its taxonomical status. A taxonomic review of *T. wingfieldii* and *C. amylolentus* classification was also proposed. *C. amylolentus* and *T. wingfieldii* exhibited a genetic distance less than half of the value determined for the *C. neoformans* varieties, and therefore should be classified as members of the

same species. Our MLST approach confirmed the phylogenetic position of *C. pinus* and *C. shivajii* as members of the Kwoniella clade.

The analyses performed based on *MAT* genes and mating assays, allowed us to confirm the tetrapolar mating-system of *Kwoniella mangroviensis* and to hypothesize similar tetrapolar mating-systems for the two newly identified species and probably also for *C. bestiolae* and *C. dejecticola*. Moreover, this study provided insight into the genetic structure of the P/R region of a *K. mangroviensis* mating-type A1 strain, by sequencing a genomic fragment containing twelve putative *MAT* genes. Comparisons with homologous regions of related species revealed an almost perfect synteny between *K. mangroviensis* and *C. heveanensis*.

Further studies are required to determine the taxonomic status of the clinical isolate, to confirm and demonstrate the tetrapolar mating-system of *Kwoniella* sp. A, *Kwoniella* sp. B, *C. bestiolae* and *C. dejecticola*, and to determine the P/R region structure of the A2 mating-type of *Kwoniella mangroviensis*.

5 - REFERENCES

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389-3402.

Bandoni, R.J. and Boekhout, T. 2011. *Tremella* Persoon (1794). *In* The Yeasts, A Taxonomic Study (C.P. Kurtzman, J.W. Fell and T. Boekhout eds), 5th ed., pp. 1567-1590, Elsevier B.V., Amsterdam.

Campbell, L.T. and Carter, D.A. 2006. Looking for sex in the fungal pathogens *Cryptococcus neoformans* and *Cryptococcus gattii*. FEMS Yeast Research 6:588-598.

Casadevall, A. and Perfect, J.R. 1998. *Cryptococcus neoformans*. ASM Press, Washington, DC.

Casselton, L.A. and Challen, M.P. 2006. The mating type genes of the Basidiomycetes. *In* The Mycota I - Growth, Differentiation and Sexuality (U. Kües and R. Fischer eds.), 2nd ed., pp 356–374, Springer Berlin Heidelberg, Berlin.

Cooper Jr., C.R. 2011. Yeasts pathogenic to humans. *In* The Yeasts, A Taxonomic Study (C.P. Kurtzman, J.W. Fell and T. Boekhout eds), 5th ed., pp. 9-21, Elsevier B.V., Amsterdam.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.

Findley, K. 2010. Evolution of the mating-type locus and insights into sexual reproduction in the *Cryptococcus* species complex. PhD thesis. Department of Molecular Genetics and Microbiology, Duke University.

Findley, K., Rodriguez-Carres, M., Metin, B., Kroiss, J., Fonseca, Á., Vilgalys, R. and Heitman, J. 2009. Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the Tremellales. Eukaryotic Cell 8:363–361.

Findley, K., Sun, S., Fraser, J.A., Hsueh, Y.-P., Averette, A.F., Li, W., Dietrich, F.S. and Heitman, J. 2012. Discovery of a Modified Tetrapolar Sexual Cycle in *Cryptococcus amylolentus* and the Evolution of *MAT* in the *Cryptococcus* Species Complex. PLoS Genetics 8:e1002528.

Fraser, J.A., Diezmann, S., Subaran, R.L., Allen, A., Lengeler, K.B., Dietrich, F.S. and Heitman, J. 2004. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS Biology 2:2243–2255.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Hawksworth, D. L. 1991. The fungal dimension of biodiversity: magnitude, significance, and conservation. Mycological Research 95:641-655.

Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., Huhndorf, S., James, T., Kirk, P.M., Lucking, R., Thorsten Lumbsch, H., Lutzoni, F., Matheny, P.B., McLaughlin, D.J., Powell, M.J., Redhead, S., Schoch, C.L., Spatafora, J.W., Stalpers, J.A., Vilgalys, R., Aime, M.C., Aptroot, A., Bauer, R., Begerow, D., Benny, G.L., Castlebury, L.A., Crous, P.W., Dai, Y.C., Gams, W., Geiser, D.M., Griffith, G.W., Gueidan, C., Hawksworth, D.L., Hestmark, G., Hosaka, K., Humber, R.A., Hyde, K.D., Ironside, J.E., Koljalg, U., Kurtzman, C.P., Larsson, K.H., Lichtwardt, R., Longcore, J., Miadlikowska, J., Miller, A., Moncalvo, J.M., Mozley-Standridge, S., Oberwinkler, F., Parmasto, E., Reeb, V., Rogers, J.D., Roux, C., Ryvarden, L., Sampaio, J.P., Schussler, A., Sugiyama, J., Thorn, R.G., Tibell, L., Untereiner, W.A., Walker, C., Wang, Z., Weir, A., Weiss, M., White, M.M., Winka, K., Yao, Y.J. and Zhang, N. 2007. A higher-level phylogenetic classification of the Fungi. Mycological Research 111:509-547.

Hsueh, Y.-P., Metin, B., Findley, K., Rodriguez-Carres, M. and Heitman, J. 2011. The mating-type locus of *Cryptococcus*: evolution of gene clusters governing sex determination and sexual reproduction from phylogenomic perspective. *In Cryptococcus*: From human pathogen to model yeast (J. Heitman, T.R. Kozel, K.J. Kwon-Chung, J.R. Perfect and A. Casadavall eds), pp 139–149, ASM Press, Washington, DC.

Hull, C.M. and Heitman, J. 2002. Genetics of *Cryptococcus neoformans*. Annual Review of Genetics 36:557–615.

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schüßler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lücking, R., Büdel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J.,

Spatafora, J.W. and Vilgalys, R. 2006. Reconstructing the early evolution of *Fungi* using a six-gene phylogeny. Nature 443:818–822.

Jukes, T.H. and Cantor, C.R. 1969. Evolution of protein molecules. *In* Mammalian Protein Metabolism, III (Munro HN, editor), pp. 21-132, Academic Press, New York.

Kozubowski, L., Lee, S.C. and Heitman, J. 2009. Signalling pathways in the pathogenesis of *Cryptococcus*. Cellular Microbiolology 11:370–380.

Kües, U., James, T.Y. and Heitman, J. 2011. Mating type in basidiomycetes: unipolar, bipolar, and tetrapolar patterns of sexuality. *In* The Mycota XIV - Evolution of Fungi and Fungal-Like Organisms (S. Pöggeler and J. Wöstemeyer eds), pp 97–160, Springer Berlin Heidelberg, Berlin.

Kurtzman, C.P., Fell, J.W. and Boekhout, T. 2011a. Definition, Classification and Nomenclature of the Yeasts. *In* The Yeasts, A Taxonomic Study (C.P. Kurtzman, J.W. Fell and T. Boekhout eds), 5th ed., pp. 3-9, Elsevier B.V., Amsterdam.

Kurtzman, C.P., Fell, J.W., Boekhout, T. and Robert, V. 2011b. Methods for Isolation, Phenotypic Characterization and Maintenance of Yeasts. *In* The Yeasts, A Taxonomic Study (C.P. Kurtzman, J.W. Fell and T. Boekhout eds), 5th ed., pp. 87-110, Elsevier B.V., Amsterdam.

Kwon-Chung, K.J. 2011. *Filobasidiella* Kwon-Chung (1975). *In* The Yeasts, A Taxonomic Study (C.P. Kurtzman, J.W. Fell and T. Boekhout eds), 5th ed., pp. 1443-1456, Elsevier B.V., Amsterdam.

Lengeler, K.B., Davidson, R.C., D'Souza, C., Harashima, T., Shen, W.C., Wang, P., Pan, X.W., Waugh, M. and Heitman, J. 2000. Signal transduction cascades regulating fungal development and virulence. Microbiology and Molecular Biology Reviews 64:746–785.

Lengeler, K.B., Fox, D.S., Fraser, J.A., Allen, A., Forrester, K., Dietrich, F.S. and Heitman, J. 2002. Mating-type locus of *Cryptococcus neoformans*: A step in the evolution of sex chromosomes. Eukaryotic Cell 1:704–718.

Lin, X. 2009. *Cryptococcus neoformans*: Morphogenesis, infection, and evolution. Infection, Genetics and Evolution 9:401–416.

Lin, X. and Heitman, J. 2006. The biology of the *Cryptococcus neoformans* species complex. Annual Review of Microbiology 60:69-105.

Liu, Y.L., Whelen, S. and Hall, B.D. 1999. Phylogenetic relationships among ascomycetes: evidence from an RNA polymerase II subunit. Molecular Biology and Evolution 16:1799-1808.

Matheny, P.B., Liu, Y.J., Ammirati, J.F., and Hall, B.D. 2002. Using *RPB1* sequences to improve phylogenetic inference among mushrooms (Inocybe, Agaricales). American Journal of Botany 89:688-698.

Metin, B., Findley, K. and Heitman, J. 2010. The Mating Type Locus (*MAT*) and Sexual Reproduction of *Cryptococcus heveanensis*: Insights into the Evolution of Sex and Sex-Determining Chromosomal Regions in Fungi. PLoS Genetics 6:e1000961.

Morrow, C.A. and Fraser, J.A. 2009. Sexual reproduction and dimorphism in the pathogenic basidiomycetes. FEMS Yeast Research 9:161–177.

Ni, M., Feretzaki, M., Sun, S., Wang, X. and Heitman, J. 2011. Sex in Fungi. Annual Review of Genetics 45:405-430.

O'Donnell, K. 1993. Fusarium and its near relatives. *In*: The Fungal Holomorph: Mitotic, Meiotic and Pleomorphic Speciation in Fungal Systematics (D.R. Reynolds and J.W. Taylor eds), pp 225–233, CAB International, Wallingford, UK.

Raudaskoski, M. and Kothe, E. 2010. Basidiomycete mating type genes and pheromone signaling. Eukaryotic Cell 9:847–859.

Rodriguez-Carres, M., Findley, K., Sun, S., Dietrich, F.S. and Heitman, J. 2010. Morphological and genomic characterization of *Filobasidiella depauperata*: a homothallic sibling species of the pathogenic *Cryptococcus* species complex. PLoS One 5:e9620.

Schmitt, I., Crespo, A., Divakar, P.K., Fankhauser, J.D., Herman-Sackett, E., Kalb, K., Nelsen, M.P., Nelson, N.A., Rivas-Plata, E., Shimp, A.D., Widhelm, T. and Lumbsch, H.T. 2009. New primers for promising single-copy genes in fungal phylogenetics and systematics. Persoonia 23:35–40.

Sheng, S., Metin, B., Findley, K., Fonseca, Á. and Heitman, J. 2011. Validation of *Kwoniella heveanensis*, teleomorph of the basidiomycetous yeast *Cryptococcus heveanensis*. Mycotaxon 116:227-229.

Statzell-Tallman, A., Belloch, C. and Fell, J.W. 2008. *Kwoniella mangroviensis* gen. nov., sp.nov. (Tremellales, Basidiomycota), a teleomorphic yeast from mangrove habitats in the Florida Everglades and Bahamas. FEMS Yeast Research 8:103-113.

Stiller, J.W. and Hall, B.D. 1997. The origin of red algae: Implications for plastid evolution. PNAS 94: 4520-4525.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences 17:57-86.

Thompson, J., Higgins, D. and Gibson, T. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22:4673-4680.

Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A.S. and Christophe, D. 1987. A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. Science 235**:**683-684.

Vilgalys, R. and Hester, M. 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. Journal of Bacteriology 172:4238–4246.

White, T.J., Bruns, T., Lee, S. and Taylor, J.W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* PCR Protocols: A Guide to Methods and Applications (M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White eds.), pp. 315-322, Academic Press, Inc., New York.

Whitehouse, H.L.K. 1949. Multiple allelomorph heterothallism in fungi. New Phytologist 48:212–244.

Xu, J. 2006. Fundamentals of fungal molecular population genetic analyses. Current Issues in Molecular Biology 8:75–90.

6 - APPENDIX

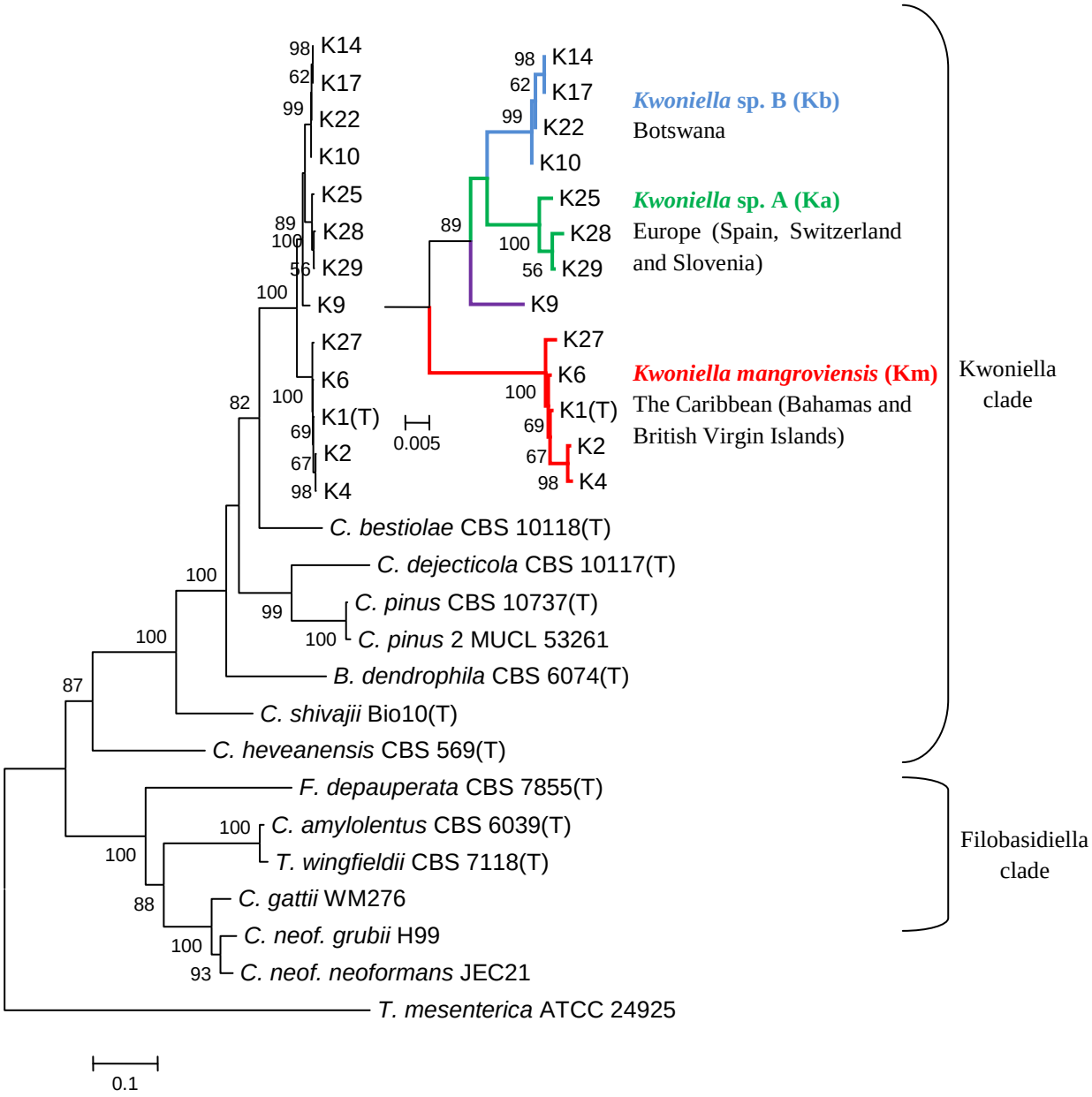


Figure 6.1 - Phylogenetic relationships among members of the Kwoniella and Filobasidiella clades based on a combined data set of concatenated sequences of five genomic loci (LSU, ITS, *RPB1*, *RPB2* and *TEF1α*). Analyses performed with 3397 positions in the final dataset. The tree was constructed using the Maximum Likelihood method implemented in the software MEGA 5.05. Numbers on branches are bootstrap values (>50) from 1000 replicates. *T. mesenterica* was used as outgroup in the analysis. (T) represents type-strain. Insert: Detailed view of the *K. mangroviensis* clades drawn to a different scale. Sequences determined in this study are listed in Table 3.1; additional sequences were retrieved from GenBank or from JGI (see also Table 2.5).