Marco António Dias Coelho

Licenciado em Biologia

Molecular mechanisms of sexual development in basidiomycetes: exploring connections with lifestyles

Dissertação para obtenção do Grau de Doutor em Biologia

Orientador: Doutor José Paulo Sampaio, Prof. Auxiliar, FCT/UNL
Co-orientador: Doutora Paula Gonçalves, Profª. Auxiliar, FCT/UNL

Júri:

Presidente: Profª. Doutora Elvira Júlia da Conceição Matias Coimbra
Arguentes: Profª. Doutora Geraldine Butler
Prof. Doutor Alexander Idnurm

Vogais: Prof. Doutor Jan Schirawski
Prof. Doutor Manuel António da Silva Santos
Prof. Doutor Álvaro Luís Afonso Moreira Rato Fonseca
Profª. Doutora Paula Maria Theriaga Mendes Bernardo Gonçalves
Prof. Doutor José Paulo Nunes de Sousa Sampaio
Inv. Principal Doutora Isabel Antunes Mendes Gordo

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Outubro de 2011

Marco António Dias Coelho

Molecular mechanisms of sexual development in basidiomycetes: exploring connections with lifestyles

Acknowledgments

I want to express my deeply gratitude to my thesis advisors, Profª. Paula Gonçalves and Prof. José Paulo Sampaio, for their encouragement, enthusiasm and thoughtful guidance along these years.

I also want to thank the members of the thesis committee, Prof. Álvaro Fonseca and Dr. Isabel Gordo, for their interest and suggestions to improve this thesis.

I would like to thank Profª. Regine Kahmann for giving me the opportunity to work at the Department of Organismic Interactions of Max Planck Institute for Terrestrial Microbiology (Marburg, Germany), under supervision of Prof. Jan Schirawski, to whom I am particularly grateful for the helpful suggestions and support.

I am indebted to current and former lab colleagues for providing a friendly lab atmosphere and helpful discussions from which I have learned so much.

I thank my friends for their support and for helping me to stay focused during these years. I greatly value their friendship and I genuinely appreciate their trust in me.

Most importantly, this work would not have been possible without the love, concern, patience and endless support of my family to whom I dedicate this dissertation. A special thank to my dearest and nearest 'girl of my life' for the immeasurable sacrifices and support and for never failing faith on me.

Finally, I appreciate the financial support from Fundação para a Ciência e a Tecnologia that funded my PhD grant (SFRH/BD/29580/2006).

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR Portugal

Resumo

Este trabalho procura contribuir para o estudo dos mecanismos moleculares de reprodução sexuada em fungos e para a compreensão da sua relevância na evolução dos diferentes modos de vida dos fungos (parasita vs. saprofítico) e na emergência de linhagens assexuadas.

O desenvolvimento sexual e a patogenicidade estão interligadas em muitos basidiomicetas fitoparasitas, sendo *Ustilago maydis* o exemplo emblemático de um grupo genericamente designado por "carvões" ou "morrões". Estes fungos causam doenças em plantas economicamente importantes como o milho e a cana do açúcar, e estão classificados no subfilo Ustilaginomycotina (uma das três linhagens principais de basidiomicetas). No entanto, as várias espécies de *Ustilago* estão filogeneticamente relacionadas com espécies do género *Pseudozyma* que são consideradas saprófitas e assexuadas. Neste trabalho, o estudo centrado na identificação dos genes que determinam a identidade sexual em fungos (genes *MAT* ou *mating type*) mostrou que *Pseudozyma prolifica* é uma espécie sexuada, fitopatogénica e biologicamente indistinguível de *U. maydis*. Para as restantes espécies de *Pseudozyma*, a análise molecular do gene *PRF1*, que codifica para uma proteína com um papel central na regulação do desenvolvimento sexual em *U. maydis*, não revelou evidências substanciais de perda de reprodução sexuada neste grupo. No entanto, alguns resultados sugerem que algumas espécies de *Pseudozyma* poderão estar a adaptar-se a um modo de vida saprofítico.

A linhagem de basidiomicetas que terá divergido primeiro (subfilo Pucciniomycotina) compreende também fitoparasitas muito revelantes (ex. as ferrugens), bem como grupos unicamente constituídos por organismos saprofíticos. Entre os últimos estão incluídas as leveduras pigmentadas (*red yeasts*) classificadas na ordem Sporidiobolales e que, como modelo de estudo, têm a vantagem de completar o seu ciclo de vida em meio de cultura laboratorial. No entanto, este grupo manteve-se praticamente inexplorado no que diz respeito à caracterização dos sistemas que determinam compatibilidade sexual (sistemas *MAT*), à identificação de genes *MAT* e às relações evolutivas entre as várias espécies sexuadas e assexuadas que abrange. Uma análise extensiva de mais de 200 estirpes pertencentes a 32 espécies de Sporidiobolales sugere que as linhagens asexuadas têm origem a partir das linhagens sexuadas, mas que não perduram tempo suficiente para formarem espécies estritamente assexuadas desprovidas de genes *MAT*. Uma investigação mais aprofundada nas leveduras *Rhodosporidium toruloides* e *Sporidiobolus salmonicolor* permitiu, pela primeira vez em Pucciniomycotina, a identificação do conjunto completo de genes *MAT*. A caracterização mais detalhada e multidisciplinar do sistema *MAT* em *S. salmonicolor* produziu resultados inesperados que levaram à proposta de um novo sistema substancialmente diferente dos dois paradigmas anteriormente existentes em fungos (sistemas bipolar e tetrapolar). Dado que o subfilo Pucciniomycotina é basal nos basidiomicetas, este novo sistema designado pseudo-bipolar constitui um contributo significativo para o estudo da evolução dos sistemas *MAT* em fungos.

Palavras-chave: basidiomicetas, reprodução sexuada, genes *MAT*, sistema pseudo-bipolar

Abstract

This work concerns the investigation of the molecular mechanisms of sexual reproduction in fungi and their possible implication for fungal lifestyles (parasitic vs. saprobic) and for the emergence of asexual fungal lineages.

The association between pathogenicity and sexuality is well-known in the basidiomycete plant parasite *Ustilago maydis* (subphylum Ustilaginomycotina), an economically important smut fungus. However, *Ustilago* species are phylogenetically interspersed with species of the genus *Pseudozyma*, which are considered saprobic and asexual. In this work, a study focused on genes involved in determining sexual identity (mating type or *MAT* genes), showed that *Pseudozyma prolifica* retains full sexual competence and pathogenicity, being therefore indistinguishable from *U. maydis*. For other *Pseudozyma* species, molecular analyses of *PRF1*, a gene that encodes a master regulator of sexual reproduction in *U. maydis*, showed no substantial evidence of loss of sexual reproduction. However, some clues were also found suggesting that some *Pseudozyma* species may be evolving towards a saprobic lifestyle.

The earliest derived lineage of Basidiomycota (subphylum Pucciniomycotina) includes also important plant pathogens (rust and anther smut fungi) as well as lineages composed solely of saprobic organisms. Among the latter, the red yeasts of the order Sporidiobolales have the advantage of completing their life cycle in culture media, but have remained very little explored concerning the characterization of mating systems, the identification of *MAT* genes and the evolutionary relationships between sexual and asexual species. A comprehensive analysis of more than 200 strains belonging to 32 species of the Sporidiobolales indicated that asexuality seems to originate frequently from sexual lineages, but does not seem to persist long enough to form truly asexual species devoid of *MAT* genes. A more in-depth investigation of the red yeasts *Rhodosporidium toruloides* and *Sporidiobolus salmonicolor* allowed the identification for the first time in the Pucciniomycotina of the complete set of *MAT* genes. A detailed and multidisciplinary characterization of the mating system in the latter species yielded surprising results. A novel mating system that differs substantially from the two mating paradigms in basidiomycetes, the bipolar and tetrapolar systems, was brought to light. Given the basal phylogenetic position of the Pucciniomycotina within the Basidiomycota, this new system designated pseudo-bipolar, constitutes a significant contribution to the study of the evolution of *MAT* systems in fungi.

Keywords: basidiomycetes, sexual reproduction, mating type genes, pseudo-bipolar system.

Abbreviations

ABC	ATP-binding cassette
AIC	Akaike information criterion
ATP	Adenosine triphosphate
bp	Base pair
cAMP	Cyclic adenosine monophosphate
CBE	Chlorazole Black E
cDNA	Complementary DNA
cM	centimorgan
CT	Cycle threshold
$CuSO_4$	Copper (II) sulphate
D1D2	Domains 1 and 2 of the 26S rDNA
DAPI	4',6-diamidino-2-phenylindole
DIC	Differential interference contrast
DMSO	Dimethyl sulfoxide
dN	non-synonymous substitutions per non-synonymous site
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dpi	Days post infection
dS	synonymous substitutions per synonymous site
HCl	Chloridric acid
HD	Homeodomain
HMG	High-mobility group
hr	Hour(s)
ITS	Internal transcribed spacer
IUPAC	International Union of Pure and Applied Chemistry
JGI	Joint Genome Institute
kb	Kilobase pair
KOH	Potassium hydroxide
LMP	Low melting point
LSU	Large subunit (ribosome)
MAP	Mitogen activated protein
MAPK	Mitogen activated protein kinase
MAT	Mating type
Mb	Megabase pair
MELs	Mannosylerythritol lipids
$MgCl_2$	Magnesium chloride
min	Minute(s)
ML	Maximum likelihood
$MnSO_4$	Manganese (II) sulphate
MP	Maximum Parsimony

MYA	Million years ago
MYP	Malt, Yeast extract and Peptone (Soytone) medium
NCBI	National Center for Biotechnology Information
OD_{640}	Optical density at 640 nm
ORF	Open reading frame
PAK	p21-activated kinase family
PCR	Polymerase chain reaction
PDA	Potato dextrose agar medium
PKA	Protein kinase A
PP	*Pseudozyma prolifica*
PREs	Pheromone response elements
rDNA	ribosomal DNA
RFLP	Restriction fragment length polymorphism
RHA	Rhodotorucine A
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rotations per minute
RT-PCR	Reverse transcription polymerase chain reaction
s	Second(s)
Tris	2-Amino2-hydroxymethyl-propane-1,3-diol
UAS	Upstream activating sequence
UV	Ultraviolet radiation
v/v	Volume per volume
w/v	Weight per volume
WGD	Whole-genome duplication
YEPS	Yeast extract, peptone and sucrose medium

Abbreviations of relevant culture collections

CBS	Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands
IAM	IAM Culture Collection, Institute of Molecular and Cellular Biosciences, The University of Tokyo, Japan (collection transferred to JCM).
JCM	Japan Collection of Microorganisms, Riken Bioresource Center, Saitama, Japan
NRRL	Agriculture Research Service (ARS) Culture collection, Peoria, Illinois, USA
PYCC	Portuguese Yeast Culture Collection, CREM, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
VKM	All-Russian Collection of Microorganisms, Pushchino Biological Research Center, Moscow, Russia.

Abbreviated species names

C. albicans	*Candida albicans*
C. cinerea	*Coprinopsis cinerea*
C. gattii	*Cryptococcus gattii*
C. heterostrophus	*Cochliobolus heterostrophus*
C. neoformans	*Cryptococcus neoformans*
D. hansenii	*Debaryomyces hansenii*
K. lactis	*Kluyveromyces lactis*
P. anserina	*Podospora anserina*
P. microspora	*Pholiota microspora*
S. cerevisiae	*Saccharomyces cerevisiae*
S. commune	*Schizophyllum commune*
S. pombe	*Schizosaccharomyces pombe*
S. reilianum	*Sporisorium reilianum*

Abbreviations of frequently used generic names

M.	*Microbotryum*
N.	*Neurospora*
P.	*Pseudozyma*
R.	*Rhodosporidium*
Rh.	*Rhodotorula*
S.	*Sporidiobolus*
Sp.	*Sporobolomyces*
U.	*Ustilago*

Index

Chapter 1

General Introduction

Chapter 2

Evolutionary relationship between sexual and asexual species of Ustilaginales – The *Ustilago* / *Pseudozyma* system

Chapter 3

Identification of mating type (*MAT*) genes in the red yeast *Rhodosporidium toruloides*

Chapter 4

The pseudo-bipolar *MAT* system of the red yeast *Sporidiobolus salmonicolor*

Chapter 5

Evidence for maintenance of sex determinants but not of sexual stages in red yeasts

Chapter 6

Concluding remarks and future directions

Appendix I

Additional information pertaining to Chapters 2 and 3

Appendix II

Additional information pertaining to Chapter 4

Appendix III

Additional information pertaining to Chapter 5

Figure Index

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Appendix III

Table Index

1. CHAPTER

General Introduction

1.1. The importance of sexual reproduction

1.1.1. Evolutionary significance of sexual reproduction in eukaryotes

The origin and maintenance of sex and the reason why most living organisms have adopted a more costly and inefficient (sexual) mode of reproduction is still a mystery for evolutionary biologists. Sexual populations have a two-fold cost compared to asexual populations (Maynard-Smith, 1978). Firstly, a sexual population consists of two sexes and only one is capable of producing offspring. By contrast, in asexual populations, each individual can produce its own progeny, which implies that an asexual population has an inherent capacity to grow more rapidly each generation. Secondly, sexually reproducing diploid organisms will only pass on one-half of their genes to any given offspring because gametes are haploid, whereas in asexual reproduction, the full genome is transmitted from a parental individual to its progeny. Finally, another example of the cost of sex is that mating partners must search for each other to mate (Maynard-Smith, 1978).
Despite its apparent disadvantages, sexual reproduction is an almost universal trend in eukaryotes (Dacks and Roger, 1999; Ramesh et al., 2005), which suggests that it must also confer important benefits. More than a century ago, August Weismann was the first to advocate that the increased genetic variation among the offspring as a product of sex would allow more efficient natural selection and in this way accelerate evolution (Weismann, 1904). Other benefits of sexual reproduction are that sex and recombination serves to purge the genome of deleterious mutations (Muller, 1964) and to generate recombinant progeny better suited in an ever-changing environment (Kondrashov, 1993). In line with this, an increased genetic diversity as a product of recombination is assumed to raise the chance for enhanced fitness, therefore greatly surpassing the efficiency of mitotic reproduction (Barton and Charlesworth, 1998; Goddard et al., 2005). Hence, sex persistence underlines its importance both for individual fitness and long-term survival of species (Hadany and Comeron, 2008).

1.1.2. Sex and lifestyle in Fungi

However, evidence that the cost of sex can be effectively counterbalanced comes from the existence of species that are capable of both sexual and asexual reproduction, as many fungal species. While in fungi, sexual reproduction is usually observed in periods of environmental instability or in sub-optimal ecological conditions, asexual reproduction is preferred when conditions are more favourable. Therefore, the co-existence of sexual and asexual reproduction in fungi might be viewed as an indication that sex offers a selective advantage.
Members of the kingdom Fungi, although often unnoticeable, occur in most of Earth's environments. With over 100 000 described species, fungi are extremely successful and play important roles in

most ecosystems (Kirk et al., 2008). Most species are saprobes and efficient decomposers of organic matter, and others evolved symbiotic associations with plants. Furthermore, and of particular interest, while about 32% of fungi are plant pathogens, only about 0.5% are clinically significant human pathogens (Kirk et al., 2008). The majority of known fungi are included in the two sister phyla Ascomycota and Basidiomycota. The Ascomycota outsize the Basidiomycota and comprises most of the plant pathogens and about 90% of the human pathogenic fungi (Morrow and Fraser, 2009). In the Basidiomycota, although the largest clade includes saprobic species (the mushrooms), about one-third of the basidiomycetes (ca. 10 000 species) are plant pathogens (the rust and smut fungi) and only close to 40 species are known to cause infection in humans and animals (Kirk et al., 2008; Morrow and Fraser, 2009).

Sexual reproduction and meiosis have been well established for many fungal species and many sexual systems have served as invaluable eukaryotic models of development (Zarnack and Feldbrügge, 2007; Zarnack and Feldbrügge, 2010), transcription regulation (Gustin et al., 1998; Basse and Farfsing, 2006) and signalling pathways (Xue et al., 2008). Research in the field has also been prompted by the fact that for many plant or animal fungal pathogens, sexual reproduction is intimately linked with pathogenicity (Sexton and Howlett, 2006), a feature particularly important in the case of dimorphic pathogenic species. These species alternate between an asexual, single-celled growth form (yeast) and a sexual stage characterized by filamentous growth (mycelia). The asexual (mitosporic) and sexual (meiosporic) forms are also known as the 'anamoph' and the 'teleomorph', respectively (Seifert and Gams, 2001). Although dimorphic plant pathogens, such as the smut fungus *Ustilago maydis*, may hypothetically persist in the asexual (saprophytic) yeast phase, sexual reproduction and concomitant host infection are indispensable for completion of the life cycle (Figure 1.1) (Bakkeren et al., 2008; Nadal et al., 2008). On the other hand, dimorphic pathogens that cause infection in the yeast stage have been also documented, like the human pathogen *Cryptococcus neoformans*. Infection caused by this organism is thought to be mediated by inhalation of spores produced in the sexual stage (basidiospores), which indicates that a dimorphic transition from a saprophytic filamentous form (sexual stage) to a yeast form (asexual stage) is crucial to initiate the infection process (Heitman, 2006; Nielsen and Heitman, 2007; Heitman, 2010) (Figure 1.1).

While sexuality is obligatory in some groups of fungi (e.g. mushrooms), the production of asexual spores is also a common feature throughout the fungal world and for about 17% of the known fungal species sexual cycles remain to be discovered (Hawksworth et al., 1995). In fact, for many animal fungal pathogens, sexual reproduction has rarely or never been observed as part of their life cycle (Sexton and Howlett, 2006; Morrow and Fraser, 2009), leading to the hypothesis that asexual reproduction is favoured by allowing pathogens to proliferate more rapidly in their hosts. It is possible that fungi take advantage of both sexual and asexual reproduction, which implies that the cost of sex in fungi is probably lower than the cost of sex in plants and animals. One interesting idea in line with this considers that the timing of sexual reproduction could have been adjusted by natural selection to minimize its costs, namely when the conditions are unfavourable for vegetative growth

and asexual spore production (Bell, 1982). These in turn are the conditions when the production of sexual survival spores is important to produce genetically variable offspring to maximize success in face of changed environmental conditions (Bell, 1982; Aanen and Hoekstra, 2007). In light of this, confirmation of strict asexuality is an invaluable information (although often difficult to ascertain), particularly when several lines of evidences suggest that asexuality is a derived character among fungi (LoBuglio et al., 1993).

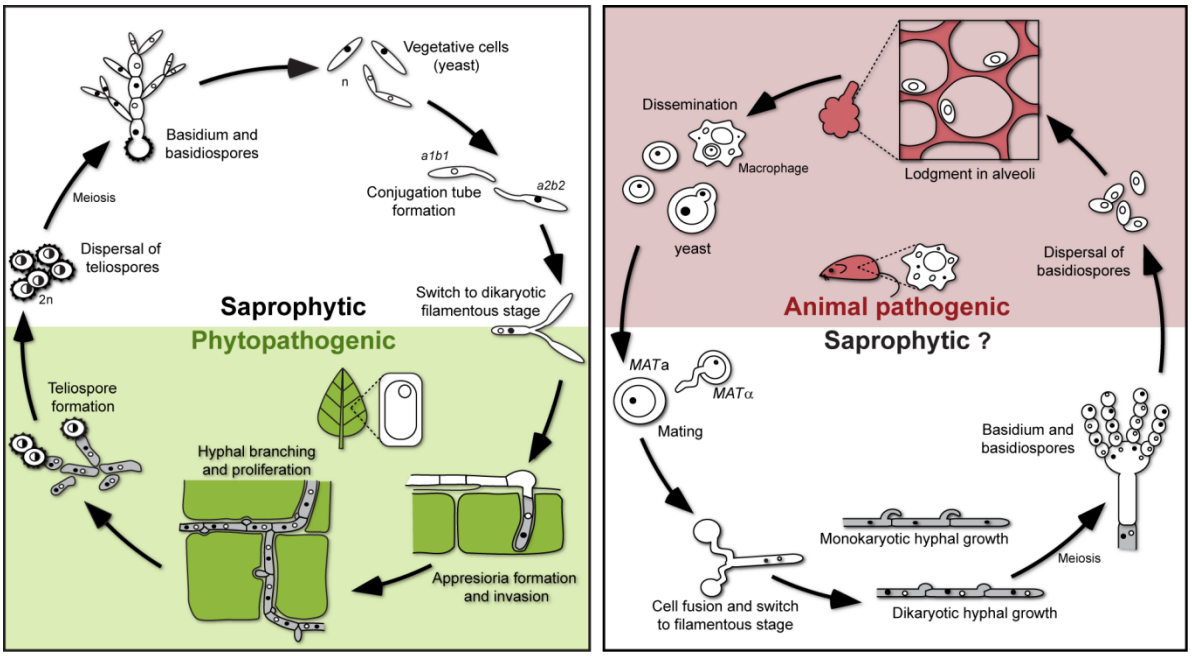

Figure 1.1. Schematic representation of the life cycle of the smut fungus *U. maydis* (on the left) and of the human pathogen *C. neoformans* (on the right). The yeast form in *U. maydis* is non-pathogenic and upon mating with a compatible partner switches to an infectious filamentous form that can invade the plant host. In *C. neoformans*, a similar dimorphic switch originates the potentially infectious propagules (basidiospores) to infect an animal host (adapted from Morrow and Fraser, 2009).

1.1.3. Genomic signatures of sex

Historically, fungal species have been differentiated by a variety of methods and concepts. Classification based on morphological characteristics, such as the size and shape of spores or fruiting body structures, has traditionally prevailed in fungal taxonomy and is still important for species recognition in some groups of fungi. The biological species concept discriminates species based on their ability to interbreed and of being reproductively isolated from other groups. However, since fungi may reproduce both sexually and asexually, and usually not at the same time, a single species can be isolated in one stage or the other which complicates its identification. Moreover, a commonly encountered situation is that of some groups of fungi with an asexual stage but for which

no sexual stage has ever been observed, implying that sex is either absent, very rare or difficult to observe in nature.

The classical and most unambiguous approach used in looking for evidence of sex is the direct observation of biological mating ability with production of viable and fertile progeny. However, failure to induce sexual reproduction in the laboratory may due to sub-optimal culture conditions rather than inability to mate, which implies that a suitable mating scenario and mating partners must be identified for a successful mating. To circumvent this problem, recent advances in DNA sequencing and genome-wide studies allowed the application of molecular analyses to access the degree of sexual recombination in natural populations as well as to search in genomes for genes involved in sexual reproduction and meiosis. Combining these strategies, convincing evidence exist that asexual fungal species have derived from sexual ancestors (LoBuglio et al., 1993; Anderson and Kohn, 1995), and that the majority of the extant presumed asexual species have phylogenetically close relatives capable of sexual reproduction (Kurtzman and Robnett, 1998; Begerow et al., 2000). Therefore, genuinely and long-standing asexual fungal lineages are indeed rare but, in some cases, asexual reproduction seems to contribute as a significant component to the genetic structure of a species population (Fisher et al., 2005). Nevertheless, even in such clonal populations genetic footprints of recombination may still be detected (Taylor et al., 1999; Fisher et al., 2005).

Of special interest, genomic approaches have been recently used with success to detect genes required for sexual reproduction and indirectly recognize sexual reproduction in some animal pathogenic species (viz. *Candida albicans* and *Aspergillus fumigatus*), which so far were believed to propagate exclusively by asexual reproduction (Tzung et al., 2001; Paoletti et al., 2005). These approaches are based on the assumption that if a process has been lost due to a fitness cost to maintain an unused trait, a reduced selective functional constraint for the genes necessary for that function is expected, and they should accumulate deleterious mutations, eventually becoming pseudogenes (Lynch and Conery, 2000; Rouquier et al., 2000). Using this reasoning, the presence of intact, non-degenerate, genes that are specifically involved in mating and sexual reproduction is a good evidence of capacity of sex (Schurko and Logsdon, 2008).

Hence, the distribution and interpretation of reproductive behaviours in fungi illustrates the difficulties to understand sexuality in evolutionary terms. In this context, discovering whether or not fungi undergo sexual reproduction is important to understand their evolution. Although the identification of fungal sexual structures remains an important and precise way to effectively recognize sex, the predictive power of genomics and population genetics is providing new insights into the reproductive modes of fungi, particularly on their inherent capacity for sex.

1.2. Mating and sexual identity in Fungi

In fungi, sexual identity is always established in the haploid stage (Billiard et al., 2011) and determined by specialized genomic regions called mating type (*MAT*) loci (Fraser and Heitman, 2005). Results of several studies have revealed that fungal *MAT* loci vary extensively in length and in genetic content, which is further translated into a remarkable diversity of mating strategies and sex-determining systems (Casselton, 2002; Fraser and Heitman, 2004; Lee et al., 2010b). Most fungal species capable of sexual reproduction are heterothallic (self-sterile) but others are homothallic (self-fertile) (Bakkeren and Kronstad, 1994). In heterothallic species, mating occurs between two sexually compatible haploid individuals that are morphologically identical but are distinguished genetically at the *MAT* locus. In homothallic species, each individual can mate with itself and several genetic mechanisms have been shown to form the basis for this sexual behaviour (Hicks and Herskowitz, 1976; Lin et al., 2005; Alby et al., 2009; Rodriguez-Carres et al., 2010). The underlying genetic mechanisms of some forms of homothallism are introduced in section 1.2.3.

1.2.1. Mating in Ascomycota

In most species within the phylum Ascomycota, sexual compatibility is determined by short DNA sequences located at a single *MAT* locus encoding key transcription factors that govern cell type identity and developmental fate (Butler, 2007). This molecular architecture was first determined in the model yeast *Saccharomyces cerevisiae*, where it was shown that two cell types (*a* and α strains) could be distinguished by different sequence information present at the *MAT* locus (Astell et al., 1981). These observations were crucial towards the elucidation of the molecular foundations of the so-called bipolar mating behaviour, which is characterized by the existence of two sexes or mating-types. Subsequent analysis of *MAT* loci in other ascomycetous species led to the identification and definition of the *MAT* 'idiomorphs' (instead of 'alleles') to denote unique sequences unrelated in structure but located at the same homologous locus in sexually compatible individuals (Glass et al., 1988; Metzenberg and Glass, 1990). Thus, the *MAT* idiomorphs differ to the extent that there is no sequence similarity between them (Glass et al., 1988; Lee et al., 1999). Recent advances in genome sequencing uncovered different genomic structures of the *MAT* locus from several species included in the two major subphyla of the Ascomycota (Pezizomycotina and Saccharomycotina), indicating that the extent of dissimilarity can vary across species (Butler, 2007). For the third sub-phylum (Taphrinomycotina) information about *MAT* loci is limited to the well known fission yeast *Schizosaccharomyces pombe* (Butler, 2007). While in *S. cerevisiae* and in related members of the Saccharomycotina the region of dissimilarity only encompasses several hundred base pairs, in filamentous ascomycetes (Pezizomycotina) this region is extended, encompassing several kilobases (Coppin et al., 1997).

1.2.1.1. Mating type (*MAT*) genes of ascomycetes

In ascomycetes, *MAT* loci encode transcription factors belonging to the α-domain, homeodomain (HD) or high-mobility group (HMG) families that regulate determinants of sexual identity and development, like peptide pheromones and pheromone receptors belonging to the G protein-coupled receptor family (Fraser and Heitman, 2003). The different sets of *MAT* genes have been lost or recruited as sex-determinant in different lineages of the ascomycetes and most of the known architectures are summarized in Figure 1.2. In general, the *MAT1-1* idiomorph (also designated as *MATα* or *MTLα*) is defined by the invariable presence of a gene encoding a protein with a α-domain motif (*MAT1-1-1* or *MATα1* or *MTLα1*), whereas the *MAT1-2* idiomorph (also known as *MATa* or *MTLa*) is usually established by a gene encoding a transcription factor with a HMG domain (*MAT1-2-1*, *MATa2* or *MTLa2*). A recent study showed that *MATα1* and *MATa2* (henceforth designated as *α1* and *a2*) genes are present in opposite mating types of almost all ascomycetes and that *α1* genes have most likely evolved from an ancestral HMG domain gene (Martin et al., 2010). These observations further support the hypothesis that alternate HMG proteins were the basis of the ancestral sex-determining system given that they are also found as integral component of *MAT* in filamentous ascomycetes and in the sex-related locus of basal fungal lineages (e.g. Mucoromycotina and microsporidia) (Idnurm et al., 2008; Gryganskyi et al., 2010; Lee et al., 2010a). Interestingly, the *a2* gene, coding for an HMG domain protein, has been recently lost in species included in the clade of *S. cerevisiae* (Butler et al., 2004), which represents a notable exception to this paradigm (Figure 1.2a). Therefore, the *MAT* locus of *S. cerevisiae* is not representative of other ascomycetes and in this clade the role of HMG genes in the determination of sexual identity has been partially replaced by HD transcription factors (a1 and α2). Among ascomycetes, the HD proteins were also identified and are known to play key roles in mating in other species of the Saccharomycotina (Stanton and Hull, 2007) and in *S. pombe* (*matPi*) (Kelly et al., 1988; Stanton and Hull, 2007) (Figure 1.2a).

1.2.1.2. Cell identity in *Saccharomyces* and *Candida* clades (Saccharomycotina)

The mating type determination circuit has been thoroughly elucidated in *S. cerevisiae* and served as the basis to explore cell identity control systems of other fungi (Herskowitz et al., 1992). In *S. cerevisiae*, cells can reproduce either asexually (clonally) by haploid cell budding or sexually by mating followed by meiosis. Three cell types exist in *S. cerevisiae*: a, α and a/α (diploid), and the identity of each cell type is specified by the activities of transcription factors encoded by each *MAT* idiomorph (Herskowitz et al., 1992). In a-cells, the *a1* gene present at the *MATa* locus encodes an HD2 class homeodomain (HD) transcription factor, which plays no role in defining the a-haploid cell type. On the other hand, in α-cells, *α1* and *α2* genes encode the α-domain transcription factor α1 and the HD1 class transcription factor (α2) that regulate the expression of α- and a-specific genes,

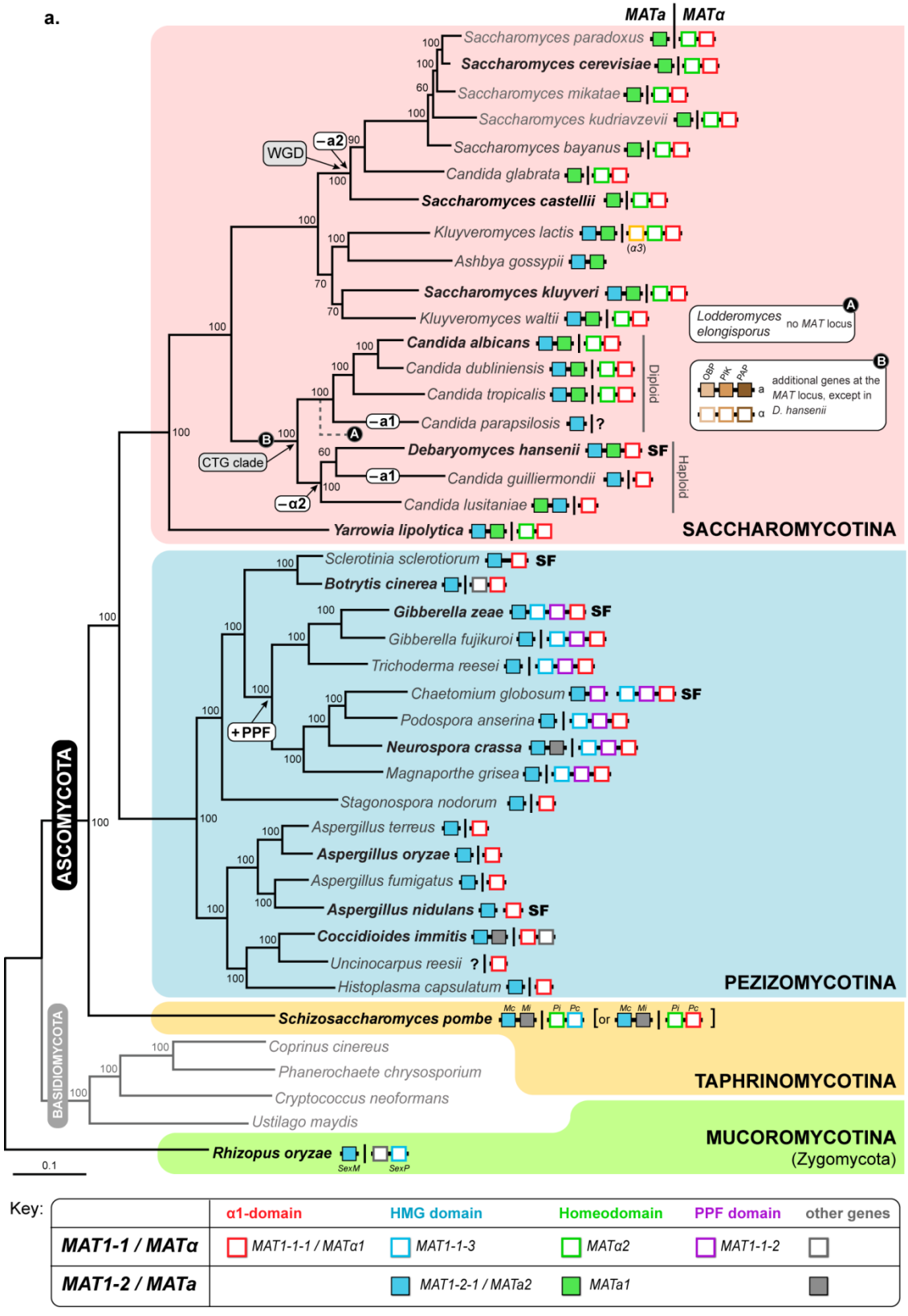

Figure 1.2. Structure and diversity of the *MAT* loci in the different ascomycete lineages. Description is given in page 10.

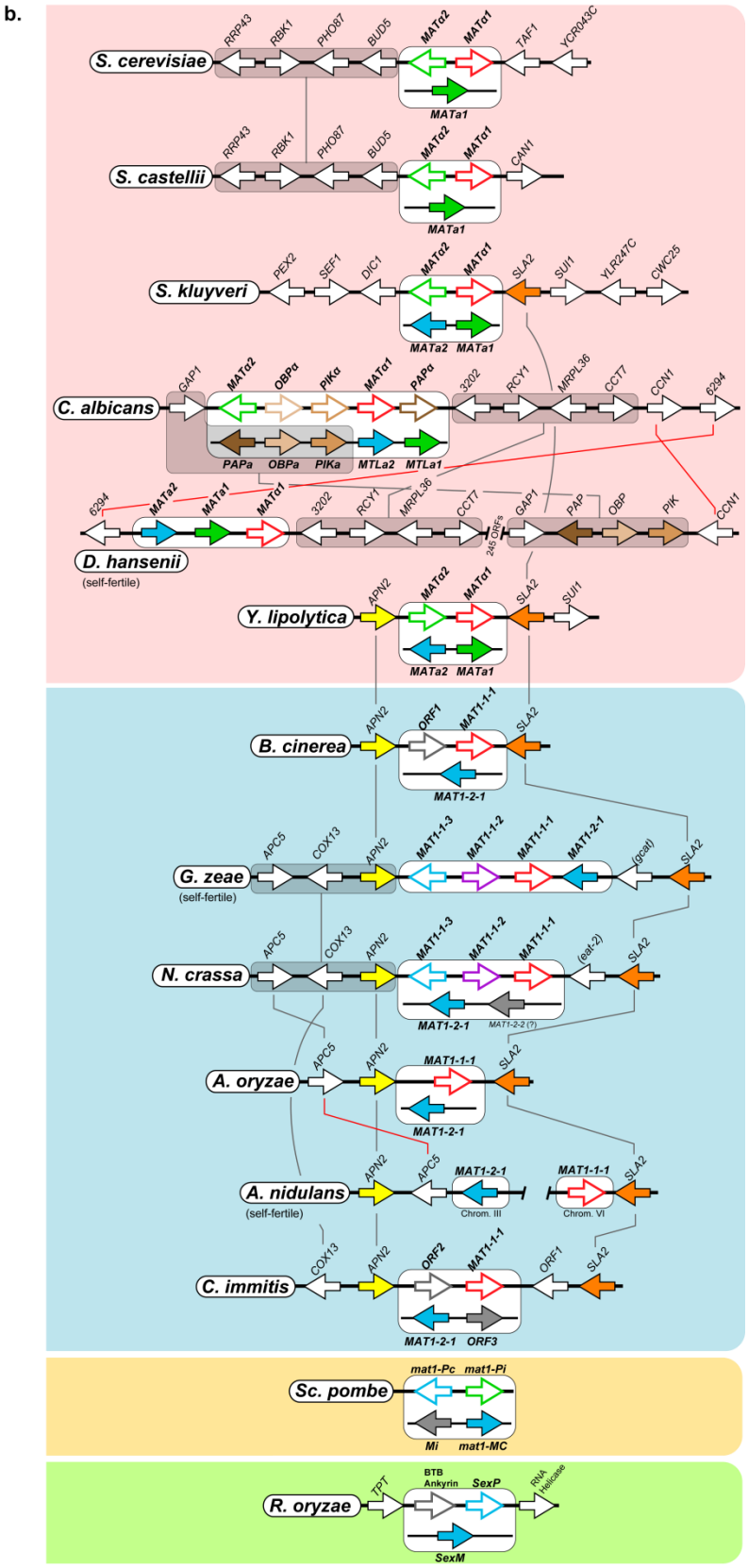

Figure 1.2. Continued. Description is given in page 10.

Figure 1.2. Structure and diversity of the *MAT* locus in the different ascomycete lineages. (a) Distribution and organization of the different sets of *MAT* genes in representatives of the three main lineages of the phylum Ascomycota – Saccharomycotina, Pezizomycotina and Taphrinomycotina (shadowed in pink, blue and yellow, respectively) – and in an emblematic species of the Mucoromycotina (shadowed in green). The phylogenetic tree was adapted from (Fitzpatrick et al., 2006). For each species, the gene content of both *MAT* idiomorphs (*MATa* and *MAT*α, separated by a vertical line) is represented as given in the key at the bottom. Genes connected by a black horizontal line are genetically linked. In self-fertile species ("SF"), genes are located either at a single genomic region or at different regions. White-rounded boxes in the tree indicate instances where genes were acquired to (+) or loss from (–) the *MAT* locus. In the CTG clade, the black circle symbolized by "A" in the tree indicate the relative placement of the species *Lodderomyces elongisporus* for which no *MAT* homologues were identified even using whole-genome information (Butler et al., 2009); the black circle "B" indicate the most likely time point when three additional genes, without known functions in mating (OBP, PIK and PAP) were recruited to the *MAT* locus [their organization is shown for *C. albicans* in (b)]. In *K. lactis* an additional gene (*α3*) is present at the *MAT*α idiomorph and codes for a transposase with an essential role in mating type switching from *MAT*α to *MATa* (Barsoum et al., 2010). For *S. pombe* it is still uncertain if the *mat-Pc* gene codes for an HMG or α1 domain protein (Martin et al., 2010). **(b)** Detailed genomic organization of the *MAT* idiomorphs of the species highlighted in boldface in the tree. Genes are depicted as arrows indicating the direction of transcription. For each species, alternate *MAT* idiomorphs are enclosed in white-rounded boxes and *MAT* genes are coloured as indicated in the key. Additional *MAT* genes in *C. albicans* and their non-*MAT* orthologues in *Debaryomyces hansenii* are coloured in the same colour. Conserved genes *SLA2* and *APN2* flanking the *MAT* idiomorphs are coloured in orange and yellow respectively. All the other flanking genes are shown in white. Orthologous genes or gene clusters (shadowed) are connected by gray or red lines if oriented in the same or opposite direction, respectively. (Based on Galagan et al., 2005; Butler, 2007; Stanton and Hull, 2007; Butler, 2010; Gryganskyi et al., 2010).

respectively (Herskowitz et al., 1992) (Figure 1.3). Mating occurs when a-cells locate α-cells via the production of compatible receptors and pheromone ligands, which are under control of the *MAT* genes but are not mating type-specific (Sprague and Thorner, 1992). For example, in α-cells α1 upregulates the a-factor pheromone receptor (*STE3*) and the α-factor pheromone genes (*MFα1* and *MFα2*), whereas the α2 transcription factor together with the MADS box transcription factor Mcm1, restrict the expression of the a-specific genes including the α-factor pheromone receptor (*STE2*) and the a-factor pheromone genes (*MFA1* and *MFA2*) (Figure 1.3) (Herskowitz et al., 1992; Tsong et al., 2007). Conversely, in a-cells, a-specific genes are constitutively activated by Mcm1 and thus the a-cell is the "default" type (Hwang-Shum et al., 1991). Upon compatible responses between mating partners, cells fuse and subsequent karyogamy occurs to form a diploid cell. In the diploid, a1 and α2 homeodomain proteins form a heterodimer that functions as a transcriptional repressor of haploid-specific genes (Galgoczy et al., 2004) (Figure 1.3). Hence, diploid cells cannot mate and alternatively undergo meiosis to generate haploid meiotic progeny (two α and two a cells).

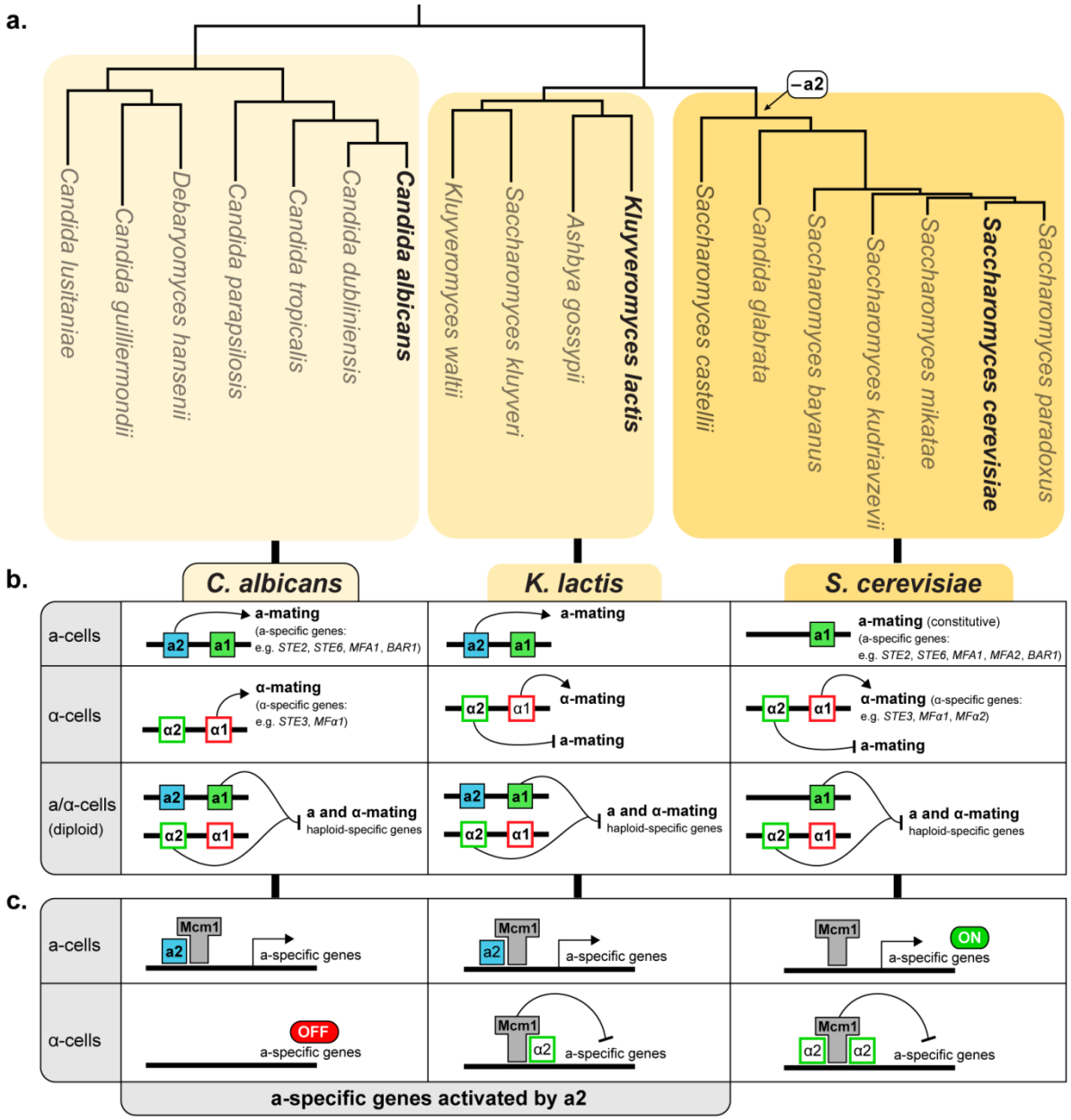

Figure 1.3. Mating-type regulation pathways in three ascomycete yeasts. (a) Phylogenetic tree showing representative species of the Saccharomycotina (based on Figure 1.2). For the species highlighted in boldface, the mating type transcriptional pathways are shown in detail in (b) and (c). Loss of the *a2* gene in the Saccharomyces clade is indicated in the tree. **(b)** For each species and cell type (a-cells, α-cells and a/α diploid) is shown which set of genes activate/ repress a- and/or α-regulated pathways. **(c)** Activation/ repression of a-specific genes is shown for the three species. In *C. albicans* and *K. lactis,* a-specific genes in a-cells are positively regulated by the *a2* gene while in *S. cerevisiae* they are "ON" by default. In α-cells, a-genes are negatively regulated by the *α2* in *S. cerevisiae* and in yeasts of the *K. lactis* lineage, but not in *C. albicans* where a-specific genes are "OFF" by default.

Although *S. cerevisiae* has served as an archetypal fungal mating system for decades, it is remarkable, that in *S. cerevisiae* and related species of the same clade the *a2* gene has been evicted (Figure 1.2 and 1.3). Only recently this question was addressed and it has been suggested that the loss of the *a2* gene in this group may be a consequence of the development of an alternative transcriptional pathway for a-specific gene expression, yet leading to an identical result (Tsong et al., 2006; Tsong et al., 2007). This circuit seems to have evolved from an ancestor transcription circuit as found in *C. albicans* and most likely through a transitional state still observed in *Kluyveromyces lactis* given that both species have maintained the *a2* gene at the *MATa* locus (Figure 1.3) (Tsong et al., 2006; Tsong et al., 2007). The most striking difference regarding the a-specific gene expression in *C. albicans*, *K. lactis* and in *S. cerevisiae* is that in the first two species, a-specific genes in a-cells are positively regulated by the a2 transcription factor, while in *S. cerevisiae*, a-specific genes are "on" by default, since activation only requires Mcm1, which is constitutively expressed. Moreover, to restrict the expression of a-specific genes in α-cells, these genes are repressed by α2 (together with Mcm1) in both *S. cerevisiae* and *K. lactis*, but not in *C. albicans* since a-specific genes are "off" by default in α-cells. Therefore, *K. lactis* activates and represses a-specific genes depending on the cell-type identity, which suggests an intermediate situation from those observed in *C. albicans* and *S. cerevisiae*. Furthermore, mapping *MAT* loci and functional evidence in a phylogenetic context indicate that activation of a-specific genes by an a2 protein represents the ancestral mode of regulation as observed in *C. albicans* (Figure 1.2 and 1.3).

1.2.1.3. Organization of *MAT* locus in filamentous ascomycetes (Pezizomycotina)

In addition to budding yeasts, the *MAT* loci of many filamentous ascomycetes have been characterized revealing considerable variation in structure and gene content. Heterothallic species of this group have a single *MAT* locus with two idiomorphs usually designated as *MAT1-1* and *MAT1-2* (Turgeon and Yoder, 2000) that correspond to *MAT*α and *MATa* idiomorphs of *S. cerevisiae*, respectively. The most simplistic architecture is defined by the presence of the *MAT1-1-1* gene (encoding a α-domain protein) in the *MAT1-1* idiomorph, while the opposite idiomorph (*MAT1-2*) houses the *MAT1-2-1* gene (encoding a HMG- domain protein). This organization is found in the corn blight fungus *Cochliobolus heterostrophus* (Turgeon et al., 1993) and its close relatives (e.g. *Stagonospora nodorum*, Figure 1.2a). A somewhat similar organization is also found in most heterothallic species of *Aspergillus* and in some dermatophyte and dimorphic pathogenic fungi (Butler, 2010; Li et al., 2010). Nevertheless, other genes may also be part of the *MAT* idiomorphs but only limited information exists about their function. For example, an additional gene designated *MAT1-1-2*, was found in the *MAT1-1* idiomorph of some filamentous fungi (viz. *Neurospora crassa*, *Magnaporthe grisea*, *Podospora anserina*). This gene encodes a protein without a known DNA-binding motif but has instead two invariant proline and one-phenylalanine residues, the so-called

PPF domain (Figure 1.2). This more complex *MAT* configuration, the classes of proteins that it contains and their ultimate functions were, nevertheless, expected to be similar to those of *C. heterostrophus*, where alternate *MAT* genes (*MAT1-1-1* and *MAT1-2-1*) are required for sexual development (Leubner-Metzger et al., 1997). However, an exception to this rule has been recently found in *Sordaria macrospora* (a species closely related to *P. anserina*, Figure 1.2a) where both *MAT1-1-1* (α-domain) and *MAT1-1-3* (HMG) appear not to play a role in vegetative growth or sexual reproduction (Klix et al., 2010). On the other hand, the *MAT1-1-2* gene (encoding a protein with a PPF domain) is essential for fruiting body and spore development (Klix et al., 2010). Hence, these "additional" genes may have important functions during mating and in fertility in species where they have been recruited to the *MAT* locus (Ferreira et al., 1998; Turgeon and Debuchy, 2007).

Whereas the *MAT* locus structure, gene number and arrangement vary in each ascomycete lineage, the position of *MAT* is unexpectedly well conserved. In general, the *MAT* locus is situated between homologues of the *APN2* (encoding a DNA lyase) and *SLA2* (cytoskeleton assembly control) genes (Figure 1.2b). Since both genes are also flanking the *MAT* locus in *Yarrowia lipolytica*, and the *SLA2* is also found at the right border of *MAT* in *Saccharomyces kluyveri*, *K. lactis* and *Pichia angusta*, (members of the Saccharomycotina, Figure 1.2b), this may represent the ancestral configuration for all ascomycetes (Butler et al., 2004). However, within the group of yeasts that have undergone whole-genome duplication (WGD) (*S. cerevisiae* and relatives) and in the *Candida* lineage (CTG clade), the genes surrounding the *MAT* locus are different (Figure 1.2) probably due to a marked genome shuffling.

1.2.1.4. *Neurospora tetrasperma* as a model for evolution of sex chromosomes

In fungi, although no straight relationship between the size of the *MAT* locus and mating ability has been reported so far, a larger *MAT* locus is hypothesized to assist the suppression of recombination of this unique region. Animal and plants have evolved various sex-determination systems but the most prominent is the acquisition of size-dimorphic sex chromosomes. This system has emerged independently many times by suppression of recombination around the sex-determining genes leading to differentiation and degeneration of the non-recombining regions of the sex chromosomes in successive steps (Charlesworth et al., 2005). There is also evidence from several taxa (mammals, birds and plants) that more than a single event of cessation of recombination can occur between sex chromosomes bringing forth chromosomal regions with different levels of sequence divergence, commonly known as "evolutionary strata" (Lahn and Page, 1999). Interestingly, complete recombination cessation in the region around the *MAT* locus has also been reported from several fungal species (Merino et al., 1996; Lee et al., 1999; Lengeler et al., 2002) but, in contrast to animal or plant dimorphic chromosomal regions, fungi usually have much smaller regions of suppressed recombination (Lengeler et al., 2002).

The *MAT* locus of the filamentous ascomycete *Neurospora tetrasperma* displays a similar organization and gene content as the closely related species *N. crassa* (Figure 1.2b). However, a remarkable feature exists in *N. tetrasperma*, since the chromosome harbouring the *MAT* locus fails to recombine over the majority of its length (~7 Mbp, or 75% of the chromosome) during meiosis, but it is flanked by normally recombining regions, analogous to the "pseudoautosomal" regions of the human sex chromosomes (Charlesworth et al., 2005; Menkis et al., 2008). Recently, comparative analyses of sequence divergence among alleles of several *MAT*-linked genes in different lineages of *N. tetrasperma* (*sensu lato*) indicate that the large region of suppressed recombination varies in length and gene content (Menkis et al., 2010). Additionally, the events leading to the suppression involved a stepwise cessation of recombination, resulting in two evolutionary strata (Menkis et al., 2010). Although the genetic mechanisms underlying these events are still unknown, they most likely include large genomic inversions and translocations that preclude pairing and crossing over during meiosis. Furthermore, the accumulation of transposable elements in these regions, although not yet explored in *N. tetrasperma*, was proposed to promote a progressively expansion of the non-homologous regions between *MAT* chromosomes (Fraser and Heitman, 2004). Finally, several events of gene conversion were detected between *MAT* alleles during the evolutionary history of this species (Menkis et al., 2010), which might provide an efficient mechanism for gene editing in the absence of recombination (e.g. the removal of deleterious mutations). The forthcoming release of the genome sequencing project of the opposite mating type of *N. tetrasperma* should help to clarify which factor(s) contribute to constrain recombination in this region.

1.2.2. Mating in Basidiomycota

Despite the enormous diversity in fungal sexual behavior, it is quite remarkable that it relies on two underlying sexual paradigms: the bipolar and the tetrapolar mating-type systems (Fraser and Heitman, 2003). Whereas the bipolar configuration is common to all species of the phylum Ascomycota, (see 1.2.1) it is less widespread throughout the Basidiomycota, where it is the tetrapolar system that occurs most frequently (Morrow and Fraser, 2009). At variance with the bipolar system where only two mating types segregate during meiosis (known as *a* and α or *MAT A1* and *MAT A2*) and are determined by genes (alleles or idiomorphs) at a single *MAT* locus, in the tetrapolar system four mating types are originated after meiosis given that two unlinked genomic regions (*MAT* loci) are determining sexual identity. Tetrapolarity has been found in at least two of the three major lineages of basidiomycetes (the sub-phyla Ustilaginomycotina and Agaricomycotina) (Morrow and Fraser, 2009) and, in the third lineage (the subphylum Pucciniomycotina), although a few reports indicate that some members may present a tetrapolar system (Yamazaki and Katsuya, 1988; Narisawa et al., 1994), in none of these the mating system has been thus far characterized at the molecular level. Therefore, studying the mating systems of basidiomycetes may provide further

knowledge into the evolutionary advantages or consequences of more than two mating types or sexes.

In a typical heterothallic basidiomycete species, mating of compatible haploid partners – homokaryotic hyphae or haploid yeast cells – originates a dikaryotic filamentous stage in which the two parental nuclei are replicated in a coordinated fashion without fusion. Karyogamy occurs in the basidia or in other specialized structures (e.g. teliospores), and basidiospores are produced after meiosis (meiospores) to restore the cycle (Hibbett et al., 2007) (Figure 1.1). Although the dikaryotic phase predominates as the vegetative growth form for many basidiomycetes (e.g. the mushrooms), it is of particular importance in pathogenic species like *U. maydis* in which the dikaryons are the infectious cell type (Banuett, 1995; Brefort et al., 2009).

In contrast to the genomic regions that determine the mating type in ascomycetes, the *MAT* loci of basidiomycetes have evolved a tremendous level of diversity concerning their gene content and organization. Nevertheless, two common and fundamental components regulate two independent molecular mechanisms of self/nonself recognition: one locus encodes pheromones and pheromone receptors (the “pheromone/ receptor locus”) and the other encodes homeodomain transcription factors (the “homeodomain locus”) (Fraser et al., 2007). For a successful mating and completion of the sexual cycle, cells that conjugate must contain different alleles of both *MAT* loci (Fraser et al., 2007; Morrow and Fraser, 2009).

1.2.2.1. The pheromone / receptor sensing system in basidiomycetes

Fungi have evolved sophisticated mechanisms to sense and respond to a variety of environmental signals, including defined substances that mediate the communication between individuals of the same species – the sexual pheromones (Karlson and Luscher, 1959; Xue et al., 2008; Raudaskoski and Kothe, 2010). To attract their mate, most fungi use mating-type specific pheromones that are secreted from a cell-type and are sensed by cells of the opposite mating type, leading to cell fusion. In basidiomycetes, this means that at the genetic level, partners must carry different alleles of the pheromone/ receptor locus that exhibits two components of the pheromone-mediated interaction system: (i) the *STE3*-like genes that encode pheromone receptors of the G protein-coupled seven-transmembrane receptor family (GPCR or 7TM receptors) and (ii) genes that encode small lipopeptide pheromone precursors (Bölker and Kahmann, 1993; Xue et al., 2008; Raudaskoski and Kothe, 2010) (Figure 1.4). An interesting aspect of the basidiomycetous pheromone/ receptor system is that only homologues of genes encoding the a-factor (pheromone) and Ste3 (receptor) of the ascomycetous system (see 1.2.1.2) have been retained (Fraser et al., 2007). Therefore, instead of the coupled sensing system typical of the ascomycetes (a-factor/Ste3 and α-factor/Ste2), cell recognition and fusion in basidiomycetes is mediated by allelic versions of the same genes (Casselton and Olesnicky, 1998). However, it is not known yet which of the two pheromone/ receptor systems (Ste2/a-factor/Ste3/α-factor or allelic variants of Ste3/α-factor) is the ancestral form.

It is interesting that all basidiomycete mating pheromones so far described, including the pheromones from the human pathogen *C. neoformans* (Moore and Edman, 1993) and those from the red yeast *Rhodosporidium toruloides* (Kamiya et al., 1978a), are hydrophobic diffusible lipopetides that contain farnesylation signals equivalent to those found in the a-factor of *S. cerevisiae* (Bölker and Kahmann, 1993). These active proteins are initially synthesized as larger precursors corresponding to the translated amino acid sequence, which subsequently undergoes posttranslational modifications. Pheromone maturation occurs by sequential events involving the carboxyl-terminal sequence CAAX (in which C is cystein, A is an aliphatic amino acid, and X is any residue) (Bölker and Kahmann, 1993). The mature pheromone is finally exported from the cell via a mechanism that involves an ATP-binding cassette (ABC) transporter (Ste6) that has been identified in both ascomycetes and basidiomycetes (Michaelis et al., 1992; Hsueh and Shen, 2005). In response to the interaction of these ligands with a suitable receptor, mating is initiated through a heterotrimeric G protein located at the plasma membrane that activates the downstream processes including a MAP kinase cascade (MAPK) signalling. This signalling pathway as well as the underlying mechanisms of a-factor processing and secretion were firstly characterized in *S. cerevisiae* and seem to have been retained in basidiomycetes (Fowler et al., 1999; Hegner et al., 1999; Olesnicky et al., 1999; Hsueh and Shen, 2005). Although the genetic and molecular mechanisms after pheromone perception are best understood in *S. cerevisiae*, significant advances have been made towards the elucidation of the same processes in basidiomycetes. These mechanisms are known in quite some detail in *U. maydis* and are described in the Chapter 2.

1.2.2.2. Homeodomain transcription factors – the second compatibility checkpoint

In basidiomycetes, following fusion of compatible partners, the progression through the sexual cycle depends on a second compatibility hurdle that is mediated by the other *MAT* locus. Analysis of this locus in model basidiomycetes revealed that it encodes, exclusively, transcription factors with a homeodomain DNA-binding motif (Figure 1.4). Therefore, the homeodomain (*MAT*) locus of basidiomycetes differs from the *MAT* locus of the ascomycetes in that, in the latter group other transcription factors (HMG or α-domain proteins) are also found as integral components of the *MAT* locus. The homeodomain is generally defined as a short amino-acid sequence that characterizes a large family of transcription factors that are ubiquitous in eukaryotes and play important roles in development, as e.g. the *Hox* genes of vertebrates that specify body segmentation (McGinnis and Krumlauf, 1992). In its elementary configuration, the homeodomain locus is composed of a pair of divergently transcribed genes encoding proteins of different homeodomain classes, one HD1 and one HD2. Remarkably, the heterodimeric homeodomain transcription factor is only active after cell fusion because dimerization is restricted to subunits (HD1 and HD2) that originate from individuals bearing different alleles of the homeodomain locus (Banham et al., 1995; Kämper et al., 1995). The

functional heterodimeric transcription factor is assumed to bind unique target sites within the promoter regions of genes whose function commits cells to a new developmental pathway. Although the homeodomain-regulated pathway is not yet defined in most basidiomycetes, it is indispensable during the latter stages of sexual development (Grandel et al., 2000; Kües et al., 2002; Wahl et al., 2010).

A final requirement of this elaborate system is that HD1 and HD2 proteins from the same cell type must not be allowed to heterodimerize or, otherwise, the homeodomain-regulated pathway would be constitutively active without mating. Several studies have reported that discrimination between self and non-self interactions is conferred by the N-terminal region of both HD1 and HD2 proteins (Yee and Kronstad, 1993; Kües et al., 1994a; Kües et al., 1994b; Yue et al., 1997). Whereas this region is highly variable between different alleles of both *HD1* and *HD2* genes the homeodomain and C-terminal regions usually display a more conserved pattern. Therefore, these genes seem to evolve under a pattern called the "divergence-homogenization duality" (Badrane and May, 1999), which is characterized, on one hand, by divergence in the N-terminal region in order to maintain mating compatibility and, on the other hand, a greater conservation in the C-terminal region to preserve the transcriptional regulatory motifs necessary to induce the homeodomain-regulated pathway (Badrane and May, 1999). Therefore, within a population of a certain species, many *HD1/HD2* alleles may exist and exhibit a high level of sequence variation. Such variation is consistent with models stating that these loci are subject to balancing selection, being therefore maintained in populations over much longer periods than would be expected for regions under neutral evolution (Wright, 1960; May et al., 1999). The current model of interaction between both HD1 and HD2 protein proposes that dimerization is achieved via a limited number of hydrophobic and polar interactions between the variable regions of the proteins (Kämper et al., 1995; Yee and Kronstad, 1998).

Hence, the pheromone/ receptor and homeodomain loci constitute the key components of the bipolar and tetrapolar mating systems of basidiomycetes. Their interaction produces an efficient yet complex mechanism of self/nonself recognition system that mediates different steps of the mating process. The following sections of this chapter are dedicated to the study of the organization of these genomic regions in model basidiomycetes both from evolutionary and phylogenetic perspectives.

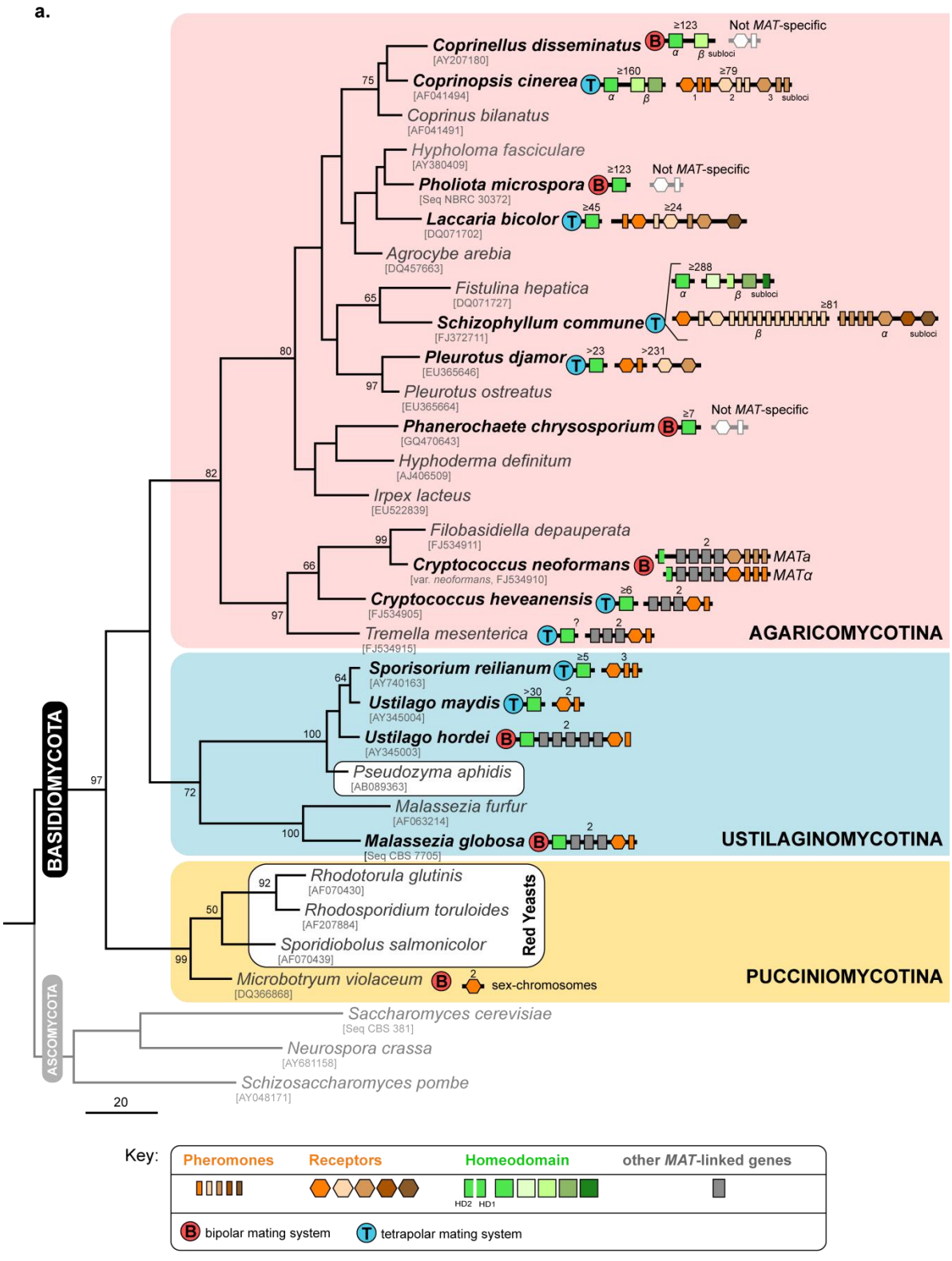

Figure 1.4. Structure and diversity of the *MAT* loci in the different basidiomycete lineages. Description is given in page 20.

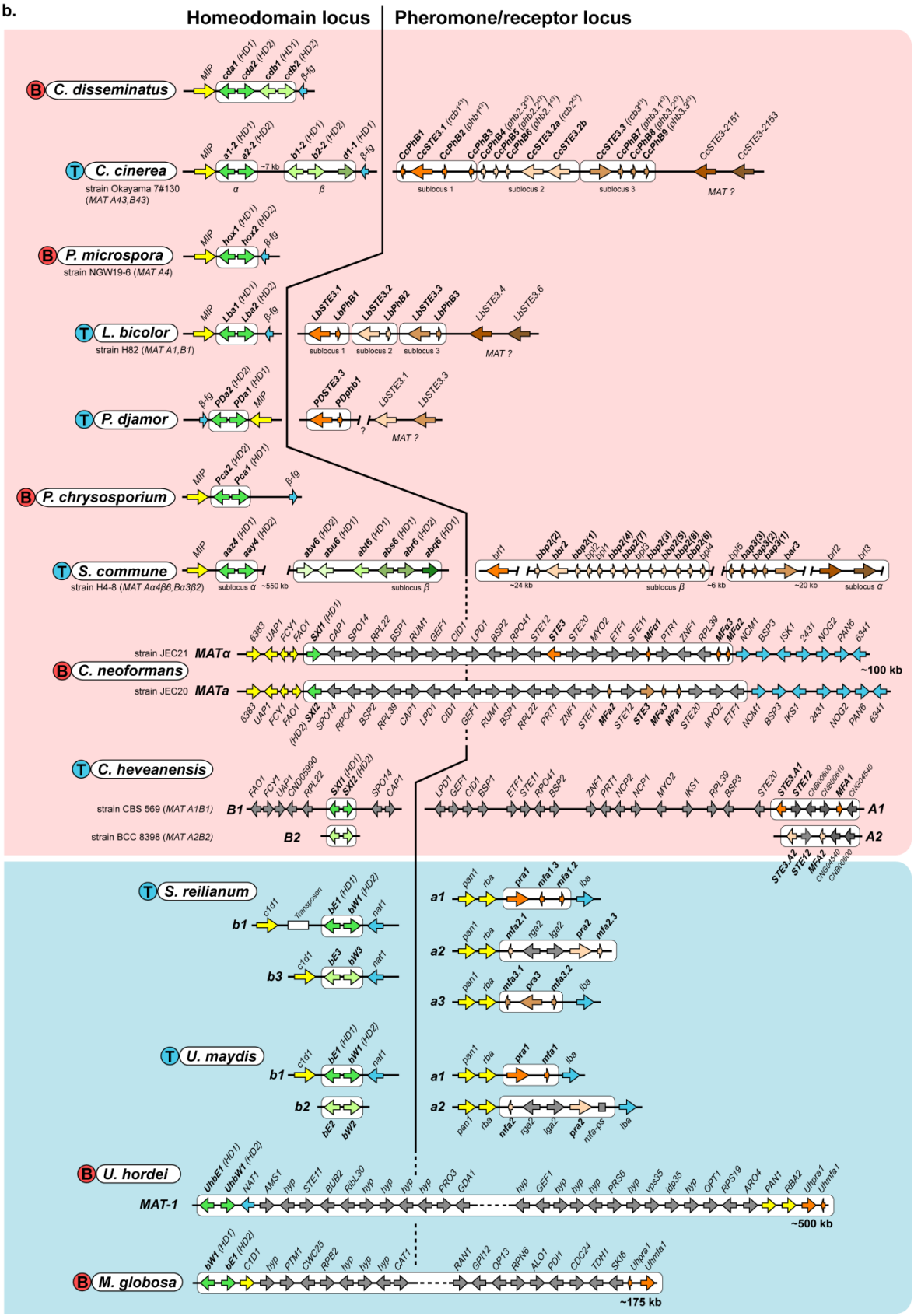

Figure 1.4. Continued. Description is given in page 20.

Figure 1.4. Structure and diversity of the *MAT* locus in the different basidiomycete lineages. (a) Distribution and organization of the different sets of *MAT* genes in representatives of the three main lineages of the phylum Basidiomycota – Agaricomycotina, Ustilaginomycotina and Pucciniomycotina (shadowed in pink, blue and yellow, respectively). The phylogenetic tree was inferred using Maximum Parsimony based on the alignment of the nucleotide sequences of the D1/D2 region of the 26S rDNA. Bootstraps values (> 50%) are shown from 1000 replicates. GenBank accession numbers given below species names (for some species, sequences were retrieved from culture collection databases). The gene content of the homeodomain and pheromone/ receptor *MAT* regions is represented for each species as shown in the key given at the bottom (different colours of each symbol stands for different allelic versions). Genes connected by a black horizontal line are genetically linked. With exception of *C. neoformans*, for tetrapolar and bipolar species, only one *MAT* allele is represented with numbers above indicating the number of *MAT* alleles determined for each species. Species enclosed in white rounded boxes in Ustilaginomycotina and Pucciniomycotina represent members of the two groups of yeasts under study in this thesis. For *M. violaceum*, only the pheromone receptor genes were so far identified. **(b)** Detailed genomic organization of the homeodomain and pheromone/ receptor *MAT* regions of the species highlighted in bold face in the tree. Genes are depicted as arrows indicating the direction of transcription. *MAT*-specific regions are enclosed in white-rounded boxes and *MAT* genes coloured as indicated in the key given in (a). When known, conserved genes flanking the *MAT* loci (coloured in yellow or blue) are shown within each group of species and have the same designation (some of the flanking genes in *U. maydis* and *S. reilianum* are included in the *MAT* locus of *U. hordei* and *M. globosa*). In *C. cinerea*, genes in brackets at the pheromone/ receptor locus refer to the nomenclature used previously for genes as identified by Riquelme et al. (2005). (Based on Bölker et al., 1992; Lengeler et al., 2002; Fraser et al., 2004; Schirawski et al., 2005; Bakkeren et al., 2006; James et al., 2006; James, 2007; Kahmann and Schirawski, 2007; Xu et al., 2007; Bakkeren et al., 2008; Niculita-Hirzel et al., 2008; Morrow and Fraser, 2009; Yi et al., 2009; Metin et al., 2010; Ohm et al., 2010; Stajich et al., 2010; James et al., 2011).

1.2.2.3. Tetrapolar mating systems – the smut fungi

As previously mentioned, for many fungal species, pathogenic development is coupled with sexual reproduction. This situation is found in the emblematic smut fungus *U. maydis*, which belongs to the order Ustilaginales of the subphylum Ustilaginomycotina. In this species, a dimorphic switch from unicellular (and saprobic) yeast-form to filamentous dikaryotic growth is a prerequisite for entry into the parasitic phase and to invade plant tissues. For this reason, it seems logical that the characterization of the mating type loci was one of the first steps carried out towards the understanding of the disease process (Banuett, 2007). In *U. maydis*, the phenotypic and developmental transition is initially mediated by pheromone-receptor interactions defined by the pheromone/ receptor locus (*a* locus) (Bölker et al., 1992). This locus exists in two forms or alleles (*a1* and *a2*) each containing a pheromone precursor gene (*MFA1* or *MFA2*) and a pheromone receptor gene (*PRA1* or *PRA2*) (Bölker et al., 1992) (Figure 1.4b). Due to its simplicity, the organization of the pheromone/ receptor locus as found in *U. maydis* was hypothesized to be the ancestral form of the smut fungi. However, recent genetic and molecular studies in *Sporisorium reilianum*, a related smut species, suggested that the ancestral form of this locus could actually be

more complex. In *S. reilianum*, a detailed analysis of the pheromone/ receptor locus showed that this locus is triallelic and not biallelic as in *U. maydis* (Schirawski et al., 2005). Furthermore, each of the three alleles encodes a single pheromone receptor and two different pheromones, each of which is recognized by one of the receptors encoded by the other alleles (Figure 1.4b) (Schirawski et al., 2005). In *U. maydis*, besides the pheromone and receptor genes, the *a2* allele harbours two additional genes (*lga2* and *rga2*) and a pheromone pseudogene (Urban et al., 1996). The *lga2* and *rga2* genes seem to play important roles in mitochondrial inheritance, sexual reproduction and virulence (Bortfeld et al., 2004; Fedler et al., 2009; Mahlert et al., 2009) and are also present in the *a2* allele of *S. reilianum*. However, the presence of a pheromone pseudogene in the *a2* allele of *U. maydis* in a genomic position where a functional pheromone gene exists in *S. reilianum* supports the hypothesis that the archetype of this locus could in fact be more similar to that of *S. reilianum* with two functional pheromone genes (Urban et al., 1996; Schirawski et al., 2005).

After fusion of compatible partners, heterozygosity at the homeodomain locus (*b* locus) is required for dikaryon maintenance and filamentous growth (Gillissen et al., 1992; Kämper et al., 1995). Each homeodomain locus is composed by two divergently transcribed genes encoding homeodomain proteins known as bE (HD1-type) and bW (HD2-type). Their functional activation follows the rules described in the previous section (1.2.2.2). However, in contrast to the biallelic and triallelic pheromone/ receptor locus of *U. maydis* and *S. reilianum*, the homeodomain locus of both species is multiallelic: in *U. maydis* more than 30 allelic versions were found in natural populations, while in *S. reilianum* five alleles were already identified (Schirawski et al., 2005). Because both pheromone/ receptor and homeodomain loci are genetically unlinked, *U. maydis* and *S. reilianum* have a tetrapolar mating system with multiple mating types.

1.2.2.4. Tetrapolar mating systems – the origin of thousands of mating types

In many members of the Agaricomycotina, the tetrapolar mating system reflects a more complex organization of both *MAT* loci that can ultimately originate up to thousands of mating types. The evolutionary strategy underlying this enormous repertoire of mating types, although very complex, has probably derived from segmental duplications of both archetypal *MAT* loci as found in *S. reilianum* followed by rounds of recombination and diversification between alleles (Riquelme et al., 2005). For example, in the ink cap mushroom *Coprinopsis cinerea* the homeodomain locus is multiallelic and is composed of two subloci (α and β). While the α subloci encodes one HD1/HD2 pair of homeodomain proteins, the β subloci encodes up to three HD1/HD2 pairs and is located ~7 kb downstream of the α subloci (Figure 1.4b). Interestingly, gene deletions have occurred during the evolution of the homeodomain locus giving rise to alleles with different numbers of homeodomain genes. Similarly, the pheromone/ receptor locus is multiallelic and consists of three tandemly arranged independent subloci (sublocus 1 to 3). Each sub-locus contains multiallelic and functionally

redundant similar genes encoding at least one pheromone receptor and one to three pheromone genes (Figure 1.4). Phylogenetic analysis of the *C. cinerea* pheromone receptors divided these proteins in two main clusters supporting an early duplication event of the ancestral gene, followed by subsequent sequence diversification (Riquelme et al., 2005).

Interestingly, even in such a complex mating system, compatibility only requires that mating partners bring together different allelic versions of genes at least in one sublocus of both *MAT* loci (Casselton and Kües, 2007). The organization of the extant *MAT* subloci is maintained because the DNA sequence that comprises each group of genes (both coding and flanking sequences) is sufficiently different to prevent homologous recombination between allelic versions. Thus, in each sublocus genes remain connected as a single unit such that intralocus recombination leading to compatible gene combinations is prevented (Casselton and Kües, 2007). However, during evolution the different subloci have been randomly mixed suggesting that at some stage they have been flanked by homologous sequences that allowed recombination between subloci (May and Matzke, 1995). Noteworthy, the ~7 kb of homologous sequence between the two first subloci in the homeodomain locus may be a relic of the ancestral configuration. A population level analysis of the homeodomain locus in *C. cinerea* identified several versions of each sublocus enough to generate 120 genetically different loci (May and Matzke, 1995). Combined with the several subloci versions also found for the pheromone/ receptor locus, more than 12 000 mating types can be generated in *C. cinerea* (Casselton and Kües, 2007). However, in other fungi (viz. *Schizophyllum commune*), even more mating types can be generated, since both multiallelic *MAT* loci are distributed in distinct subloci that can freely recombine (Brown and Casselton, 2001; Fowler et al., 2004) (Figure 1.4b). Taking advantage of the genome sequencing projects, it was possible to examine and compare the *MAT* loci structure of several other mushroom species. These analyses revealed that synteny (conserved gene order) around the homeodomain locus is maintained across several mushroom species (Figure 1.4b). By contrast, the pheromone/ receptor locus is usually located in a genomic region prone to diversification driven by gene duplications, translocations and transposon insertions (Niculita-Hirzel et al., 2008).

Another particularity of *C. cinerea* and other tetrapolar mushroom fungi (e.g. *Schizophyllum commune*), is that the control of cell-cell attraction and fusion by the pheromone/receptor system has been abandoned (Casselton and Olesnicky, 1998; Kamada, 2002). When any two monokaryotic filaments meet, they undergo cell-cell fusion irrespective of whether or not they are compatible at the pheromone/ receptor locus. However, both the pheromone/ receptor and homeodomain loci are required for the development and maintenance of the dikaryotic state. Interestingly, in both *C. cinerea* and *S. commune*, a difference in only one of the *MAT* loci is sufficient to activate the respective homeodomain- or pheromone/ receptor-regulated developmental pathway. Thus, based on the phenotypes derived from these partial mating interactions, it was possible to determine the developmental pathway controlled by genes at each *MAT* locus. Genes encoded at the homeodomain locus were found to be necessary to repress asexual sporulation (Tymon et al., 1992), to regulate pairing of nuclei within the dikaryotic tip cell and to coordinate nuclear division,

clamp cell formation and septation from the subapical cell (Swiezynski and Day, 1960; Raper, 1966; Casselton and Olesnicky, 1998) (Figure 1.5). On the other hand, the genes encoded by the pheromone/ receptor locus initiate nuclear exchange and migration to the apical cell and regulate fusion of the clamp cell with the subapical cell (Casselton and Olesnicky, 1998; Kües, 2000; Kamada, 2002) (Figure 1.5).

Hence, a tetrapolar multiallelic *MAT* system generates thousands of mating types ensuring that nearly every fusion event produces a compatible interaction. Therefore, this system maintains a high outcrossing frequency while restricting inbreeding. In this context, it is tempting to speculate that in species with a very large number of mating types, the search for suitable mating partners by means of a pheromone-mediated chemoattraction has been abandoned, since any two individuals in a natural population are most likely compatible mates.

1.2.2.5. Evolution of bipolar systems – multiallelic bipolar basidiomycetes

Although tetrapolar systems predominate in the phylum Basidiomycota, bipolar species also exist and, most significantly, do not seem to be uniformly distributed. For example, about 25% of mushroom species are estimated to have a bipolar mating system and these occur interspersed with tetrapolar relatives (Raper, 1966). Hypotheses for the foundations of the bipolar mating systems of basidiomycetes have been firstly proposed by John Raper. One such hypothesis assumes that, mutations occurring in either tetrapolar *MAT* loci leading to self-compatible interactions could turn the tetrapolar system into a bipolar system. Remarkably, mutations leading to self-compatibility were observed in both natural and laboratory strains of the tetrapolar species *C. cinerea* (Kües et al., 1994b; Olesnicky et al., 2000).

In line with this hypothesis, a substantial body of evidence, including recent genomic data, indicate that in the mushroom species *Coprinellus disseminatus*, *Pholiota microspora* (*nameko*) and *Phanerochaete chrysosporium* the transition to bipolarity was due to loss of the association of the pheromone and receptors to the *MAT* locus (Martinez et al., 2004; James et al., 2006; Yi et al., 2009; James et al., 2011) (Figure 1.4). However, it remains to be clarified, whether the pheromones and receptors are self-compatible and/or if the pheromone receptor-regulated pathway is constitutively activated.

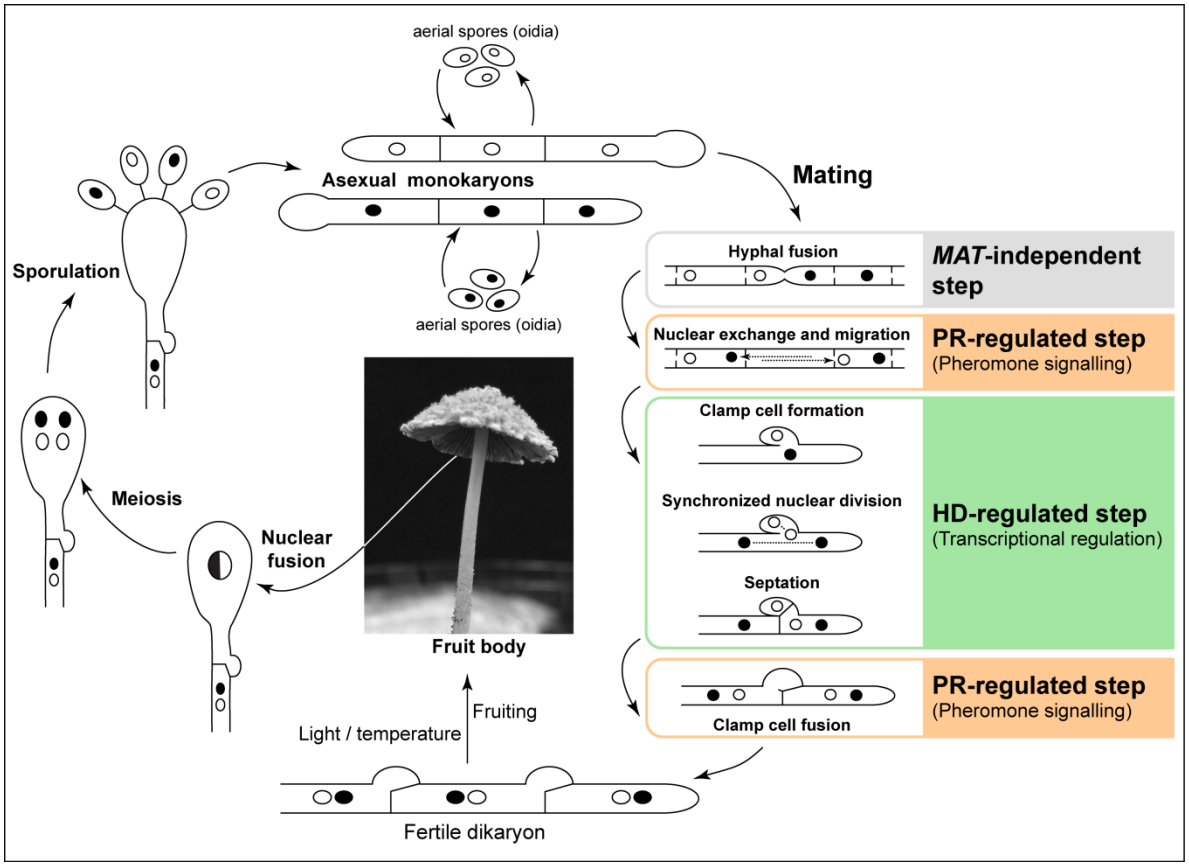

Figure 1.5. Life cycle of *Coprinopsis cinerea*. Roles of the pheromone/ receptor (PR) and homeodomain (HD) *MAT* genes on the developmental processes of *C. cinerea*, particularly during the formation and maintenance of the dikaryon. [(Based on Casselton and Olesnicky, 1998; Kües, 2000); the photograph of *C. cinerea* was obtained from (Stajich et al., 2010)].

Moreover, the observation that the homeodomain locus is the only mating incompatibility factor in these bipolar species raises some questions on the relative role of the homeodomain genes versus the pheromone/ receptor genes in controlling processes usually dependent of the pheromone/ receptor loci, such as nuclear migration. Interestingly, a very recent report in the bipolar mushroom *P. microspora* showed that compatibility at the homeodomain locus is sufficient for both clamp cell formation and fusion (Yi et al., 2010). This suggests that the homeodomain genes may be, at some extent, regulating processes usually dependent of genes at the pheromone/ receptor locus, like clamp cell fusion. This data evokes another of Raper's hypothesis proposing that the function of one of the *MAT* loci could be gradually assumed by the other (Raper, 1966).

Finally, given that the homeodomain locus is multiallelic, this bipolar system maintains a high outcrossing frequency while simultaneously increasing the odds of inbreeding (50%) when comparing with the tetrapolar system (25%).

1.2.2.6. Evolution of bipolar systems – biallelic bipolar basidiomycetes

A final hypothesis proposed by Raper for the origin of the bipolar system, considers that bipolar species could arise from tetrapolar ancestors as a result of coalescence of the two *MAT* loci into a single non-recombining *MAT* region (Raper, 1966). Examples that support this model have indeed been identified in model species from at least two different phylogenetic lineages: (i) in the barley smut fungi *Ustilago hordei* classified in the Ustilaginomycotina and (ii) in the human pathogen *Cryptococcus neoformans* included in the Agaricomycotina.

In *U. hordei*, the role of the sexual cycle in pathogenesis is similar to that of *U. maydis* outlined above (see 1.2.2.3). However, while *U. maydis* has a tetrapolar mating system defined by two unlinked *MAT* loci located in two different chromosomes, *U. hordei* has a bipolar system, despite their close phylogenetic relationship (Begerow et al., 2000). The underlying reason for the bipolar mating behaviour of *U. hordei* was addressed in a landmark study by Bakkeren and Kronstad, where it was found that the two equivalent *MAT* loci were fused into a non-recombining region and segregated together after meiosis (Bakkeren and Kronstad, 1994) (Figure 1.4b). Evidence that recombination is suppressed across the entire *MAT* region came from the analysis of random progeny derived from independent spores, revealing that combinations of the *MAT* loci different from the parental genotypes were not observed in any of the analysed progeny. Furthermore it was found that the *MAT* locus of *U. hordei* is unexpectedly expanded up to ~0.5 Mb, with the pheromone/ receptor locus at one end and the homeodomain locus at the other (Bakkeren and Kronstad, 1994) (Figure 1.4b). While sequencing of pheromone receptor and homeodomain genes from strains isolated worldwide confirmed that only two allelic versions of the *MAT* locus exist in natural populations (Bakkeren and Kronstad, 1994), the annotation of the genes included in the region of suppressed recombination in the *MAT-1* locus of *U. hordei*, revealed a remarkable accumulation of repetitive elements and transposons, in addition to genes not obviously related to mating or pathogenicity (Bakkeren et al., 2006). Finally, comparison between the *MAT-1* region of *U. hordei* with the corresponding genomic regions of *U. maydis* showed substantial rearrangements, including inversions and translocations, which have occurred since the divergence of the two species. However, given that synteny in the vicinity of the *MAT* genes located at both loci was maintained between the two species, the authors suggested that the tetrapolar configuration was most likely present in the common ancestor of these fungi and that a chromosomal translocation fused the two *MAT* regions to establish the bipolar *MAT* locus (Bakkeren et al., 2006). Hence, in *U. hordei* every successful fusion event mediated by compatible genes at the pheromone/ receptor loci ensures a compatible interaction at the homeodomain loci and completion of the life cycle.

Recently, the whole genome sequence of *Malassezia globosa*, the causing agent of dandruff disease, and also belonging to Ustilaginomycotina, provides an independent example in which the tetrapolar *MAT* loci have become linked to constitute a bipolar mating system (Xu et al., 2007; Bakkeren et al., 2008) (Figure 1.4b). Nevertheless, the sexual cycle of this species remains to be discovered.

A transition from tetrapolar to bipolar mating system also occurred in *C. neoformans* (teleomorph: *Filobasidiella neoformans*), a dimorphic human-pathogenic fungus that infects immunocompromised individuals and causes meningoencephalitis (Kwon-Chung, 1976). The bipolar mating system in *C. neoformans* is defined by two mating types (*MATa* and *MAT*α) and mating follows an identical pattern as referred for *U. maydis* and is controlled by the same sets of *MAT* genes, encoding pheromones and receptors and homeodomain transcription factors (Hull et al., 2002; Hull et al., 2005; Stanton et al., 2010). Results from several studies showed that the *MAT*α locus of *C. neoformans* harbours a single homeodomain gene and the pheromone/receptor cluster is linked to this homeodomain gene spanning about 100 to 120 kb (Moore and Edman, 1993; Hull et al., 2002; Lengeler et al., 2002) (Figure 1.4b). Interestingly, and in contrast to other *MAT* loci of basidiomycetes where both homeodomain genes *HD1* and *HD2* are encoded in each allele of the *MAT* locus, in *C. neoformans* only one of these genes is encoded by the biallelic *MAT* locus: the *MAT*α encodes only the transcription factor Sxi1α (HD1 class) and the the *MATa* locus encodes only the Sxi2a (HD2 class) (Hull et al., 2005). Another interesting feature is that the *Cryptococcus MAT* locus also encodes many genes linked to both pheromone/ receptor and homeodomain genes, resembling the *MAT* locus structure of *U. hordei*. However, in *C. neoformans*, many of these additional genes seem to be directly associated with different steps of the sexual cycle, including mating (e.g. genes encoding components of the MAPK pathway), meiosis and sporulation (Lengeler et al., 2002; Fraser et al., 2004; Nichols et al., 2004) (Figure 1.4b).

The current knowledge on the evolution of the *Cryptococcus MAT* locus can be summarised in three major steps: (i) new genes related to mating and meiosis were acquired into an ancestral tetrapolar *MAT* loci encoding the homeodomain transcription factors and the pheromone/receptor components at two unlinked genomic regions; (ii) a chromosomal translocation linked the two expanded *MAT* loci in one of the mating types leading to a non-recombining region and generating an intermediate tripolar system; (iii) recombination between the two alleles finally yielded a bipolar mating system (Fraser et al., 2007). Strong experimental evidence for this model was obtained by different strategies. Firstly, accessing the reproductive patterns of genetically manipulated strains where both *SXI1*α and *SXI2a* genes were deleted from *MAT* and reintroduced to unlinked genome positions (Hsueh et al., 2008) and, secondly, by recent comparative genomics approaches using more divergent *Cryptococcus* species (Findley et al., 2009), like *Cryptococcus heveanensis*. In this species, a tetrapolar mating system was uncovered from the genomic analysis of the *MAT* regions since they were located in two different and unlinked clusters corresponding to the pheromone/ receptor and homeodomain loci (Metin et al., 2010). In addition, homologues of genes that were recruited into the *MAT* locus in *C. neoformans* were also found in both genomic regions flanking the key *MAT* genes in *C. heveanensis*, which further supports that a similar tetrapolar system was the ancestral state of this group (Metin et al., 2010) (Figure 1.4b).

1.2.2.7. Mating systems in Pucciniomycotina – *Microbotryum violaceum*

Until the present work, knowledge regarding *MAT* systems in Pucciniomycotina [the earliest diverged subphylum of Basidiomycota (Stajich et al., 2009)] was limited to the anther-smut fungus *Microbotryum violaceum*. In this species, two mating types exist (*MAT A1* and *MAT A2*) each strictly correlated with the presence of a size-dimorphic sex chromosome ranging from 2.2 to 4.2 Mb, as identified by karyotype analyses (Hood, 2002). The larger chromosome is found in *MAT A2* strains and contains a lower density of functional genes, suggesting that the acquisition of noncoding sequences is most likely the cause for the different sizes (Hood et al., 2004). Furthermore, analysis of randomly isolated DNA fragments from the two sex-chromosomes of *M. violaceum* indicated a higher accumulation of repetitive elements and transposons than in autosomes, which are expected to restrict or even suppress recombination (Hood, 2002). Recently, sequence analysis and genetic mapping of the same set of DNA fragments demonstrated that the actual non-recombining region in this species is ~25% (~1 Mb) of the chromosome length and that the divergence between mating type-linked homologues in this region varies between 0 and 8.6%, thus resembling the evolutionary strata of the sex chromosomes of animal and plants (Votintseva and Filatov, 2009).

Conjugation between cells of opposite mating type is again required prior to infection of the host plant indicating that cell-to-cell signalling should be mediated by pheromone/ receptor system and that sexual development and pathogenicity are also interconnected in *M. violaceum* (Schäfer et al., 2010). The genes encoding pheromone receptors were indeed found in *Microbotryum* and are mating type-specific: the pheromone receptor A1 (Pr-MatA1) was restricted to *MAT A1* strains whereas the Pr-MatA2 was only found in *MAT A2* strains (Yockteng et al., 2007). Moreover, phylogenetic analysis using sequences of the two pheromone receptors from different *Microbotryum* species showed trans-specific polymorphism, i.e., the sequences from the Pr-MatA1 of different *Microbotryum* species were more akin to each other than to Pr-MatA2 sequences of their own species. This suggests that the two pheromone receptors have been maintained for a long period of time by balancing selection since their divergence, which is expected for mating type genes because mating can only occur between individuals carrying different alleles (May et al., 1999; Richman, 2000). Surprisingly, the divergence between both pheromone receptors was estimated to be at least 370 million years old (MYA), which represents the oldest known trans-specific polymorphism identified so far in any group of organisms (Devier et al., 2009). In the context of fungal evolution, this means that the divergence of both pheromone receptors in *Microbotryum* is very ancient and may have occurred soon after the separation of ascomycetes and basidiomycetes (~450 MYA). Yet, both pheromone and homeodomain transcription factors genes remain to be discovered, and their role in determining sexual compatibility in *Microbotryum* is still uncertain.

In conclusion, the characterization of the genomic architecture of the *MAT* gene complexes across different groups of species provides a fine tool to explain the differences between the *MAT* systems and to retrace their evolutionary histories. Within the basidiomycetes, it is not yet clear if the bipolar system is the ancestral, a derived state, or even both. In this context, the characterization of *MAT* systems in early-diverged lineages may provide new and interesting findings.

1.2.3. Homothallism in Fungi

Sexual reproduction in fungi is typically classified into heterothallism and homothallism. While heterothallic (self-sterile) species require fusion of compatible individuals for sexual reproduction to occur, in homothallic (self-fertile) species each individual (single spore or cell) can undergo sex with itself and produce sexual progeny. The fact that homothallism is present in at least three main fungal lineages (Mucoromycotina, Ascomycota and Basidiomycota) implies that this sexual behaviour should have a selective advantage under certain environmental conditions, for instance, when physical barriers prevent strains of different mating types to get together (Lin and Heitman, 2007). Furthermore, many self-fertile species are not hampered from exchanging genetic information with genetically distinct strains. In these cases, a single spore is capable of completing the life cycle, but can also be outcrossed, therefore providing a balance between genomic stability (by inbreeding) and increased diversity (by outcrossing). In addition, homothallic fungi can also benefit from DNA repair mechanisms conferred by recombination in order to maintain their genomic integrity, despite not always generating distinct progeny (Lin and Heitman, 2007).

Transitions between homothallism and heterothallism have frequently occurred during fungal evolution and results from several authors indicate that they have occurred in both directions (Raper, 1966; Yun et al., 1999; Galagan et al., 2005). Homothallic individuals (or single cells) must contain all the necessary genetic information to initiate sexual development and complete the life cycle. However, self-fertility can be attained by distinct molecular mechanisms: (a) the presence of compatible *MAT* genes in one genome; (b) packaging of two compatible nuclei into a single spore – pseudohomothallism; (c) monokaryotic fruiting or same-sex mating; and (d) mating-type switching. In the context of this thesis, only the first three mechanisms will be briefly described in the following sub-sections of this chapter and are summarized in Figure 1.6.

1.2.3.1. Primary homothallism – compatible *MAT* genes in a single genome

In contrast to the heterothallic species in which *MAT* alleles (or idiomorphs) are present in different mating types, many of the homothallic ascomycetous species examined retained both *MAT* genes in a single genome. However, the *MAT* genes can occur in two different configurations: (i) both *MAT*

genes are fused or closely linked or (ii) are unlinked and located in different regions of a single genome (Figure 1.2 and 1.6a). In these cases, it is thought that the interaction between gene products encoded by the two *MAT* idiomorphs in a single genome mirrors the interaction that generally occurs after recognition and cell fusion between compatible individuals in heterothallic fungi. Therefore, genes are maintained in this configuration and ensure efficient post-fusion developmental steps (Lin and Heitman, 2007). Examples of homothallic species having this configuration are given in Figure 1.6a.

1.2.3.2. Secondary homothallism – pseudohomothallism

Another form of homothallism is achieved by packaging nuclei with compatible mating type information into the same spore (Figure 1.6b). This form of homothallism is well described in *N. tetrasperma* where it is also understood that the formation of the self-fertile ascospores derives from programmed meiotic spindle alignment, nuclear movement and ascospore delimitation, in addition to efficient segregation of the *MAT* genes at the first meiotic division (Merino et al., 1996). Upon germination, the resulting ascospores give rise to a self-fertile heterokaryotic mycelium with nuclei of compatible mating types. Thus this form of homothallism is functionally similar to heterothallism (Raju, 1992). Since this process is also found in unrelated fungal species suggests that pseudohomothallism has arisen several times independently (Raju and Perkins, 1994).

1.2.3.3. Secondary homothallism – monokaryotic fruiting or same-sex mating

In contrast to many homothallic fungi that have preserved the ability to outcross, others exist that are heterothallic but are capable to self. The basidiomycete *C. neoformans* is heterothallic and mating relies on the conjugation of two individuals (a and α) carrying different alleles of the biallelic *MAT* locus (section 1.2.2.6) (Figure 1.6c). However, the first indication that other forms of sexual reproduction could exist in this species came from the unexpected frequency of the mating types in world-wide populations, where in about 99% of the natural and clinical isolates the α mating type prevailed (Kwon-Chung and Bennett, 1978). The answer for this biased distribution relies on the fact that α strains are also able to grow filamentously to produce a monokaryotic mycelium and spores (monokaryotic fruiting) (Figure 1.6c). Interestingly, although this form of reproduction was initially considered mitotic and asexual, it was more recently shown that, in fact, is an altered sexual process that involves cells of the same mating type (same-sex mating) and meiosis (Lin et al., 2005). The critical step during this process is the diploidization which can occur either by mating of cells of the same mating type (mainly α cells) or by endoreplication (replication of the DNA without immediate reduction), and can take place at any stage before sporulation. Subsequently, meiosis occurs in the

basidium to produce recombinant progeny (Lin et al., 2005) (Figure 1.6c). Furthermore, definite evidence that same-sex mating occurs in nature was provided after characterization of naturally occurring αADα hybrids derived from fusion between two α strains of different serotypes (A and D) (Lin et al., 2007). In addition, fusion between strains of the same serotype to produce αAAα hybrids was also observed to occur in nature (Lin et al., 2009).

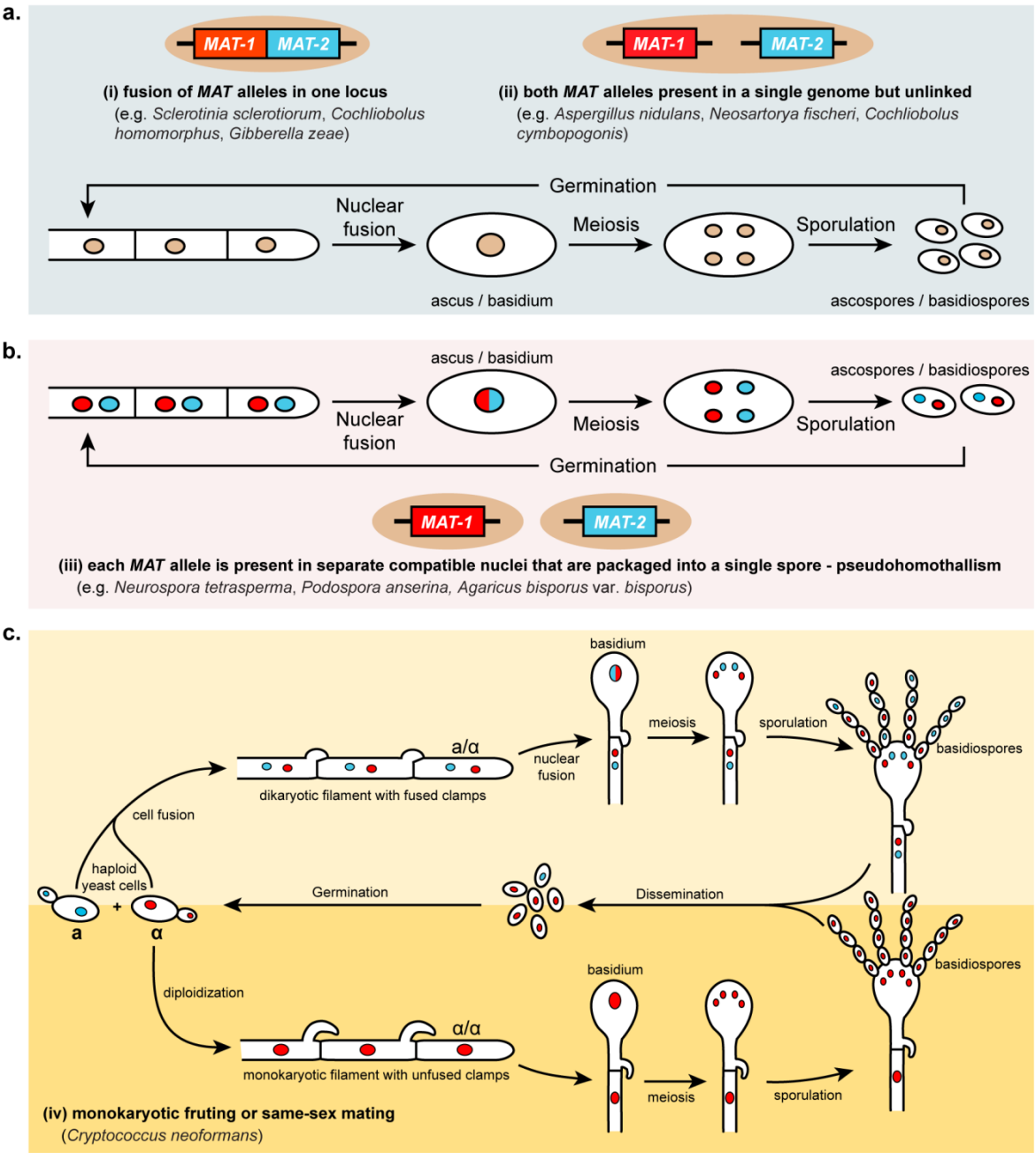

Figure 1.6. Mechanisms of homothallism in fungi. (a) Primary homothallism. **(b)** Secondary homothallism – pseudohomothallism. **(c)** Heterothallic (light yellow panel) and monokaryotic fruiting (dark yellow panel) in *Cryptococcus neoformans*. See text for details.

Several recent results indicate that both heterothallic and homothallic sexual systems in *C. neoformans* have some overlapping regulatory networks, including most of the elements of the pheromone response pathway, but other processes that are specific for same-sex mating also appear to exist (recently reviewed by Wang and Lin, 2011). Although it remains to be fully elucidated how same-sex mating is genetically controlled and what are the environmental signals leading to the diploidization, this mechanism can provide a selective advantage in the absence of a compatible mating partner. This is particularly relevant in *C. neoformans* since the sexual spores (basidiospores) are small enough to reach alveoli (Velagapudi et al., 2009), and may thus function as the infectious propagules.

1.3. Phylogenetic and ecological relationship between parasitic and saprobic basidiomycetous yeasts studied in this thesis

Among the fungi, the basidiomycetes include an incredibly diverse assemblage of species and, as already mentioned, many are plant parasites. For these pathogens, classified in the subphyla Ustilaginomycotina and Pucciniomycotina, pathogenic development is coupled to sexual development. While several of the underlying mechanisms of this relationship have long been determined in the maize smut *Ustilago maydis* (Ustilaginomycotina), considerably less is known for the plant parasites belonging to Pucciniomycotina. Nevertheless, this subphylum also comprises many plants parasites such as the rust fungi (e.g. the wheat rust *Puccinia graminis*) and the anther smuts (e.g. *Microbotryum violaceum*), as well as many saprobic yeast taxa (Hibbett et al., 2007). The latter include the so-called “red yeasts” of the genera *Rhodosporidium* and *Sporidiobolus*, which are classified in the order Sporidiobolales (Sampaio et al., 2003). One of the major limitations of *U. maydis* as a model organism is that, although both maize and anther smuts have very similar lifestyles and depend on a host plant to complete the life cycle, their evolutionary distance prevents direct extrapolation of the vast findings obtained with *U. maydis* to the anther smuts. In contrast, red yeasts are phylogenetic close relatives of the anther smuts and, as experimental model have a major advantage over their strictly parasitic counterparts, because they complete their life cycle on culture media (Sampaio, 2011a; Sampaio, 2011c). Furthermore, they share several characteristics such as a dimorphic life cycle, switching from a haploid yeast phase to a dikaryotic filamentous stage where the sexual structures are formed (teliospores) (Schäfer et al., 2010; Sampaio, 2011a; Sampaio, 2011c). Finally, red yeast species harbour an interesting variety of sexual behaviors, including self-sterility (heterothallism), self-fertility (homothallism) and asexuality (Hamamoto et al., 2011; Sampaio, 2011a; Sampaio, 2011c; Sampaio, 2011b), which further highlights their potential as model organisms to study the evolution of sexual reproduction modes and their biological significance in the Pucciniomycotina.

Another interesting phylogenetic relationship exists between the plant parasites (smuts) of order Ustilaginales (viz. *U. maydis*) and their close relatives included in the genus *Pseudozyma*. *Pseudozyma* species represent the asexual states (anamorphs) of the Ustilaginales and are interspersed with *Ustilago* and other smut fungi species (e.g. *Sporisorium* and *Macalpinomyces*) (Begerow et al., 2000). However, species of *Pseudozyma* are not plant pathogens and seem to behave as saprobes. They can be found in habitats quite distinct from the typical *Ustilago* host plants, like for example, seawater or Antarctic lakes. Therefore, the *Ustilago*/*Pseudozyma* system also offers an interesting platform for studying transitions from sexuality to asexuality and their ecological and evolutionary consequences.

1.4. Objectives and outline

The objective of the work described in this thesis was to bring to light the content and organization of the *MAT* loci in red yeasts and to uncover their sexual reproductive systems, in order to assist the elucidation at the molecular level of important stages of the life cycle in both saprobes and closely related plant parasites. To a lesser extent, this work also aimed to contribute to a better understanding of the so far unclear evolutionary relationship between saprobic (non-pathogenic) species and their parasitic counterparts using *Ustilago*/*Pseudozyma* as the model system. **Chapter 1** consists of a general introduction, which highlights the importance of sexual reproduction in fungal lifestyles and evolution, and revises the molecular mechanisms underlying the different modes of reproduction and sex-determining systems of fungi. **Chapter 2** describes the isolation and characterization of key mating genes in *Pseudozyma* and *Ustilago*-related species using the extensive knowledge available for *Ustilago maydis*. Furthermore, a special emphasis is given to the phylogenetic relationship between *U. maydis* and *P. prolifica*. **Chapter 3** describes the identification and isolation of *MAT* genes in the red yeast species *Rhodosporidium toruloides* using a genome-walking approach, followed by a comparative genomic analysis with two related red yeast species. **Chapter 4** describes the characterization of the putative *MAT* locus of the red yeast *Sporidiobolus salmonicolor*, including the first time isolation in Pucciniomycotina of the genes encoding the homeodomain transcription factors, and the unveiling of a new fungal sex-determining system. In **Chapter 5** the content and organization of *MAT* homologues is explored in an expanded set of red yeast species using degenerate PCR. Connections between the presence of *MAT* genes and the reproductive behaviour are put forward for both sexual and presumably asexual red yeast species. In **Chapter 6**, some concluding remarks of the research described in this thesis are presented together with future perspectives.

2. CHAPTER

Evolutionary relationship between sexual and asexual species of Ustilaginales – The *Ustilago* / *Pseudozyma* system

Part of the work presented in this Chapter was performed at the Max Planck Institute for Terrestrial Microbiology (MPI, Marburg, Germany) under supervision of Prof. Dr. Jan Schirawski.

2.1. Introduction

The smut fungi include about 1450 species (Kirk et al., 2008) some of which are plant pathogens of economical importance. Most of the smut species have a restricted host range and are highly specialized parasites of members of the Poaceae (grasses), including crops such as maize, wheat, barley, sorghum and sugar cane. *Ustilago maydis* belongs to the order Ustilaginales and is the best-studied smut fungi. It is also the most attractive model organism among plant pathogenic basidiomycetes since is one of few that form a true biotrophic interaction that endures during fungal development inside its host, the maize plant (Brefort et al., 2009). Furthermore, *U. maydis* is amenable to efficient reverse genetics and cell biology approaches (Brachmann et al., 2004; Kämper, 2004; Steinberg and Perez-Martin, 2008) and, together with the availability of an annotated genome (Kämper et al., 2006), provides new insights towards the understanding of the disease cycle and of the complexity of the biotrophic interaction (Doehlemann et al., 2008; Brefort et al., 2009).

The most prominent feature of *U. maydis* life cycle is the dimorphic transition occurring upon mating of compatible haploid yeast cells, which allows the fungus to invade its host plant (see 1.2.2.3 and Figure 1.1). Most smut fungi invade plants during the seedling stage causing a systemic infection that leads to a massive fungal proliferation and spore (teliospores) formation in the female and/or male inflorescences of the host (Brefort et al., 2009). However, *U. maydis* may also infect in a later stage of development of the plant, giving rise to tumours in all aerial parts of the host that become filled with wind-borne teliospores. These and other symptoms (e.g. leaf chlorosis) that characterize this smut disease can be scored out to determine the degree of the infection severity.

In *U. maydis* the dimorphic switch is controlled by a tetrapolar mating system defined by a biallelic pheromone/ receptor locus (*a* locus) and a multiallelic homeodomain locus (*b* locus) (Banuett, 2007; Kahmann and Schirawski, 2007) (see 1.2.2.3). Additionally, other processes regulated by components of the *MAT* loci have been analysed in quite some detail in *U. maydis*, the most remarkable being the activation of a signalling pathway upon binding of the mating pheromone to its receptor. Pheromone perception is then transmitted by two interconnected signalling pathways: a cAMP-dependent protein kinase A (PKA) and a mitogen-activated protein kinase (MAPK) (Nadal et al., 2008) (Figure 2.1). The point of convergence of the two pathways is the pheromone response factor (Prf1), which is an HMG-box transcription factor that recognizes and binds to pheromone response elements (PREs) located in the regulatory regions of pheromone-induced genes (Hartmann et al., 1996; Urban et al., 1996). Interestingly, Prf1 is required for basal as well as induced transcription of *MAT* genes and is subjected to autoregulation since two PREs are present in the promoter region of the corresponding gene (Hartmann et al., 1996; Hartmann et al., 1999) (Figure 2.1). It has been shown that full expression of genes at both pheromone/receptor and homeodomain loci is achieved by modulation of the phosphorylation state of the Prf1, which has six MAPK phosphorylation sites and five PKA sites. While posttranslational phosphorylation of the Prf1

by the Adr1 (PKA protein) via cAMP signalling is required for expression of the pheromone/ receptor genes (*a* genes), phosphorylation by both Adr1 and by a MAPK (Kpp2) triggers the expression of the homeodomain transcription factor genes (*b* genes). Thus, Prf1 is able to combine signals from two different pathways and activate distinct transcriptional responses simply by shifting its phosphorylation status (Kaffarnik et al., 2003; Nadal et al., 2008; Zarnack et al., 2008; Brefort et al., 2009) (Figure 2.1). In addition to these complex transcription and posttranscriptional regulators, other signals exist, such as carbon sources that affect *PRF1* gene expression via an upstream regulatory element (UAS) (Hartmann et al., 1999).

A still poorly understood issue is the relationship between the plant pathogenic smuts and their phylogenetic closest relatives of the genus *Pseudozyma* that includes 17 species that are apparently non-pathogenic. Despite being a relative small group, members of *Pseudozyma* are achieving a considerable attention in the scientific literature as recent studies have brought to light exceptional properties of these yeasts. Some examples include (i) potential in biodegradative processes (Middelhoven, 1997); (ii) activity against industrially and medically important fungi (Hajlaoui and Bélanger, 1993; Buzzini and Martini, 2000; Hammami et al., 2011) and (iii) production of mannosylerythritol lipids (MELs) (Arutchelvi et al., 2008). Of these, MELs are the most promising surface active compounds belonging to the glycolipid class of biosurfactants, and have been increasing in demand due to their pharmaceutical applications and biochemical properties, including antitumor and activity against human leukaemia cells (Isoda et al., 1997; Zhao et al., 2001). Notably, it was found that while *Pseudozyma* produce these compounds in relevant quantities, *Ustilago* species produce them only as a minor component (Haskins et al., 1955; Kitamoto et al., 1990).
Some species of *Pseudozyma* are indeed very closely related to *Ustilago* species (Boekhout and Fell, 1998; Begerow et al., 2000). Based on the phylogenetic analysis of six unlinked loci, Munkasci and coworkers suggested that *P. prolifica* (the type species of the genus) should be considered a synonym of *U. maydis* (Munkacsi et al., 2007) if the phylogenetic species concept (sensu Taylor et al., 2000) is employed. However, since no sexual cycle was observed, these authors suggested that *P. prolifica* could be an asexual, saprobic derivative of the sexual *U. maydis*, in line with other examples previously observed for some asexual ascomycetous fungi (Taylor et al., 1999). Furthermore, since all the other species of *Pseudozyma* subsist as saprobes and sexual reproduction has never been observed, a series of experiments presented in this chapter were conducted aiming at shedding some light on the following basic questions: (i) Besides its phylogenetic resemblance with *U. maydis*, is *P. prolifica* a truly mating and pathogenic strain of *U. maydis*? (ii) Are the remaining *Pseudozyma* species cryptic sexual and parasitic smuts or does the genus contain truly asexual species derived from pathogenic and sexual ancestors as a results of loss or change in function of sex-related genes?

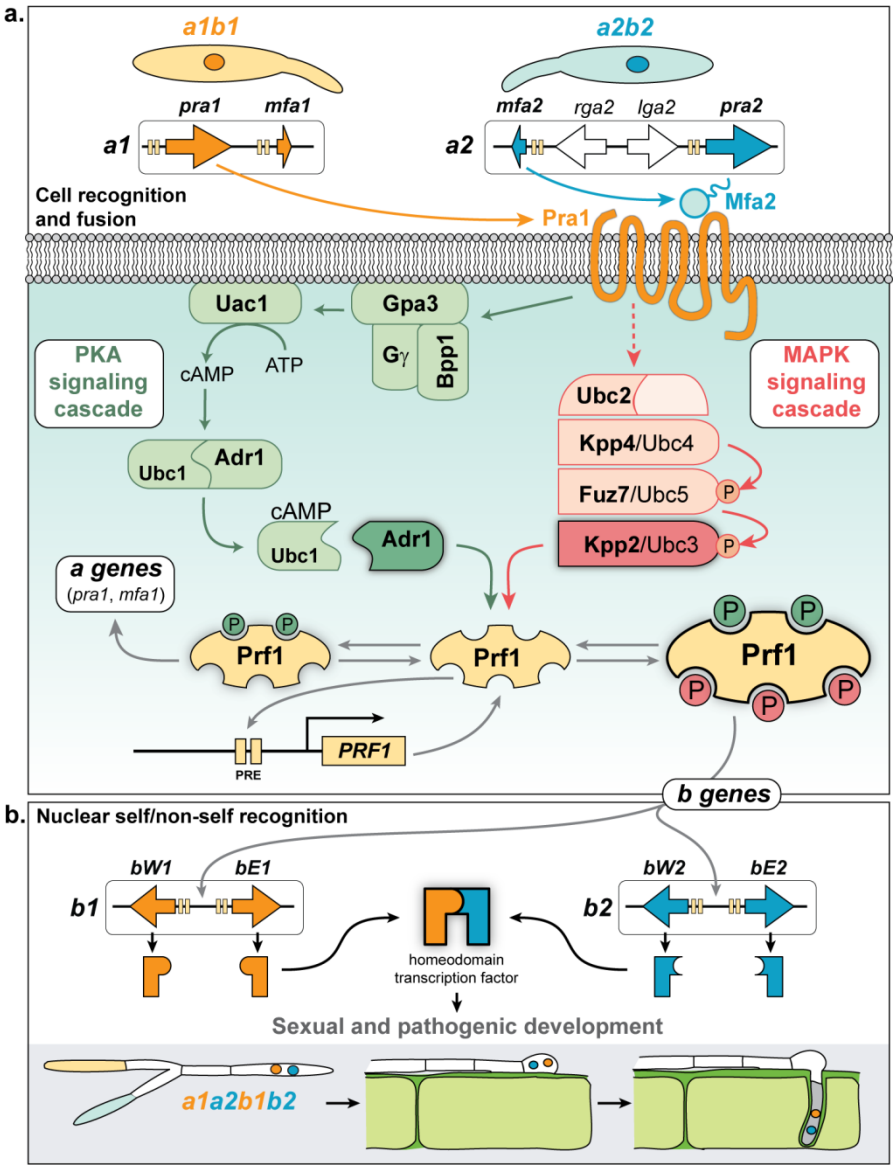

Figure 2.1. PKA and MAPK signalling during mating in *Ustilago maydis*. (a) Genomic regions encoding cell-specific pheromones and receptors are depicted for each mating type. Components required for mating and signalling pathways are represented for a haploid *a1b1* cell. Mfa2 pheromone binds to the transmembrane pheromone receptor Pra1 (orange) and induces both PKA (green) and MAPK (red) signalling pathways, with their interactions indicated by green or red arrows, respectively. Phosphorylation is symbolized by small circles labelled 'P' and coloured according to the respective kinase pathway. Characterized components of the PKA pathway are the α (Gpa3) and β (Bpp1) subunits of a heterotrimeric G-protein; the adenylate cyclase Uac1; and the inhibitory and catalytic subunits of PKA (Ubc1 and Adr1, respectively). Components of the MAPK pathway are the Ste50p-like adaptor protein Ubc2 (the N-terminal part of Ubc2 is essential only for mating, whereas the C-terminal part is required only for pathogenic development); the MAPK module consist of the kinases Kpp4 (Ubc4 or MAPKKK), Fuz7 (Ubc5, MAPKK) and Kpp2 (Ubc3, MAPK). Prf1 (light yellow) binds to PREs (light yellow boxes on promoter regions) in the *PRF1* promoter. Phosphorylated Prf1 induces expression of the *a* (pheromone/ receptor) and *b* (homeodomain) mating type genes. **(b)** After cell fusion an active homeodomain transcription factor is formed (bE/bW heterodimer) that regulates sexual and pathogenic development. (Adapted from Brefort et al., 2009).

2.2. Material and methods

2.2.1. Yeast strains, growth conditions and mating tests

Strains used in this chapter are listed in Table 2.1. *Ustilago maydis* strains were kindly provided by Prof. Dr. Jan Schirawski. Strains were routinely grown in Potato dextrose Agar (PDA) medium (Difco) at 28ºC or, alternatively, in YEPS-light media at 28ºC in a shaking incubator at 150 rpm, unless stated otherwise. YEPS-light medium contains 1% (w/v) yeast extract, 0.4% (w/v) peptone and 0.4% (w/v) sucrose. To confirm sexual compatibility, cells were grown overnight in 5 ml of YEPS-light at 28ºC, centrifuged at 4500 rpm during 4 min and washed in sterile water. The pellet was then resuspended in 50 μl of sterile water. Each cell suspension (10 μl) was co-spotted in charcoal-containing PDA plates that were sealed with Parafilm and incubated at room temperature for 48 hr. Plates were then observed by eye or with the help of a stereomicroscope for the presence white fuzzy colonies indicative of dikaryotic mycelia (fuzz phenotype).

2.2.2. PCR detection of *MAT* genes in *P. prolifica*

Diagnostic PCRs with primers based on nucleotide sequences flanking or within the *MAT* loci of *U. maydis* were carried out to directly identify these regions in *P. prolifica* strain CBS 319.87. PCR products were purified and directly sequenced with the same primers. PCR reactions were performed in a final volume of 50 μl with the following components: 1× Taq Buffer (Fermentas, Canada), 0.20 mM of each of the four deoxynucleoside triphosphates (GE Healthcare), 0.2 μM of each primer, 100 ng of genomic DNA, and 0.5 U Taq DNA polymerase (Fermentas, Canada). Thermal cycling consisted of a 5-minute denaturation step at 95°C, followed by 35 cycles of denaturation at 95°C for 30 s, 30 s at the annealing temperature (variable), and extension at 72°C (variable time). Annealing temperatures, extension times and primer sequences are given in Table 2.2. A final extension of 7 min at 72°C was performed at the end of each reaction.

Table 2.1. Strains used in Chapter 2 and relevant information pertaining to them. Sexual species are indicated by 'S' and those presumably asexual are indicated by 'AS'. For strains of *U. maydis*, the *MAT* genotype is indicated.

Species name	Strain	Sexuality	Isolation source and remarks
Macalpinomyces eriachnes	56574 (M)	S	*Eriachne helmsii*, Australia
Moesziomyces bullatus	Ust. Exs. 833	S	*Paspalum distichum*, India
Pseudozyma antarctica	PYCC 5048	AS	Sediment from lake Vanda, Antarctica
Pseudozyma aphidis	A116	AS	Seawater 20 Km of Faro, Algarve, Portugal
Pseudozyma aphidis	A236	AS	Seawater 20 Km of Faro, Algarve, Portugal
Pseudozyma aphidis	A259	AS	Seawater 20 Km of Faro, Algarve, Portugal
Pseudozyma aphidis	A272	AS	Seawater 20 Km of Faro, Algarve, Portugal
Pseudozyma aphidis	A81	AS	Seawater 20 Km of Faro, Algarve, Portugal
Pseudozyma aphidis	JCM 10318	AS	Secretion of Aphididae, Germany
Pseudozyma flocculosa	JCM 10321	AS	*Trifolium pratense*, Canada
Pseudozyma fusiformata	CBS 423.96	AS	*Brassica oleacea*, UK
Pseudozyma graminicola	CBS 10092	AS	*Phleum pratense*, Russia
Pseudozyma hubeiensis	AS2.2493	AS	Wilting leaf, China, Hubei Province
Pseudozyma parantarctica	JCM 11752	AS	Human blood, Thailand
Pseudozyma prolifica	CBS 319.87	AS	Litter of *Scirpus microcarpus*, Canada
Pseudozyma rugulosa	JCM 10323	AS	Leaf of *Zea mays* infected with moulds
Pseudozyma shanxiensis	AS 2.2523	AS	Leaf of *Quercus mongolica*, China
Pseudozyma thailandica	JCM 11753	AS	Human blood, Thailand
Pseudozyma tsukubaensis	JCM 10324	AS	Flower, Mont Tsukuba, Japan
Ustilago avenae	MP 2043	S	*Avena barbata*, unknown location
Ustilago cynodontis	DSM 4457	S	*Cynodon dactylon*, Japan
Ustilago hordei	Ust. Exs. 784	S	*Hordeum vulgare*, Iran
Ustilago kolleri	F 947/GD 1300	S	*Avena sativa*, Spain
Ustilago maydis	BUB7	S (a1b3)	(Basse et al., 2002)
Ustilago maydis	BUB8	S (a2b4)	(Schulz et al., 1990)
Ustilago maydis	FB1	S (a1b1)	FB1, FB2, FB6a and FB6b are segregants isolated from the same teliospore (Banuett and Herskowitz, 1989)
Ustilago maydis	FB2	S (a2b2)	
Ustilago maydis	FB6a	S (a2b1)	
Ustilago maydis	FB6b	S (a1b2)	
Ustilago spermophora	H.U.V. 13634	S	*Eragrostis ferruginea*, Unknown
Ustilago striiformis	H.U.V. 18286	S	*Alopecurus pratensis*, Germany
Ustilago vetiveriae	H.U.V. 17954	S	*Vetiveria zizanioides*, Unknown
Ustilago xerochloae	Ust. Exs. 1000	S	*Xerochloa imberbis*, Australia

Table 2.2. List of primers and specific PCR conditions used in Chapter 2. Degenerate nucleotides follow the IUPAC nomenclature.

Primer	**Sequence** (5' – 3')	**Annealing/ extension**	**Remarks**
oMW15	AGATTCGAGGCTGGAATTGC	59ºC / 2 min	*a1* locus-specific region
oBH3	TTCACATTTGGACGCATCG		
oJS463	GTCAGGTAGGCGTTACGACTGC	64ºC / 40 s	amplification of *pan1* gene fragment
oMW12	GTGAACCCTGCCCAGTTTGC		
oJS6	CTTCATACAGACGCTCTGAGGTG	63ºC / 70 s	*a2* locus-specific region
oJS7	GATTCGACGAGCACAGTCACAG		
oJS461	CCAAAGCGGATGAAGACG	58ºC / 40 s	amplification of *nat1* gene fragment
oJS462	TCACTCTGCAGGGATTCG		
oBH17	GTTGGGTTGATTCCGTAGG	59ºC / 2 min	amplification of the *bW-bE* region
E-HIIIb	GATGAACCATAGCGTGAGCTGATG		
oJS326	GTCGCGTATGTACTGTTGG	58ºC / 40 s	amplification of *c1d1* gene fragment
oJS327	GGGCTTATCATAGCCTTCG		
MC076	ATCAGCCGCATCATYGCSCACATG	61ºC / 2 min	amplification of *PRF1* gene fragment
MC078	GTCATGGGYGAGAYGAGCATCGA		
MC093	AGCCTTCATCCCACCTACCC	57ºC / 30 s	*PRF1* expression analysis in *U. maydis*
MC094	CTTGCGACCTTTCCACCTTG		
MC107	CCGCTCGTTCCCTTCTATTC	57ºC / 30 s	*PRF1* expression analysis in *P. aphidis*
MC108	GTCCCTGCTTTGGTCGATTTG		
MC105	GCCGATTGTGCCATTCTCAT	57ºC / 30 s	*EF1*α expression analysis (reference)
MC106	TTGGTGGTGTCCATCTTGTTG		

2.2.3. Plant pathogenicity tests

To assay pathogenicity, maize seedlings (*Zea mays*, variety 'Early Gold Bantam') were grown on T-type potting soil (Frühstorfer Erde, Lauterbach/ Fulda, Germany) in a greenhouse under controlled conditions (14 hr light, at 28ºC; 10 hr dark, at 20ºC) until reaching a three-leaf stage (~8 days after sowing). Strains of *U. maydis* and *P. prolifica* were grown overnight at 28ºC in YEPS-light medium, centrifuged for 3 min at 4500 rpm and washed twice in sterile water. Strains to be crossed were mixed in equal amounts in sterile water to obtain a total concentration of about 10^7 cells/ml. Approximately 1.0 – 1.5 ml of the resulting cell suspension was injected into a plant in the opposite side of the cotyledon (about 2.5 cm above soil level) using a sterile hypodermic needle. The plants were not watered for a 24 hr period.

2.2.4. Microscopic observation of fungal growth and proliferation *in planta*

To detect the presence of fungal material on the leaf surface, samples of the third leaf of infected maize plants were obtained by removing the leaves at the stalk. Leaves were rinsed briefly in sterile water and then immersed in calcofluor solution (100 μg/ml) for approximately 1 min. After rinsing briefly in sterile water, the leaf was cut to size, placed on a glass slide covered with a cover slip and observed under microscope using the DAPI filter.

Calcofluor stock solution contained 10 mg/ml of fluorescent brightener 28 (Sigma F3543) in DMSO and the working solution was obtained by diluting the stock 1:100 in 0.2 M Tris-HCl pH 8.0.

For visualization of fungal hyphae proliferation in leaves, an appropriate fraction of the third leaf (up to 3 cm below the injection hole) collected ~48 hr after infection, was soaked in ethanol overnight. The leaf segments were rinsed once in sterile water and a 10% (w/v) KOH solution was added after which it was left in incubation at 90ºC for about 3-4 hr. After carefully discarding the KOH solution and rinsing once with sterile water, the leaves were incubated in Chlorazole Black E (CBE) staining solution overnight at 60ºC. The CBE was discarded and the leaf samples were distained in 50% (v/v) glycerol for at least 24 hr. Samples were mounted on glass slides and visualized by under microscope using the differential interference contrast (DIC) filter. The CBE solution contained 0.03% (w/v) Chlorazole Black E in 1:1:1 solution of water, lactic acid and glycerol.

2.2.5. Spore germination and segregation analysis

Infected maize plants were assessed for their symptoms at 9-days post infection (9 dpi) using a standardized disease index. Tumours excised from the plant leaves were dried at 37ºC for two weeks. The “spores-plant leaves” mixture was mashed and mixed in sterile water and centrifuged at 13,000 rpm for 5 min. The supernatant was discarded together with plant debris and the spores were resuspended in a 1.5% (w/v) $CuSO_4$ solution during ~1 hr to destroy residual vegetative cells. The $CuSO_4$ solution was removed after centrifugation in the same conditions and the spores were washed twice in sterile water. Spore suspensions were serially diluted and plated onto PDA plates containing ampicilin, chloramphenicol, kanamycin and tetracycline to inhibit bacterial contamination. Plates were then incubated at 28ºC for several days to allow for spore germination. Colonies were streaked out twice on new PDA plates to obtain single-celled derived colonies. Mating tests were performed as indicated in section 2.2.1 using *U. maydis* test strains of known mating type.

2.2.6. PCR amplification and sequence of *PRF1* homologues

Degenerate primers MC076 and MC078 were designed based on conserved regions of the Prf1 protein sequence obtained from *U. maydis*, *S. reilianum* and *U. hordei*. PCR reactions were performed as indicated in section 2.2.2. The annealing temperature and extension time were adjusted for this pair of primers as shown in Table 2.2. PCR fragments were subsequently sequenced using the same primers.

2.2.7. Sequence data and phylogenetic analysis

For the species tree (Figure 2.7), nucleotide sequences of the D1/D2 region of the 26S rDNA (LSU) and ITS regions were retrieved from GenBank and from Stoll et al. (2005) (sequence accession numbers are listed in section I.1, Appendix I). Sequences of both regions were concatenated and subsequently aligned using ClustalW 1.4 (Thompson et al., 1994) included in the BioEdit software (Hall, 1999). The Phylogenetic tree based on this dataset was inferred by Neighbour-joining (Saitou and Nei, 1987) with the Kimura-2-parameter (Guindon and Gascuel, 2003). To root the phylogenetic tree, D1/D2 and ITS sequences of *Leucocintractia scleriae* and *Ustanciosporium taubertianum* were used. Amino acid sequences of Prf1 were deduced from the nucleotide sequences. No putative introns were predicted using the AUGUSTUS software (Stanke et al., 2008). Prf1 sequences were aligned and the putative phosphorylation sites were searched by comparison with the sites previously identified in *U. maydis* (Kaffarnik et al., 2003). The phylogenetic tree based on amino acid sequences of Prf1 (Figure 2.8) was constructed using Neighbour-joining and the evolutionary distances were computed using the JTT matrix-based method (Jones et al., 1992). Trees were obtained in MEGA4 software (Tamura et al., 2007).

2.2.8. Quantification of the *PRF1* expression by Real-Time RT-PCR

Cells of *P. aphidis* (strains JCM 10318^T, A81 and A116) and *U. maydis* (strains FB1 and FB2) were initially grown in MYP agar medium [0.7% (w/v) malt extract, 0.05% (w/v) yeast extract, 0.25% (w/v) soytone-peptone, 1.5% (w/v) agar] at 25ºC and 150 rpm for ~24 hr. These cultures were separately inoculated in 50 ml of fresh MYP media for each condition to be tested. The same concentration of cells was used in all experiments (initial OD_{640} ~0.015) but in the dual culture experiments, the same OD_{640} was obtained by mixing equal proportions of a pair of strains. Cultures were harvested at OD_{640} ~0.8 and immediately frozen in liquid nitrogen. Isolation of total RNA and purification steps were performed as described in section I.2 (Appendix I). The relative expression of *PRF1* expression was determined by one-step Real-Time RT-PCR using the MasterAmp™ Real-Time RT-PCR Kit

(EPICENTRE® Biotechnologies, Madison, Wisconsin, USA) according to the manufacturer's instructions. Total RNA samples previously treated with DNase were tested for genomic DNA contamination by conventional PCR. Pure RNA samples were used in Real-Time RT-PCR with primers for *P. aphidis* (MC107 and MC108) and for *U. maydis* strains (MC093 and MC094) (Table 2.2). Each reaction was performed in a total volume of 12.5 μl and contained: 1× Green RT-PCR PreMix, 0.5M $MnSO_4$, 0.2 μM of each primer, 1.25 U RetroAmp RT DNA polymerase and 25 ng/μl of RNA template. Real-Time RT-PCR reactions were performed in a Rotor-Gene™ 3000 device (former Corbett Research Company, Sydney, Australia). The first strand synthesis consisted of a 30-minute step at 60ºC (reverse transcription). Subsequent cycling conditions were: 3 min at 94ºC, followed by 40 cycles of 30 s at 94ºC, 30 s at 57ºC, 30 s at 72ºC. Expression data was normalized against *EF1α* (elongation factor-1 α gene) expression using primers MC105 and MC106. A threshold value was chosen for each gene that allowed clear differentiation of samples and the corresponding cycle threshold (CT) values were determined. The efficiency of each reaction was determined by a calibration curve using dilution series of RNA samples of known concentration. The CT values were read out and plotted versus the logarithm of the sample's concentrations. The efficiency was calculated from the slope by the equation:

$$E = 10^{\left(-\frac{1}{\text{slope}}\right)} - 1$$

where *E* is efficiency. An efficiency estimate of ~ 90% was obtained for all reactions. The relative expression of two genes (X and Y) in a same sample is given by:

$$\frac{N_X}{N_Y} = \frac{(1+E_Y)^{CT_Y-1}}{(1+E_X)^{CT_X-1}}$$

Considering that the *Y* gene as the *EF1α* (the reference gene in this case), and since the efficiency was the same for both genes (*E* = 0.9), the relative initial number of cDNA copies of the *PRF1* gene can be given by the equation:

$$\frac{N_{\text{PRF1}}}{N_{\text{EF1}\alpha}} = \frac{(1+0.9)^{CT_{\text{EF1}\alpha}-1}}{(1+0.9)^{CT_{\text{PRF1}}-1}} = [N_0]^{\text{Rel}}_{PRF1} = (1+E)^{(CT_{\text{EF1}\alpha}-CT_{\text{PRF1}})}$$

2.3. Results and discussion

2.3.1. Identification of *MAT* loci in *P. prolifica*

The first attempt to clarify the sexual status of *P. prolifica* was based on the detection of *MAT* genes since they represent genomic footprints of sex. This strategy was preferred above others since several studies indicated that *P. prolifica* CBS 319.87 (the type strain) was an asexual strain (Boekhout and Fell, 1998; Begerow et al., 2000). Accordingly, several genomic regions located either inside or flanking the *MAT* loci of *U. maydis* were selected. Primers previously designed for these regions based on the genomic information of *U. maydis* were used to search by PCR for the corresponding regions in *P. prolifica* (Figure 2.2a). Without exception, amplicons of genes flanking both *MAT* loci (*pan1*, *nat1* and *c1d1*) were obtained from both *U. maydis* control strains FB1 (*a1b1*) and FB2 (*a2b2*) and from *P. prolifica*. Moreover, a 0.9 kb fragment of the selected *a2*-specific region was obtained from both *P. prolifica* and *U. maydis* FB2 using the primers oJS6 and oJS7 that anchored, respectively, in the *rga2* gene and in the promoter region of the *a2* pheromone gene (*mfa2*) (Figure 2.2b). Conversely, no amplification was obtained for the latter strains when the *a1*-specific region was screened with primers oMW15 and oBH3 that anchored, respectively, in the *a1* receptor gene (*pra1*) and in the *rba1* gene (Figure 2.2b). This indicates that *P. prolifica* has an *a2*-type locus similar to that of *U. maydis* FB2.

The other component of the tetrapolar mating system of *U. maydis* corresponds to the multiallelic homeodomain transcription factor genes (*bW* and *bE*). Amplification and sequencing in *P. prolifica* of the 5' end and common intergenic regions of these genes using primers oBH17 and E-HIIIb (Figure 2.2) revealed that a *b9* allele 100% identical to that of *U. maydis* (GenBank: AJ630065) was present in *P. prolifica*. Together, these results confirm that *P. prolifica* has an *a2b9 MAT* genotype.

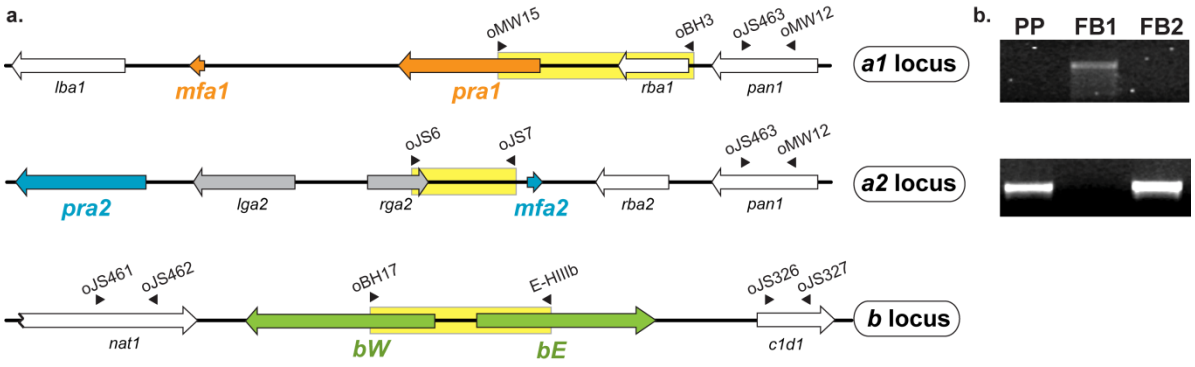

Figure 2.2. Genomic regions encompassing both *MAT* loci of *Ustilago maydis*. (a) In both loci, the selected *MAT*-specific region is highlighted in yellow. *a1*- and *a2*-specific pheromone (*mfa1*/*2*) and receptors (*pra1*/2) genes are shown by orange or blue arrows, respectively, and *b* genes (*bW*/*bE*) are represented by green arrows. Genes flanking the *MAT* loci are shown as empty arrows and primer annealing sites are indicated by arrowheads. **(b)** Amplicons of *a1*- and *a2*-specific regions obtained from *P. prolifica* (PP) and *U. maydis* (FB1 and FB2).

2.3.2. Life cycle and sexual behaviour of *P. prolifica*

Next, since both *MAT* loci were detected in *P. prolifica*, it was essential to determine if this species has retained sexual competence or is a mating-deficient derivative of *U. maydis*. Hence, *P. prolifica* and different strains of *U. maydis* were co-spotted in PDA-charcoal media plates to check for compatible interactions. Mycelia-like colonies (Fuzz phenotype), indicative of positive interactions, were obtained when crossing *P. prolifica* (*a2b9*) with both *U. maydis* FB1 (*a1b1*) and FB6b (*a1b2*) strains, suggesting that *P. prolifica* is mating competent (Figure 2.3).

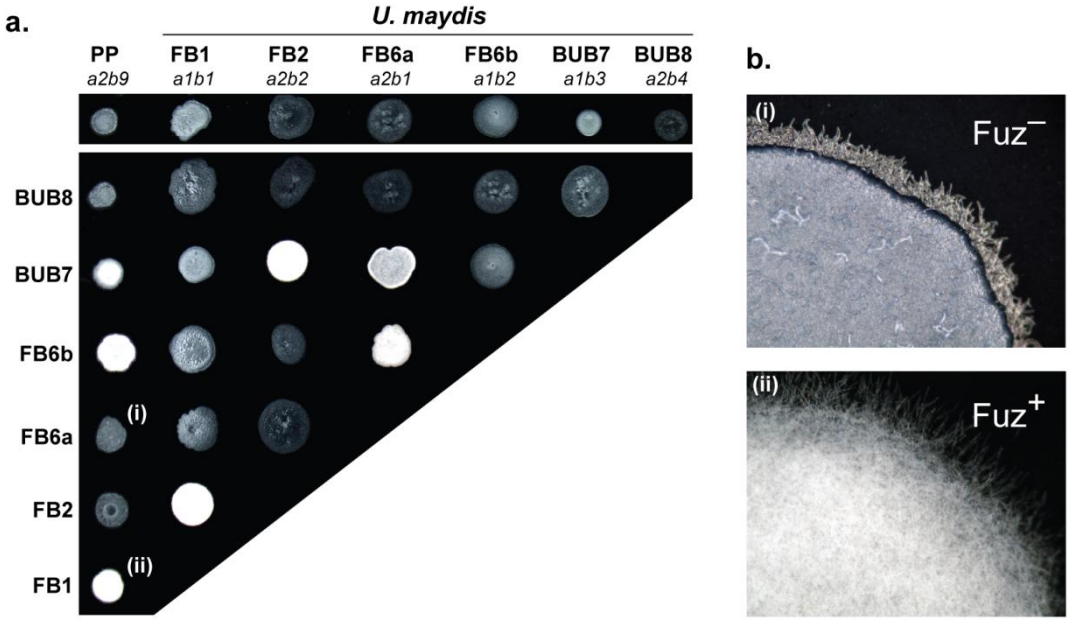

Figure 2.3. Mating tests between *P. prolifica* (PP) and *U. maydis* strains on PDA-charcoal plates. (a) mating reactions between pairs of strains with different *MAT* genotypes: (i) and (ii) represent a negative and a positive mating reaction (white fuzzy colony), respectively. **(b)** magnified views of (i) and (ii).

However, to validate *P. prolifica* as an *Ustilago*-like species, it was important to confirm if sexual development in this species followed the same pattern as that of *U. maydis*. In the latter species, sexual structures (dikaryotic mycelia and teliospores) are formed only after infection of maize plants. Therefore, it was required to verify if the compatible interaction between *P. prolifica* × *U. maydis* FB1 (PP × FB1) strains could also induce disease *in planta*. To examine this, cells were grown separately and then mixed in equal proportions immediately before infection of maize seedling. At 1 day post infection (dpi), filamentous proliferation on leaf surface was microscopically examined by staining fungal hyphae with calcofluor, revealing substantial fungal growth and appressoria formation, indicative of a positive mating interaction (Figures 2.4a – b). Subsequent staining of fungal hyphae with Chlorazol Black E in leaves collected at 2 dpi further revealed a massive invasion of plant cells and proliferation of clamped mycelia (Figures 2.4c – d). At 9 dpi plants were scored for infection

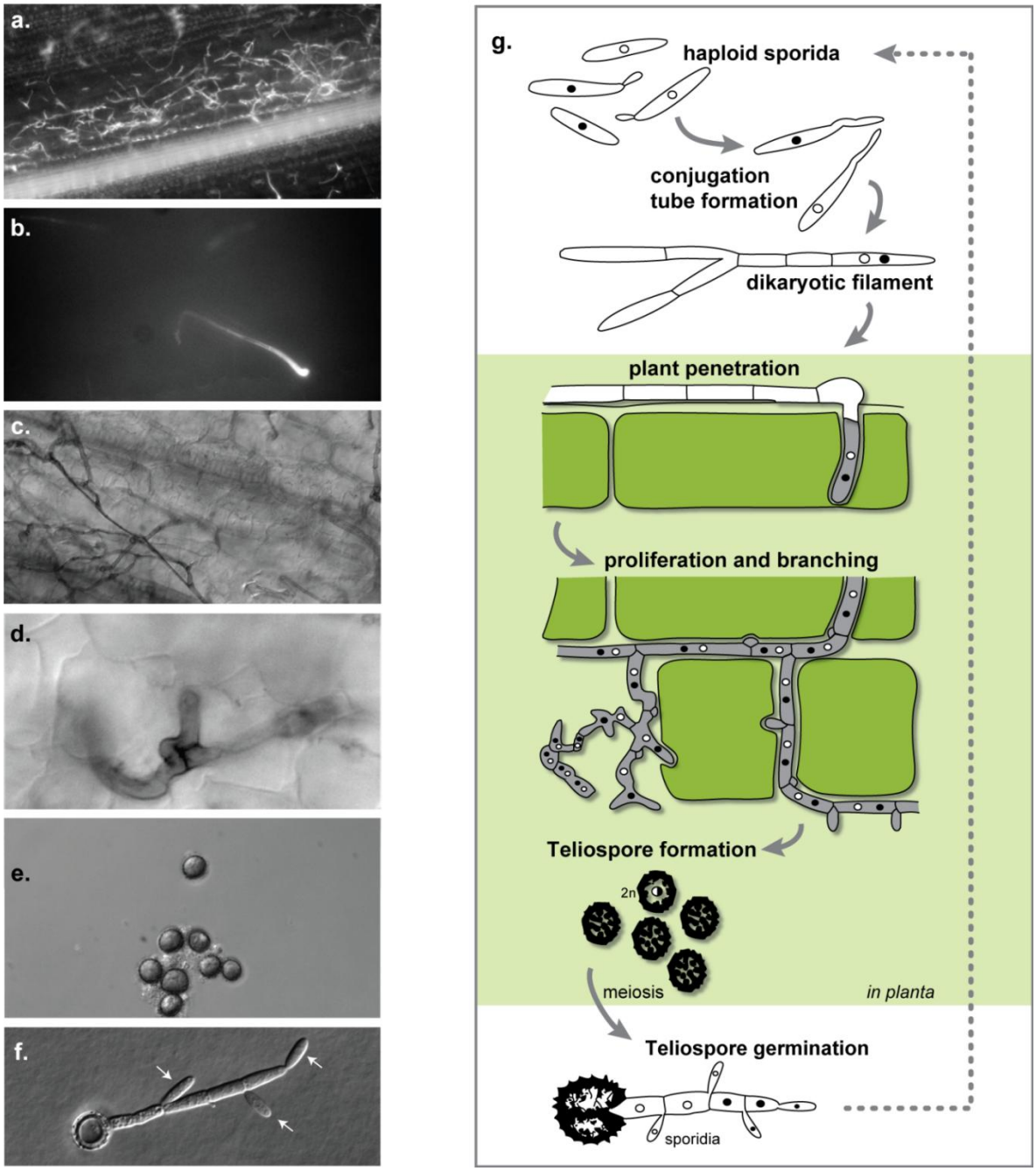

Figure 2.4. Time course microscopic observation of the infection steps of maize seedlings by PP × FB1. **(a)** Filamentous proliferation on leaf surface 1 dpi. **(b)** Appressoria formation. **(c)** Invasion and proliferation in plant cells. **(d)** Formation of dikaryotic and clamped mycelia. **(e)** Mature teliospores. **(f)** Teliospore germination with the formation of a four-celled basidium and basidiospores (arrows). **(g)** Schematic representation of the life cycle of *U. maydis* depicted for comparison purposes. The area shaded in light green indicates the part of the life cycle that takes place inside the plant (life cycle adapted from Kämper et al., 2006).

symptoms and results showed that infection with PP × FB1 induced severe disease symptoms in all aerial parts of the plants (e.g. tumours on leaves and stem) comparable to those of the control infection (FB1 × FB2) (Figure 2.5). On the other hand, no symptoms other than leaf chlorosis were detected in plants infected exclusively with *P. prolifica* (PP) (Figure 2.5). This demonstrates unequivocally that *P. prolifica* can infect and induce disease in maize plants upon mating with *U. maydis* compatible strains.

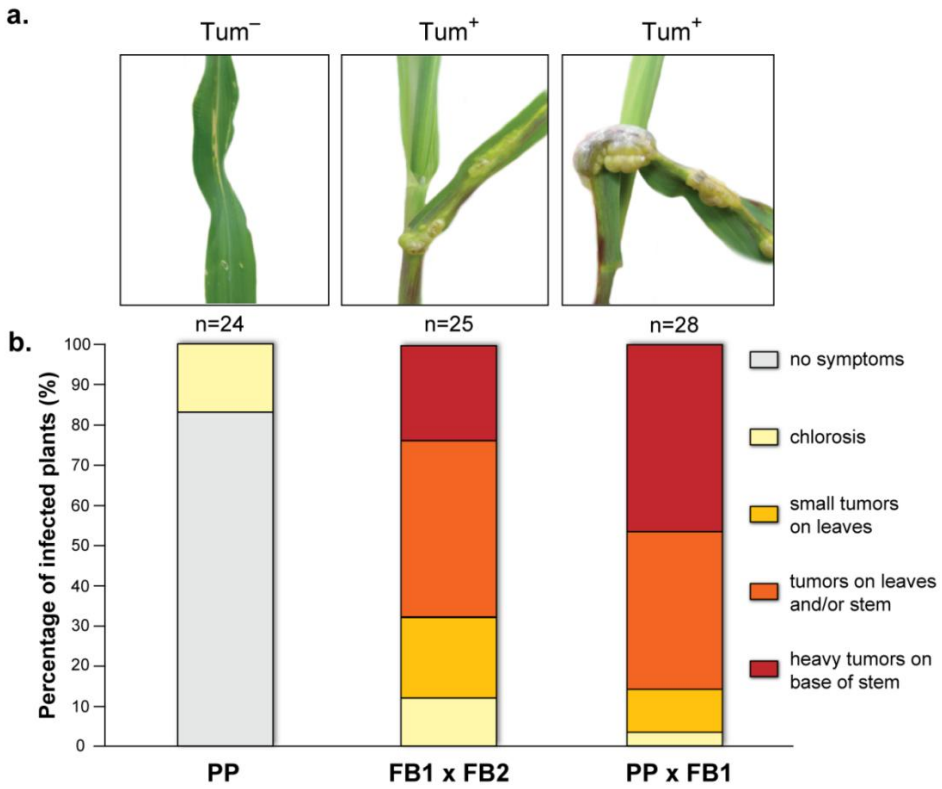

Figure 2.5. Plant pathogenicity resulting from infection with PP × FB1. (a) Tumour morphology (Tum^+) observed at 9 dpi maize plants is depicted for infections with *U. maydis* strains (FB1 × FB2) and with *P. prolifica* × *U. maydis* (PP × FB1). No tumours developed (Tum^-) in plants infected exclusively with *P. prolifica* (PP). **(b)** Disease symptoms and their rating are shown for the three tested infections. The total number of infected plants is indicated above each column and the colour code for disease rating is given on the right.

Finally, teliospores obtained from tumours of plants infected with FB1 x PP germinated in PDA plates and gave rise to basidia and basidiospores, completing the life cycle (Figures 2.4e – g). Strains derived from this cross showed no obvious defects in vegetative growth and a normal meiosis seemed to have occurred since four different mating types were found in the progeny as expected for a tetrapolar mating system (sporidia S15, S20, S25 and S28, Figure 2.6).

Taken together, these results show conclusively that *P. prolifica* is a sexual species and should be regarded as a synonym of *U. maydis*.

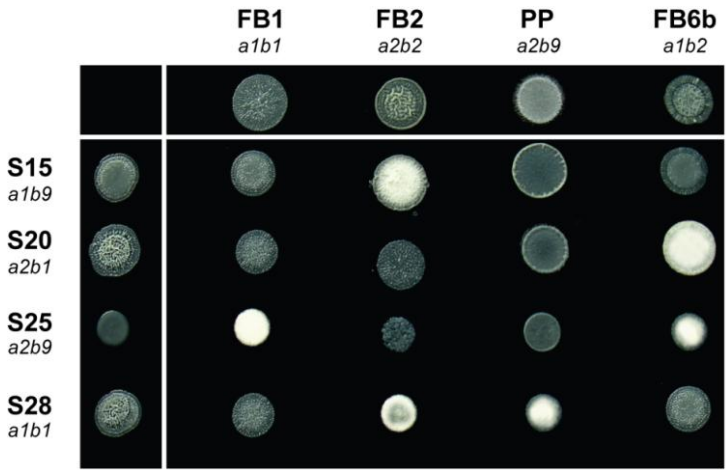

Figure 2.6. Progeny derived from the cross PP × FB1. Four sporidia (S15, S20, S25 and S28) were co-spotted with the indicated test strains of *U. maydis* of known mating type. *MAT* genotypes are shown above the strain names. White fuzzy colonies depict positive mating reactions.

2.3.3. Evolution of the pheromone response factor (*PRF1*) in *Pseudozyma*

In light of the previous results, it is conceivable that other species of *Pseudozyma* may have retained the ability to undergo sexual reproduction, even if reproduction is primarily clonal. On the other hand, a transition to strict asexuality could have occurred as a consequence of a change in the lifestyle (e.g. a shift from pathogen to saprobe). Interestingly, phylogenetic analysis revealed that with the exception of *U. spermophora* and *P. tsukubaensis*, no relationships analogous to that of *U. maydis* and *P. prolifica* were so far observed in this group (Figure 2.7). Instead, the remaining *Pseudozyma* species are interspersed with their sexual relatives but represent distinct phylogenetic entities (Begerow et al., 2000) (Figure 2.7).

To get some insight in sexual proficiency of *Pseudozyma* species, a PCR-based approach was used to amplify in *Pseudozyma* partial sequences of genes that are known to have a central role in sexual reproduction in *U. maydis*. Different sets of degenerate primers were firstly designed based on the conserved regions of the mating type-specific pheromone receptor genes. However, due to the limited genomic information available, primers were only based on sequences retrieved from three species: the receptor genes *PRA1* and *PRA2* from *U. maydis*, *U. hordei* and *S. reilianum* and the additional *PRA3* receptor gene present exclusively in *S. reilianum*. Consequently, although several attempts were made using the different sets of primers, none of them allowed amplification of the receptor genes in any of the species. Therefore, an alternative approach was devised. This consisted in the characterization in *Pseudozyma* species of a well-conserved gene, *PRF1*, which encodes a master regulator of sexual reproduction in *U. maydis*. The presence of a functional copy of *PRF1* in *Pseudozyma* can be an indication of sexual competence.

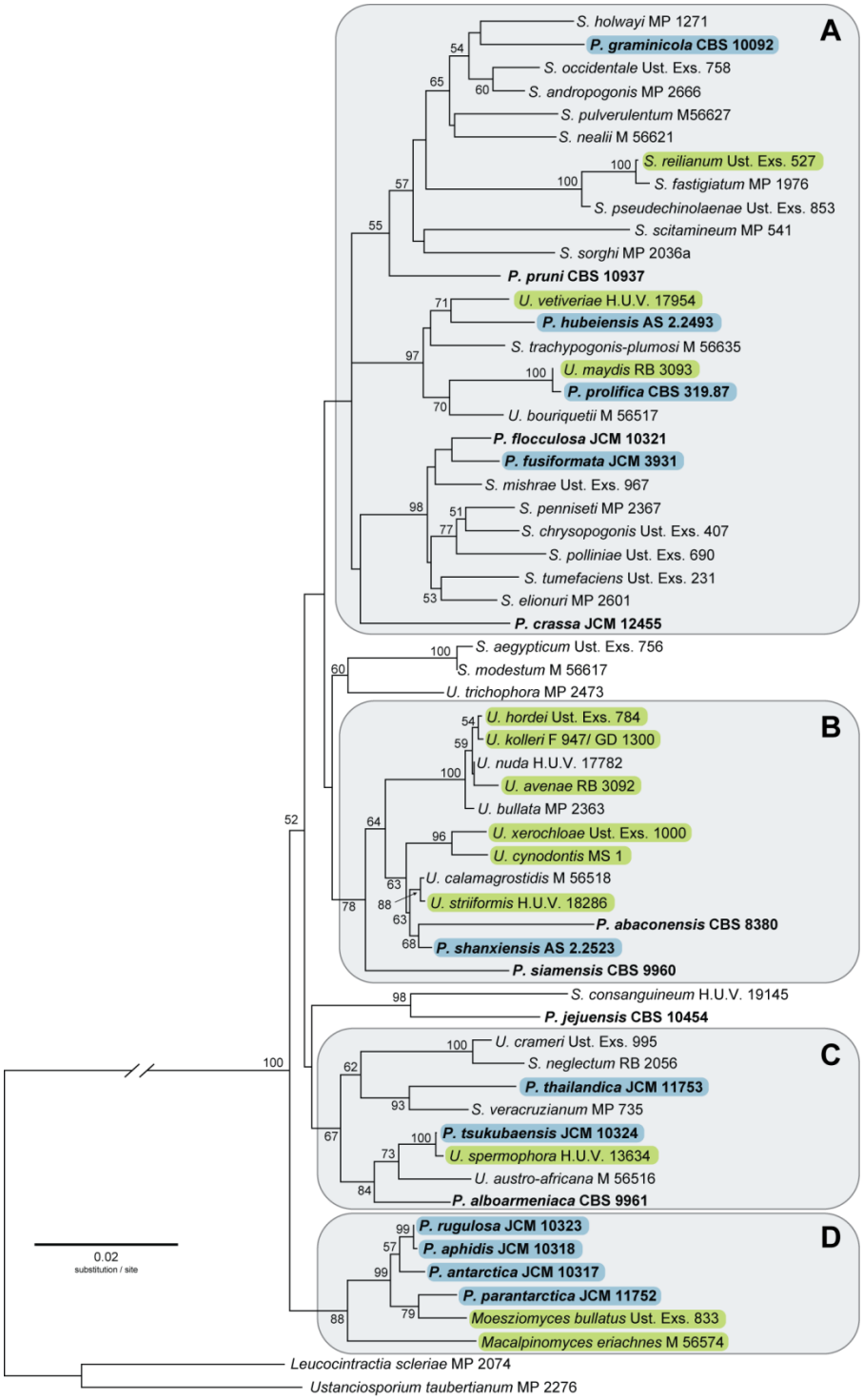

Figure 2.7. Phylogenetic relationships between sexual and asexual species of Ustilaginales. Tree based on the concatenated dataset of the ITS and LSU rDNA sequences of 60 species, inferred by Neighbour-joining and using the Kimura-2-parameter model. Species of *Pseudozyma* are indicated in boldface and in those highlighted in blue (asexual) or green (sexual) boxes *PRF1* homologues were characterized. Strain designations are indicated after the species name. Groups A to D are the same as in Figure 2.8 and 2.9. Abbreviations: *P.*, *Pseudozyma*; *U.*, *Ustilago*; *S.*, *Sporisorium*. Numbers on branches are bootstrap values from 1000 replicates (values less than 50% are not shown).

A comparison of the *PRF1* gene sequences from three sexual species (*U. maydis*, *U. hordei* and *S. reilianum*) revealed a high level of sequence identity (> 70%). Degenerate primers MC076 and MC078 (Table 2.2) were designed based on the conserved regions of the Prf1 protein and used to search for *PRF1* homologues in selected sexual and asexual species of this group (Figure 2.7 and Table 2.1). Fragments of about 1 kb were amplified from 11 (of the 12) species of *Pseudozyma* analysed and from closely related sexual species (Figure 2.7), and subsequently sequenced with the same primers. In all cases, no putative introns or termination codons were predicted in the sequenced region by in silico analyses. The deduced Prf1 amino acid sequences were aligned and used to construct a phylogenetic tree that was found to reproduce the phylogenetic relationships previously established for these species (Figure 2.7 and 2.8). A unique insertion of about 20 amino acids was additionally found in species of the *P. aphidis* clade (Figure 2.8).

In *U. maydis,* Prf1 is posttranslationally regulated by phosphorylation mediated either by PKA or by MAPK signalling pathways, to induce pheromone/ receptor as well as the homeodomain transcription factor genes. Therefore, the newly obtained Prf1 sequences were scrutinized for the presence of conserved PKA and MAPK phosphorylation sites important for Prf1 function (Müller et al., 1999). The sequenced region encompassed the first two PKA phosphorylation sites (P1 and P2) in addition to the first four MAPK sites (M1-M4) identified in *U. maydis* Prf1 (Figure 2.9a). One putative phosphorylation site fitting the PKA consensus sequence R/K R/K X S/T (Taylor et al., 1990) was found across all species (P2, Figure 2.9a). On the other hand, although the first two putative MAPK phosphorylation sites were absent in the majority of the sequences (M1 and M2), the two central MAPK sites were conserved (M3 and M4) (Figure 2.9b). Remarkably, only the central phosphorylation sites were found to be essential during mating in *U. maydis* as previously demonstrated by replacing the wild-type *PRF1* allele with alleles carrying different combinations of mutations that blocked protein phosphorylation (Kaffarnik et al., 2003). Hence, these results suggest that *PRF1* is present and apparently functional in *Pseudozyma* and that a similar posttranslational regulation mechanism may be operating in the analysed species.

2.3.4. Expression analysis of the *PRF1* homologue of *P. aphidis*

While most of the *Pseudozyma* species are found intermingled with many *Ustilago* or *Sporisorium* species, which are sexual, the clade that includes *P. aphidis*, *P. rugulosa* and *P. antarctica* is placed in a basal position of the species tree where relatively less sexual related species were found so far (Figure 2.7 and 2.8). Moreover, most *Pseudozyma* strains of this clade have been isolated from sources that are different from those of their plant pathogens relatives. For example, *P. antarctica* was originally isolated from lake sediments in Antarctica and other strains of this species, as well as the type strain of *P. parantarctica* were isolated from blood samples of patients in Thailand (Sugita et al., 2003). *P. aphidis*, another species included in the same clade, was also reported in a case of

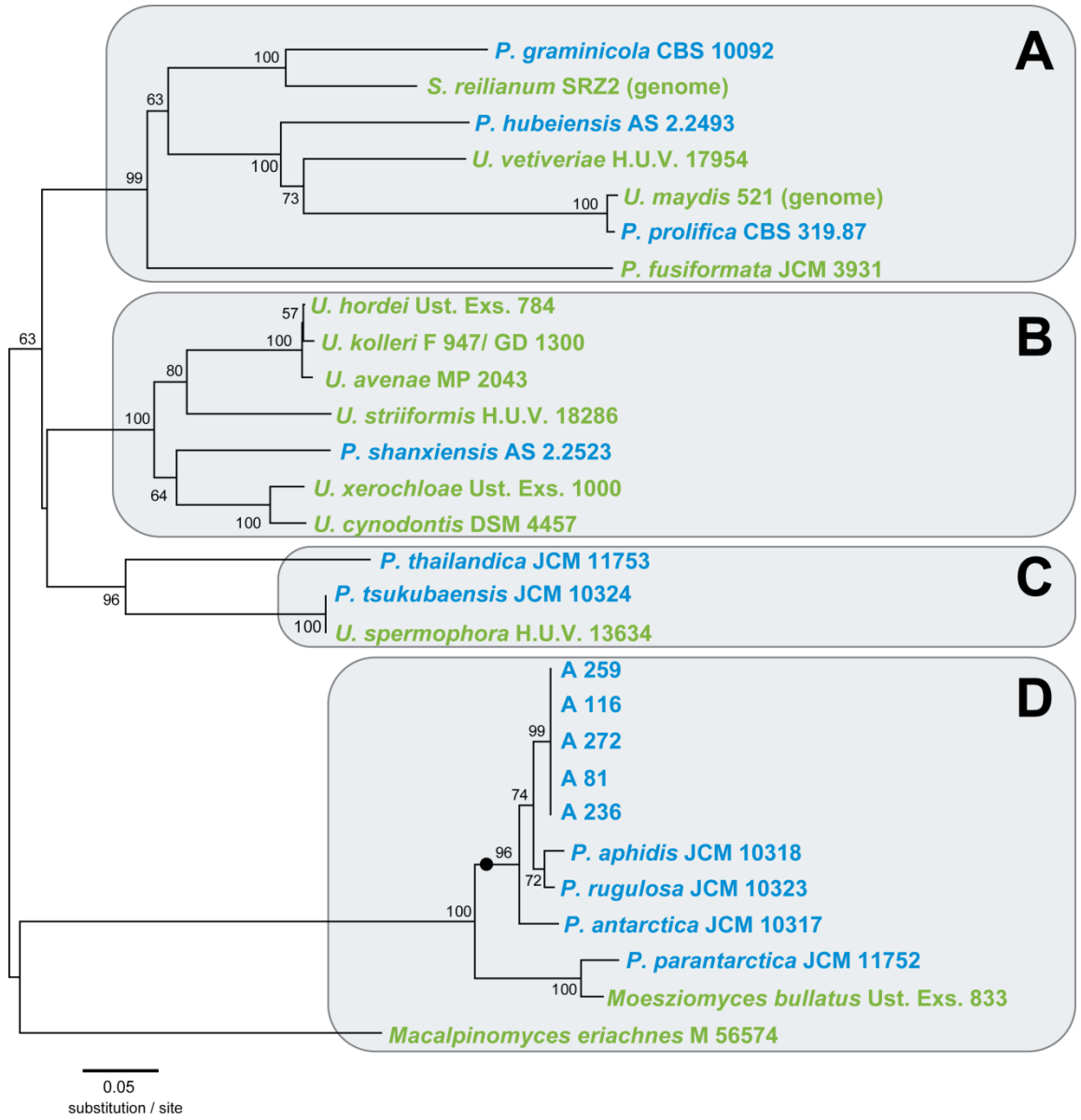

Figure 2.8. Phylogenetic tree based on the amino acid sequences of Prf1 obtained from the several Ustilaginales species. The tree was inferred by Neighbour-joining and the evolutionary distances were computed using the JTT matrix-based method. Sexual and putatively asexual species are depicted in green and blue, respectively. Groups A to D and abbreviations are the same as in Figure 2.7. A 20 amino acid insertion gain was found in the clade indicated by a black circle. The tree is unrooted and the numbers on branches are bootstrap values from 1000 replicates (values less than 50% are not shown).

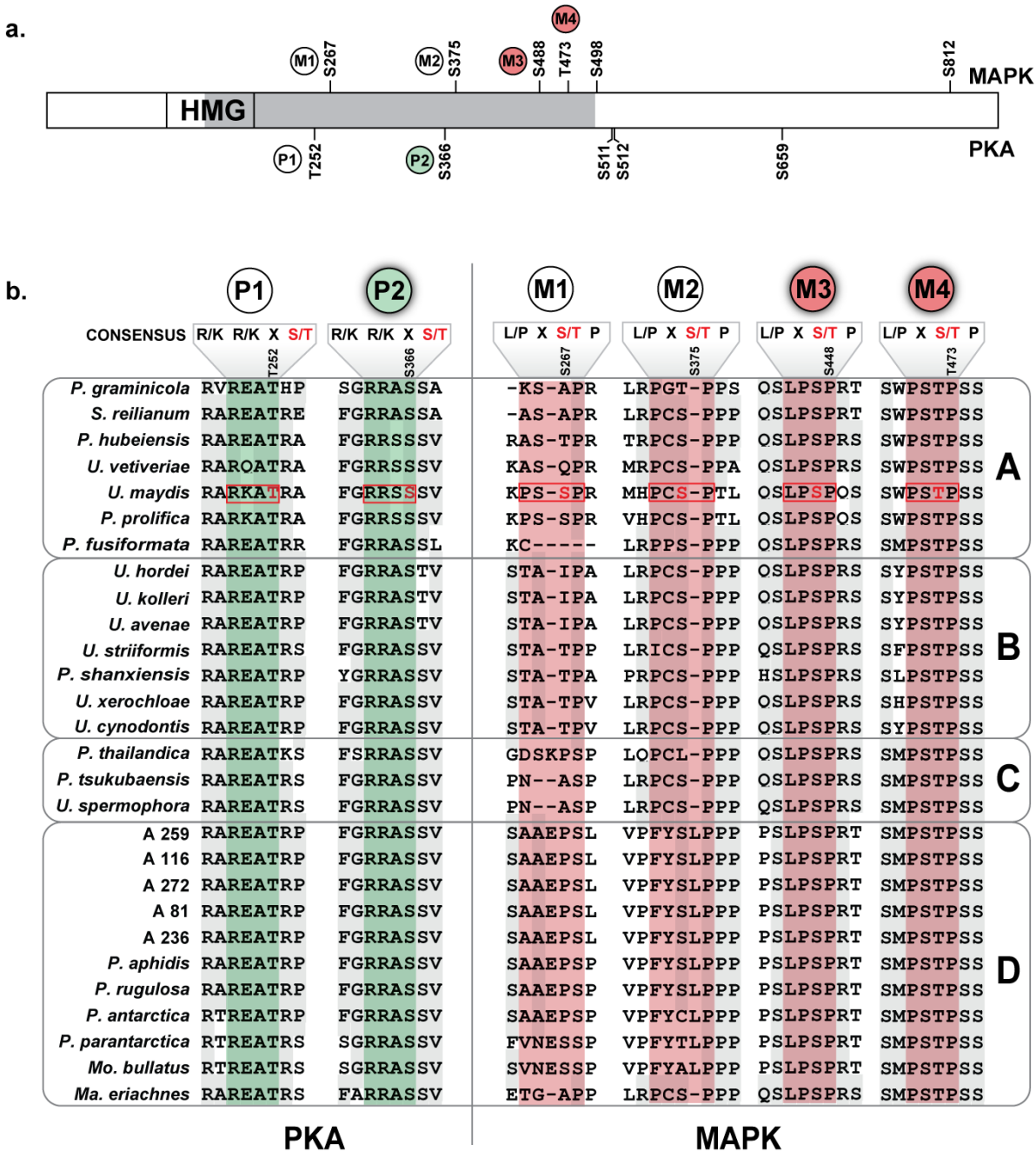

Figure 2.9. Conservation of PKA and MAPK phosphorylation sites in Prf1. (a) Diagram representing the Prf1 protein of *U. maydis*. The region analysed in all the selected species is shadowed in gray. Amino acid positions phosphorylated by PKA and MAPK are indicated, and those depicted in more detail in the bottom alignment are symbolized by labelled circles ("M" and "P" for MAPK and PKA sites, respectively). **(b)** Alignment of the region surrounding the phosphorylation sites. Green and pink shadowed regions indicate the phosphorylation recognition sequence for PKA and MAPK, respectively. For *U. maydis*, the residues phosphorylated are shown in red. Only the phosphorylation sites indicated by coloured circles are conserved across species. Groups A to D are the same as in Figures 2.7 and 2.8.

fungaemia in the USA (Lin et al., 2008) but this species was originally isolated from secretions of aphids on leaves of *Solanum pseudocapsicum* (Henninger and Windisch, 1975). Additionally, *P. aphidis* was consistently isolated from seawater collected at depths of 700 m, 20 Km off the coast of Faro (Portugal), which suggests that this species may have a primary marine origin (Gadanho et al., 2003). Although the abovementioned strains were collected from different sources, in none of them the sexual state was observed. Therefore, it was interesting to explore in more detail the mode of reproduction of this group, particularly for *P. aphids* since several isolates were available.

In order to investigate if isolates of *P. aphidis* were mating competent, they were crossed in different combinations in PDA-charcoal and Water-agar plates. No compatible mating reactions were observed but this could also be a consequence of mating type incompatibilities between isolates. Alternatively, it would indicate that other factors or culture conditions are required to induce sexual reproduction in this species. Prf1 has two major functions during the development in *U. maydis*: one is to stimulate pheromone and receptor gene transcription during mating to promote cell fusion, and the other, is to maintain expression of homeodomain transcription factor genes during pathogenic development.

Based on this knowledge and since *PRF1* is upregulated in *U. maydis* upon pheromone stimulation, an alternative approach was devised to detect a positive mating interaction between strains of *P. aphidis*. The rationale for this experiment was the following: if compatible mating strains of *P. aphidis* are co-cultivated, the expression level of *PRF1* should increase as a result of pheromone perception; on the other hand, strains cultivated separately would only express *PRF1* at basal levels. To test this, three *P. aphidis* strains were used: the type strain (JCM 10318^T) and two strains obtained from the marine environment (A81 and A116). Total RNA was extracted from strains cultivated either separately or mixed in pairs in equal proportions and the relative expression was subsequently determined by Real-Time RT-PCR. Expression of *PRF1* in both single and dual cultures of *U. maydis* was also measured and served as a control (see methods in section 2.2.8).

When A81 and A116 cells were grown separately, *PRF1* expression was low but comparable to the basal expression level detected in cells of *U. maydis* FB1 (Figure 2.10) (Hartmann et al., 1999). On the other hand, while the expression level increased ~5-fold in co-cultured cells of *U. maydis* (FB1 × FB2), no significant induction was observed in dual cultures of *P. aphidis* A81 × A116 (Figure 2.10). These results suggest that pheromone-mediated interactions occurred between strains of *U. maydis* but not between strains of *P. aphidis*, indicating that the latter strains are sexually incompatible or sexually incompetent. However, an interesting result was obtained for the type strain of *P. aphidis* (Paph^T). In this case, single-cultured cells exhibited ~5- to 6-fold higher expression of *PRF1* when comparing to the expression level in cells of other *P. aphidis* strains grown in the same conditions. This higher level of *PRF1* expression in cells of Paph^T was similar to the expression level obtained in the *U. maydis* FB1 × FB2 dual cultures. Nevertheless, when co-culturing Paph^T with another *P. aphidis* strain (A81 or A116) no further induction was observed (Figure 2.10). In the dual culture experiments, one has to take into account that results should reflect the expression of *PRF1* in about

50% of cells of each strain. However, in contrast to the expect expression level of *PRF1* obtained in the A116 × Paph^T experiment, indicating that a similar number of cells was present in the dual culture, the A81 × Paph^T experiment exhibited lower expression. The most likely explanation for this is that A81 × Paph^T cultures had a higher number of cells of strain A81 at the end of growth.

The most prominent conclusion of these experiments was that the expression of *PRF1* gene in the type strain of *P. aphidis* is higher than expected. While this may indicate that Prf1 has acquired a different functional role, further experiments are still required to confirm this hypothesis. Given the absence of information concerning the molecular mating type of all the species involved and whether they possess functional *MAT* genes, it was not possible to derive conclusions concerning sexual capabilities of these strains.

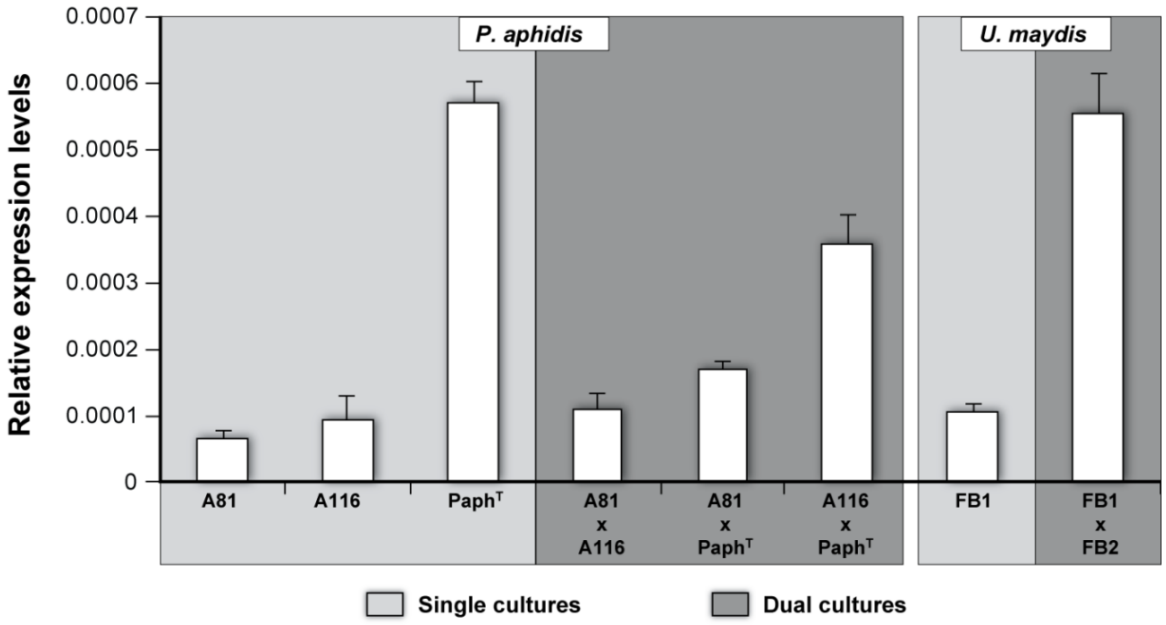

Figure 2.10. Relative expression of the *PRF1* gene as determined by Real-Time RT-PCR analysis. Expression data obtained from cultures grown either alone (single cultures) or mixed in pairs (dual cultures) is depicted from two independent experiments (standard deviation is shown on columns) for both *P. aphidis* and *U. maydis* (control). The expression data was normalized against the *EF1*α gene.

2.4. Final remarks

In summary, molecular, phylogenetic and biological data showed that strain CBS 319.87 (*P. prolifica*) is a mating competent strain of *U. maydis* that causes disease in maize seedlings and is able to complete the life cycle. Moreover, most of the remaining species of *Pseudozyma* showed no substantial evidence of loss of sexual reproduction: (i) apparently functional homologues of *PRF1* gene of *U. maydis* were found in all the analysed *Pseudozyma* species and (ii) Prf1 seemed to be regulated at posttranscriptional level by a similar mechanism by maintaining the functionally relevant phosphorylation sites. Nevertheless, two main observations suggest that a different situation may be evolving in *P. aphidis* (and possibly in other members of the same clade). First, an insertion of about 20 amino acids was found in Prf1 sequences of *P. aphidis*, *P. antarctica* and *P. rugulosa* but not in any other species. Secondly, abnormal high levels of *PRF1* expression were found for the type strain of *P. aphidis*. Although, further studies are still needed, it is tempting to consider that an adaptation to a lifestyle different from that of their closely related plant pathogens may be associated with a change of *PRF1* function in these species.

3. CHAPTER

Identification of mating type (*MAT*) genes in the red yeast *Rhodosporidium toruloides*

The work presented in this chapter was published in:

Coelho M. A., Rosa, A., Rodrigues, N., Fonseca, Á. and Gonçalves, P. 2008. Identification of mating type genes in the bipolar basidiomycetous yeast *Rhodosporidium toruloides*: first insight into the *MAT* locus structure of the Sporidiobolales. Eukaryot Cell 7:1053-1061.

Contributions: André Rosa and Nádia Rodrigues assisted in the construction of the genome-waking libraries of *MAT A1* and *MAT A2* strains of *Rhodosporidium toruloides* (PYCC 4416 and PYCC 4661), leading to the identification of mating type-specific *STE20* and *RHA2.A2* genes.

3.1. Introduction

Fungi from the subphylum Pucciniomycotina have remained virtually unexplored as to the content and organization of the *MAT* loci. Preliminary work with the bipolar anther smut *Microbotryum violaceum* provided the first clues on the gene content of *MAT* loci in this group, but the respective genetic organization remained to be elucidated (Yockteng et al., 2007) (see 1.2.2.7).

The Pucciniomycotina also comprises many saprobic yeast taxa including the red yeasts of the genera *Rhodosporidium* and *Sporidiobolus* (Sampaio et al., 2003). All know heterothallic *Rhodosporidium* species were isolated as haploid yeasts and have a bipolar mating behavior, i.e., their strains belong to either one of two complementary mating types, designated *A1* and *A2* or *A* and *a* (Abe et al., 1975; Fell and Statzell-Tallman, 1998; Sampaio, 2011a). Therefore, based on the current understanding of basidiomycete *MAT* loci, it was hypothesized that the genes encoding the pheromones, pheromone receptors, and the homeodomain transcription factors would be part of a single *MAT* locus in *Rhodosporidium* species. Pioneering work by Japanese researchers led by S. Fukui in the late 1980s identified three genes (*RHA1*, *RHA2* and *RHA3*) encoding the pheromone precursors in a *MAT A1* (or *A*) strain of *Rhodosporidium toruloides* (Akada et al., 1989a). These genes were not detected in a complementary *MAT A2* (or *a*) strain (Akada et al., 1989b) and have been suggested to be part of the *MAT* locus (Bölker and Kahmann, 1993). Fukui and coworkers had previously characterized the early stages of the mating process in *R. toruloides* (Abe et al., 1975) and suggested that *MAT A1* cells constitutively produced a pheromone (*A*-factor, later named rhodotorucine *A*), which induced the production of *a*-factor by *MAT A2* cells. The studies of Akada et al. (Akada et al., 1989a; Akada et al., 1989b) were not continued, and no additional information on the *MAT* loci of *Rhodosporidium* species was published since then.

This chapter presents evidence that a gene encoding a p21-activated kinase (PAK; Ste20-like) possibly involved in pheromone signaling is *MAT*-specific and that, along with the multiple pheromone precursor genes, it belongs to the *R. toruloides MAT* locus. It was also found that this genomic region displayed a high level of synteny with homologous genomic regions from the closely related yeast *Sporobolomyces roseus*, whose genome sequence is presently available at Joint Genome Institute (JGI). Genomic fragments of a close relative of the latter species, *Sporidiobolus salmonicolor*, were also available in form of Trace Archives (short unassembled genomic DNA fragments) and provided an additional source of sequence data for comparison. Analysis of the data on *R. toruloides* together with that of the *Sp. roseus* genome and of the *S. salmonicolor* provided the first look into the *MAT* locus structure in red yeasts.

3.2. Material and methods

3.2.1. Yeast strains, culture conditions and mating tests

The following strains of *R. toruloides*, obtained from the Portuguese Yeast Culture Collection (CREM; Portugal), were studied in detail: PYCC 4416 (CBS 14; *MAT A1*), PYCC 5082 (*MAT A1*), PYCC 4417 (CBS 5745; *MAT* A2), and PYCC 4661 (*MAT A2*). Cultures were grown on MYP agar (see 2.2.8) at room temperature. To check sexual compatibility, pairs of 2- to 4-day-old cultures were mixed on an MYP agar plate, incubated at room temperature, and examined microscopically after 1 week for the production of mycelium with clamp connections and teliospores, using phase-contrast optics. An enlarged set of strains was used to check mating type specificity of pheromone genes by PCR, in addition to the four strains mentioned above, namely, CBS 350, CBS 315, MSD 231, A421, and A415 (*MAT A1*); CBS 349, PYCC 4786, PYCC 4943, PYCC 5109, PYCC 5081, A412, A401, A413, and A456 (*MAT A2*). MSD and A strains are environmental isolates obtained by Gadanho et al. (Gadanho et al., 2006).

3.2.2. DNA extraction and common settings for PCR amplification

Isolation of genomic DNA was performed essentially as previously described (Sampaio et al., 2001), with a few modifications. After centrifugation, the cell extracts were submitted to digestion with proteinase K (1 mg/ml; 1 h, 37°C) and, subsequently, to phenol (pH 8.0) and chloroform-isoamyl alcohol (24:1) extractions. Nucleic acids were dissolved in 100 μl of distilled water after precipitation. PCRs were performed in a final volume of 25 μl and contained the following components (unless stated otherwise): 2 mM of $MgCl_2$, 0.25 mM of each of the four deoxynucleoside triphosphates (GE Healthcare), 0.8 μM of primer, 5 μl of template (genomic DNA was diluted 1:750), and 1 U Taq DNA polymerase (Fermentas, Canada). Thermal cycling consisted of a 5-minute denaturation step at 95°C, followed by 35 cycles of denaturation at 94°C for 30 s, 30 s at the annealing temperature (variable), and extension at 72°C (variable time); the annealing temperatures and extension times used in each reaction, as well as the sequences of all the primers, are listed in Table 3.1. A final extension of 7 min at 72°C was subsequently performed. Amplification products were purified using the GFX PCR DNA and gel purification kit (GE Healthcare) and either cloned into the pMOSBlue vector with the pMOSBlue blunt-ended cloning kit (GE Healthcare) and sequenced or sequenced directly (sequencing was performed by STABVida, Portugal).

3.2.3. Mating type specificity of *RHA* genes

The *MAT A1* specificities of the *RHA1*, *RHA2*, and *RHA3* pheromone precursor genes were checked by PCR amplification with gene-specific primers based on the sequences previously determined (Akada et al., 1989b) and using genomic DNA from both *MAT A1* and *MAT A2* strains as template. One forward primer common to the three genes (RHAFw) was used together with one reverse primer specific for each gene (RHA1Rev, RHA2Rev, and RHA3Rev) (Table 3.1). Following the identification of the *RHA2.A2* gene, mating type specificity was checked in an enlarged set of strains using reverse primers specific for either the *RHA2* or the *RHA2.A2* coding regions (MC075 and MC071, respectively) and a common forward primer based on a conserved *STE20* region in the two mating types (Ste20Dw2A2) (Table 3.1).

3.2.4. Genome walking

A genome walking approach using the Universal Genome Walker kit (BD Biosciences Clontech) was employed to characterize the *MAT* loci of *R. toruloides*. This method consists of preparing a set of genomic DNA libraries each requiring the ligation of adapters to the ends of restriction fragments obtained by complete digestion with a particular enzyme (blunt cutters). These libraries are subsequently screened using combinations of gene-specific and adapter primers (AP1 and AP2). For both *MAT* A1 (PYCC 4416) and *MAT A2* (PYCC 4661) strains, the genomic libraries were prepared using EcoRV, PvuII, and StuI. Each Genome Walker library was used as a template in independent amplification reactions. Primer design and PCR conditions followed the manufacturer's instructions. The fragment upstream of the *RHA2* gene (fragment 1; Figure 3.2) was obtained from the EcoRV library, whereas the two fragments encompassing part of the *STE20* gene in *MAT A2* strains (fragments 5 and 6; Figure 3.2) were amplified from the StuI *MAT A2* library.

3.2.5. PCR amplification and sequencing of the *STE20* alleles

A fragment encompassing the complete region between the *RHA2* and *RHA3* genes was obtained by PCR using gene-specific primers RHA3Up2 and SCP2Up (Figure 3.2 and Table 3.1) and genomic DNA from strain PYCC 4416 (*MAT A1*) as a template. The primer pairs Ste20CRFw with Ste20CRRev and NCARFw with NCARRev were based on regions of the *MAT A1 STE20* allele that were found to be conserved in the closely related species *Sp. roseus* and *S. salmonicolor*. These primers were used to amplify most of the *RtSTE20* gene in the remaining strains under study (*MAT A1* and *MAT A2*). The PCR products were purified and directly sequenced by primer walking.

Table 3.1. List of primers and specific PCR conditions used in Chapter 3. Degenerate nucleotides follow the IUPAC nomenclature. [a]primers supplied in the Universal Genome-Walker kit (BD Biosciences Clontech). All the others were obtained from MWG Biotech AG. Primers inside brackets were used in multiple amplifications.

Primer Name	**Sequence** (5' – 3')	**Annealing / Extension**	**Remarks**
RHAFw	CCTTGTTCTCGGTGGCTATTT	55ºC / 90 s	*RHA* forward common primer
RHA1Rev	GGCGGATGGATTGGTTGAGC		*RHA1*, *RHA2*, *RHA3* reverse specific primers; Confirm mating type specificity of *RHA* genes
RHA2Rev	CTTGGCCCGCCCTATGTCTG		
RHA3Rev	CTCCGCTGTCGTGTACTGATG		
RHA2Up	GCAGACGAACAAACGATCATGGAACA	65ºC / 3 min	Amplify the region upstream of the *RHA2* gene (Genome-walking libraries of PYCC 4416, *MAT A1*)
(AP1)[a]	GTAATACGACTCACTATAGGGC		
RHA2Up2	AAGGGCAGACGACAAGAACAAAGGCC	65ºC / 3 min	Amplify the region upstream of the *RHA2* gene (nested PCR) **[fragment 1, Figure 3.2]**
(AP2)[a]	ACTATAGGGCACGCGTGGT		
RHA3Up2	GAGAACCACGGACTGTCGAGGTCGAG	63ºC / 3 min	Amplify the region between the *RHA2* and *RHA3* genes **[fragment 2, Figure 3.2]**
SCP2Up	GGCTGCGGAAGACAACGACTG		
STE20CRFw	ACAGACATCATCGACAGCAAC	55ºC / 1 min	Amplify the conserved regions of the *RtSTE20* alleles (*MAT A1* and *MAT A2*) **[fragments 3 and 4, Figure 3.2]**
STE20CRRev	CGTACTCCTTCTGCTTGACGA		
NCARFw	CGRGAAGACCTGAARCCGTT	59ºC / 90 s	
NCARRev	CATGWACTCCATGACGACCCA		
STE20DwA2	CGAGTGCTGACAGTCGCATCTTTCCG	67ºC / 3 min	Amplify the region downstream of the *RHA2.A2* gene (Genome-walking libraries of PYCC 4661, *MAT A2*)
(AP1)[a]	GTAATACGACTCACTATAGGGC		
(STE20Dw2A2)	CGAAGCTGACCGACCAGAAGTCGAAG	65ºC / 3 min	Amplify the region downstream of the *RHA2.A2* gene (nested PCR) **[fragment 5, Figure 3.2]**
(AP2) [a]	ACTATAGGGCACGCGTGGT		
STE20UpA2	ACCGCAAGGCGAGGTAGAAGGTTCGA	67ºC / 3 min	Amplify the region upstream of the *RtSTE20.A2* gene (Genome-walking libraries of PYCC 4661, *MAT A2*)
(AP1) [a]	GTAATACGACTCACTATAGGGC		
STE20Up2A2	AGGTAGAAGGTTCGAGCCGGTGTCGT	65ºC / 3 min	Amplify the region upstream of the *RtSTE20.A2* gene (nested PCR) **[fragment 6, Figure 3.2]**
(AP2) [a]	ACTATAGGGCACGCGTGGT		
NCAR3Fw	TGGCATTGGCGAGCMCGMAA	55ºC / 90 s	Amplify part of the *STE3* homologue in *R. kratochvilovae* (PYCC 4818, *MAT A1*)
NCAR4Rev	GRCCRAAGAAGASGAARAAGA		
NCAR5STE3Fw	CATCTTCTGGTGTKTGGTCAMCGT	60ºC / 1 min	Amplify part of the of the *STE3* homologue in *R. toruloides* (PYCC 4416, *MAT A1*)
NCAR6STE3Rev	GGGASGSTGTAGGCGATTTG		
(STE20Dw2A2)	CGAAGCTGACCGACCAGAAGTCGAAG	66ºC / 90 s	Amplify a *MAT A1*-specific region (3' end of *RtSTE20* gene to the CDS of *RHA2)*
MC074	GGGTCCACGAGATTTCGGGGTACTTG		
(STE20Dw2A2)	CGAAGCTGACCGACCAGAAGTCGAAG	63ºC / 90 s	Amplify a *MAT A2*-specific region (3' end of *RtSTE20* gene to the CDS of *RHA2.A2)*
MC071	GGTGCATGCCATGTACGAGGTGACA		

3.2.6. PCR amplification of putative *MAT A1* pheromone receptor (*STE3.A1*)

For the identification of pheromone receptor genes in *R. toruloides*, the rcb1 gene from *Coprinopsis cinerea* (GenBank: Y11082) was used to perform a Tblastx search in the *Sp. roseus* genome sequencing project database (JGI; http://genome.jgi-psf.org/Sporo1/Sporo1.home.html). Positive hits were retrieved and used to search the NCBI Trace Archive sequences of *S. salmonicolor*, yielding an assembled sequence for the *SsSTE3* gene. The *SsSTE3* and *SrSTE3* genes were in turn aligned with other basidiomycete pheromone receptor genes [*PDSTE3.3* gene from *Pleurotus djamor* (GenBank: AY462110), *bbr1* from *Schizophyllum commune* (GenBank: U74495), and *STE3a* from *Cryptococcus neoformans* var. *neoformans* (GenBank: XM_570116)], and the conserved regions were used to design degenerate primers (NCAR3Fw and NCAR4Rev, Table 3.1). These degenerate primers were used in PCRs with genomic DNA from *R. toruloides* but failed to yield an amplification product. A similar reaction using as template genomic DNA from a closely related species (*Rhodosporidium kratochvilovae* PYCC 4818; *MAT A1*) (Sampaio et al., 2001) produced an amplicon which was directly sequenced and found to encode a partial open reading frame (ORF) of a *STE3*-like pheromone receptor. The new sequence was used to refine the design of a second set of primers for *R. toruloides*. These primers (NCAR5STE3Fw and NCAR6STE3Rev) (Table 3.1) permitted the amplification of part of the *STE3* homologue in *R. toruloides*. The position and orientation of the *STE3* homologue relative to the *RHA1* gene in *R. toruloides* were checked by PCR with gene-specific primers (NCAR5STE3Fw and RHA1Rev) (Table 3.1). PCRs were performed in a final volume of 50 μl and contained the following components: 1× Long PCR buffer, 2 mM of $MgCl_2$, 0.2 mM of each of the four deoxynucleoside triphosphates (GE Healthcare), 1.0% DMSO, 1.0 μM of primer, 5 μl of template (genomic DNA diluted 100-fold), and 2.5 U Long PCR enzyme mix (Fermentas, Canada). Thermal cycling consisted of a 3-minute denaturation step at 94°C, followed by an initial round of 10 cycles of denaturation at 95°C for 20 s, annealing at 60°C for 30 s, and extension at 68°C for 6 min and a second round of 25 cycles, increasing the extension time 2 s in each cycle. A final extension of 10 min at 68°C was performed. The nucleotide sequences of the ends of the amplified fragment were obtained by direct sequencing using the same primers.

3.2.7. In silico analysis

Primers were designed using Primer Premier 5.0 software (Premier Biosoft International). The assembly of DNA sequences was performed using SeqMan II software (v.6.1) from DNASTAR. Gene models were constructed by combining nucleotide homology searches in the NCBI database, using Blastn and tBlastx and in silico predictions of putative introns and coding sequences using AUGUSTUS software (http://augustus.gobics.de/submission) (Stanke et al., 2008). Blastp searches in NCBI were used to check for conserved domains within the deduced amino acid sequences and

to assign putative gene functions. The genomic DNA sequences of *Sp. roseus* (strain IAM 13481) showing a high degree of identity to the *STE20* homolog of *R. toruloides* PYCC 4416 (*MAT A1*) were retrieved from the first release of the *Sp. roseus* genome sequencing project (scaffold 9, contig 11). Annotation of the 40-kb region encompassing the *STE20* gene was completed using Tblastx. The NCBI Trace Archive sequences of *S. salmonicolor* (strain IAM 12258) contained sequences with a high percent sequence identity to that of the *STE20* gene of *Sp. roseus*. The relevant *S. salmonicolor* fragments were assembled stepwise from the *STE20* region. Following homology searches using the *SsRHA2* gene, other two genes (*SsRHA1* and *SsRHA3*) of *S. salmonicolor* were also retrieved and assembled (Trace Archive sequences are listed in section I.3, Appendix I). Nucleotide and amino acid sequences were aligned using ClustalW 1.4 and T-Coffee (Poirot et al., 2003), respectively. Tblastx was used to search the entire genome of *Sp. roseus* for homologues of the homeodomain transcription factors (HD1 and HD2). The GenBank accession numbers for the novel *R. toruloides* DNA sequences obtained in this Chapter are listed in section I.4 (Appendix I).

3.3. Results and discussion

3.3.1. *MAT* status of *R. toruloides* strains

To characterize genomic regions involved in mating in *R. toruloides*, the four strains to be studied (two of each mating type) were first tested for their ability to mate and form hyphae with clamp connections and teliospores, characteristic of the sexual cycle in this yeast (Figure 3.1a). All four strains were fertile, *MAT A1* isolates mated only with *MAT A2* isolates, and none of the strains were self-fertile. The mating results were in accord with the presence of the pheromone precursor genes previously described by Fukui and coworkers (Akada et al., 1989b) (*RHA1*, *RHA2*, and *RHA3*), since they were detected by PCR with gene-specific primers only in strains that mated as *MAT A1* in crossing experiments (Figure 3.1b).

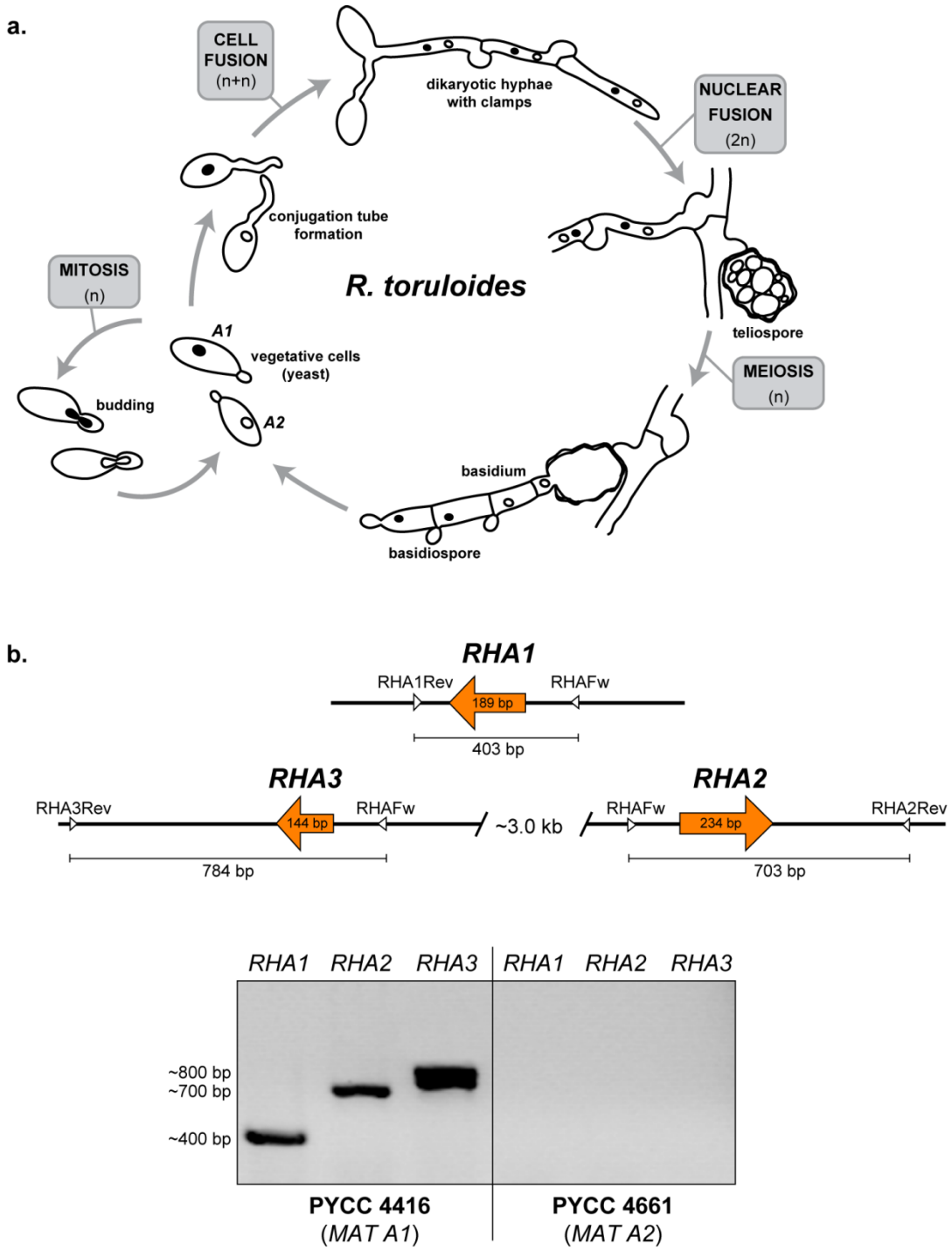

Figure 3.1. Mating behaviour of *Rhodosporidium toruloides*. (a) Diagram representing the sexual and saprobic life cycle of *R. toruloides*. **(b)** Mating type-specificity of the pheromone precursor genes. In the upper part, the organization of *RHA1*, *RHA2* and *RHA3* genes is depicted for strain PYCC 4416 (*MAT A1*). Coding regions are shown as orange arrows, indicating the direction of transcription. Gene-specific primers are indicated by white arrowheads that amplify *RHA1*, *RHA2* and *RHA3* gene-specific regions. Results of the PCR amplification of the three gene-specific regions are shown on the bottom for two strains of opposite mating type.

3.3.2. Genomic regions in the vicinity of the pheromone genes in *MAT A1*

Two of the pheromone precursor genes previously identified (*RHA2* and *RHA3*) were found to be located close to each other (~3.0 kb) on opposite DNA strands (Akada et al., 1989b), while the location of the third gene (*RHA1*) in relation to the former remained to be established (Figure 3.1b). Since in basidiomycetes the genes encoding pheromone precursors have so far been found almost invariably inside the *MAT* loci, sequences of the three *RHA* genes were used as anchors to explore the adjacent regions, presumed to be part of the *MAT A1* locus in this species. To this end, a genome walking approach was designed that resulted in the amplification of a 0.7-kb genomic DNA fragment upstream of the *RHA2* gene (Figure 3.2, fragment 1). Sequencing of this fragment revealed two notable features: first, a region of 204 bp located immediately upstream of the *RHA2* gene was found to be identical to the corresponding region upstream of the *RHA3* gene previously sequenced (Akada et al., 1989b); secondly, a region of approximately 200 bp at the 5' end of the fragment turned out to encode a peptide with strong similarity to the catalytic domain of PAK kinases (Hofmann et al., 2004). To investigate whether this was part of a PAK kinase gene, the entire intervening genomic region between the *RHA2* and *RHA3* genes was amplified and sequenced by primer walking (Figure 3.2, fragments 1 and 2). This region was indeed found to contain a putative PAK kinase gene (named *RtSTE20*), homologous to genes found in *U. maydis*, *C. neoformans*, and *Pneumocystis carinii* (Smulian et al., 2001; Wang et al., 2002; Smith et al., 2004). An in silico analysis predicted the presence of at least four introns in this gene.

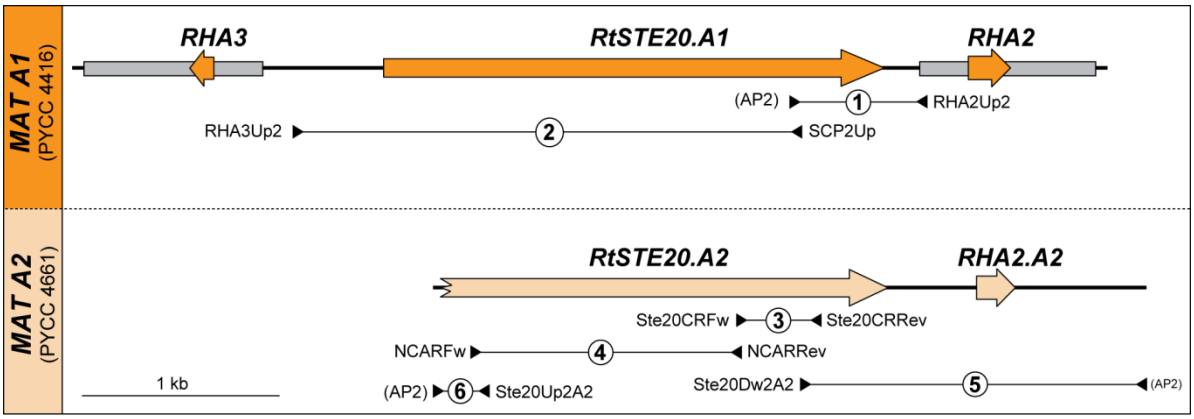

Figure 3.2. Structure of the *STE20* region of the *MAT A1* and *MAT A2* loci of *R. toruloides*. Coding regions are shown for the indicated *MAT A1* and *MAT A2* strains as orange or light orange arrows, respectively, indicating the direction of transcription. The transcribed regions of *RHA2* and *RHA3* genes previously identified (Akada et al., 1989b) are shown as gray boxes. Genomic DNA fragments 1 to 6 were obtained either by genome walking or by PCR with gene-specific primers. The primers used in each case are indicated by black arrowheads (see Table 3.1).

Since the *RtSTE20* gene was found between two mating type-specific genes, it was anticipated that it could be part of the *MAT A1* locus. In this case, its sequence would be expected to exhibit mating type-specific polymorphisms. To investigate this, the *RtSTE20* gene was amplified and sequenced almost completely in the four strains under study (the regions sequenced, approximately 1.7 kb, correspond to fragments 3 and 4 depicted in Figure 3.2). The sequences were indeed found to be almost identical between alleles obtained from strains of the same mating type (eight nucleotide differences between the two *MAT A1* strains and three differences between the two *MAT A2* strains). However, sequences obtained for the opposite mating types had approximately 80 single nucleotide differences, which would lead to twelve amino acid changes in the conserved domains of the RtSte20 protein (Figure 3.3). The large majority of the nucleotide differences between the two alleles were, nevertheless, located in less-conserved regions of the *RtSTE20* gene. This observation supports the conclusion that *RtSTE20* is a mating type-specific gene in *R. toruloides*.

3.3.3. Identification of a *MAT*-specific region in *MAT A2* of *R. toruloides*

To explore the organization of the *MAT A2* locus of *R. toruloides*, the region around the *RtSTE20* gene in *MAT A2* strain PYCC 4661 was analyzed using an identical genome walking approach. A fragment of 2 kb downstream of the *STE20* gene was amplified and sequenced in this strain (fragment 5, Figure 3.2) revealing a completely dissimilar sequence when compared to the same region in a *MAT A1* strain (encoding the *RHA2* gene). To ascertain if this region also encoded a pheromone precursor, the sequence was examined for the presence of an open reading frame exhibiting characteristics in keeping with the general features of previously characterized pheromone precursor genes (Bölker and Kahmann, 1993; Caldwell et al., 1995). Indeed, a region was found potentially encoding two identical copies of a 13-amino-acid peptide and exhibiting at the C-terminus a sequence of four amino acids (CAAX motif) very similar to that of the *MAT A1* pheromone precursors (Figure 3.4). The most likely initiation codon is preceded by two CT-rich regions (Figure 3.4), one of which is identical to that found upstream of the transcription initiation site of the *RHA* (*MAT A1*) genes (Akada et al., 1989b). Therefore, this novel gene, named *RHA2.A2*, is proposed to encode a precursor of the peptide moiety of the *MAT A2* pheromone (rhodotorucine *a*) of *R. toruloides*.

In order to check the association between mating behaviour and the presence of the newly identified *RHA2.A2*/*STE20* region, a specific primer based on the *RHA2.A2* coding region was used in a *MAT A2* diagnostic PCR in combination with a *RtSTE20* forward primer. A total of 14 additional strains (five *MAT A1* and nine *MAT A2*), whose mating behaviour was previously checked, were tested for the presence of the putative *MAT A2* locus sequences. As expected, an amplicon was obtained only for *MAT A2* strains (Figure 3.5). Conversely, when the same *RtSTE20* primer was used in combination with a primer based on the *RHA2* (*MAT A1*) coding sequence in a PCR assay with the same strain set, only *MAT A1* strains yielded an amplification product.

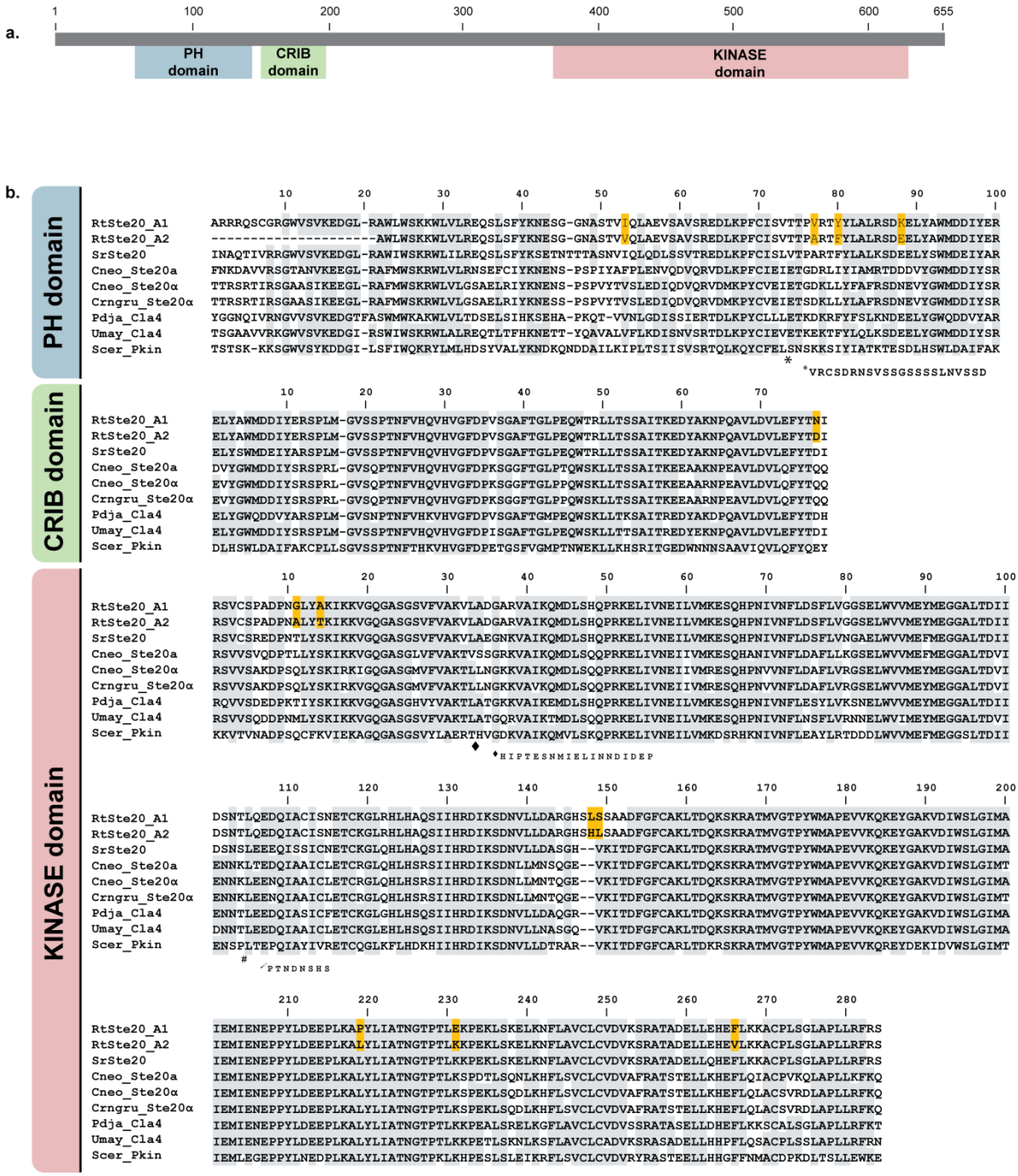

Figure 3.3. Conserved domains of fungal PAK kinases. (a) The position of the three conserved domains is shown along the *R. toruloides* RtSte20_A1 protein. **(b)** Alignment of the three conserved domains of several fungal PAK kinases: RtSte20_A1 and RtSte20_A2 (*Rhodosporidium toruloides*, *MAT A1* and *MAT A2*); SrSte20 (*Sporobolomyces roseus*), Cneo_Ste20a and Cneo_Ste20α (*Cryptococcus neoformans*), Cngru_Ste20α (*C. neoformans* var. *grubii*), Pdja_Cla4 (*Pleurotus djamor*), Umay_Cla4 (*Ustilago maydis*) and Scer_Pkin (*Saccharomyces cerevisiae*). Insertions that occur only in the Scer_Pkin sequence are indicated by *,♦,#. Identical amino acids are shaded in gray and those that differ between the two mating types of *R. toruloides* are highlighted in yellow. Sequence accession numbers are given in Figure 3.7.

```
caaccgtctcttgtgcttcgttcctcccgcgctggcctcctccgctctcatgcgtcgtctccctttgttc -247
ctctcacgacggtcgagattggctcggcaccactggcacaatagtgcttcctcgattgtgctctggcgca -177
caacggattctacagacaaaggccgacttgcgcacaaaggcatgctcctccccttttgtcctctcaggcta -107
caaaaaggtcggtttcaccgtctctcctctctccttcctcacacctcctcaccctcactcgtcttcgtct  -37
tcctccaccacccatctcccacttccaaccgaaacaATGGCCCCCATCACCCTCGCCGCCAACCAGGAGG   34
                                    M  A  P  I  T  L  A  A  N  Q  E

TCGCTGCCGCCGCCCCTGCTCTCGTCAACGAGGAGCACAACCCGACTGTCACCTCGTACATGGCATGCAC  104
V  A  A  A  A  P  A  L  V  N  E  E  H  N  P  T  V  T  S  Y  M  A  C  T

CGTCTCGCGCAAGGACGTCTCGGTCGAGATGGTTGACGAGGAGCACAACCCGACCGTCACCTCGTACATG  174
   V  S  R  K  D  V  S  V  E  M  V  D  E  E  H  N  P  T  V  T  S  Y  M

GCTTGCACGATCGCCTAAtcctccgtaagtccctctgcttcctcatggcagtcctgtgctaacttcgtcc  244
A  C  T  I  A  *
cattggcagcgttgcagggtggtcgcaagctgcgagacatgaggattgaggggaagatttccggggcgga  314
agcgaggatgtactcaagcccctagcatggctgtggttcttttgtatgcaacatcttgtcgtacacttcc  384
```

Figure 3.4. Nucleotide sequence of the genomic region encoding the *RHA2.A2* gene, showing the predicted pheromone precursor sequence. CT-rich regions possibly involved in transcription initiation are underlined. The two repeats proposed to represent the peptide moiety of the mature pheromone are shadowed. The sequences resembling the CAAX prenylation motif are enclosed in boxes.

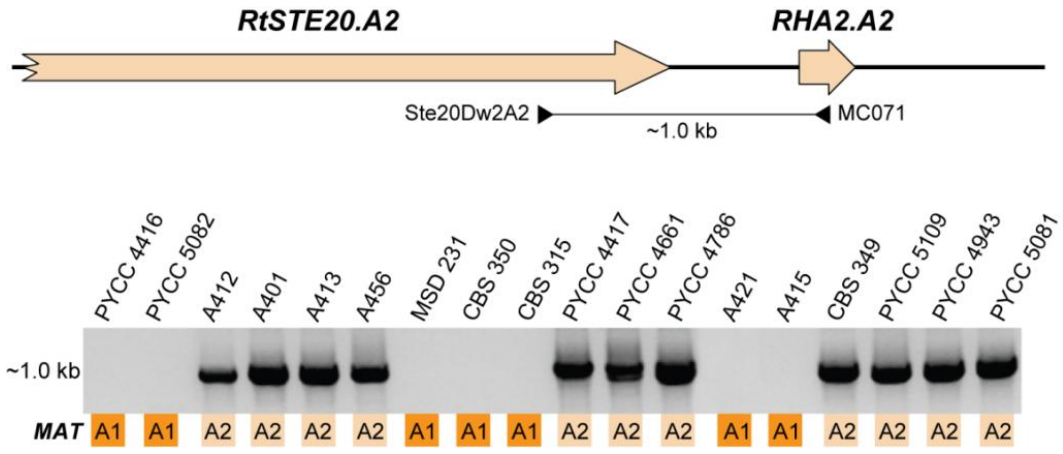

Figure 3.5. Mating type-specificity of the *RHA2.A2*/*STE20* region. Coding regions are shown as light orange arrows, indicating the direction of transcription. Primers sites specific for this region are indicated by black arrowheads. Results of the PCR amplification of *RHA2.A2*/*STE20* region are shown for several *R. toruloides* strains of both mating types (*MAT A1* or *MAT A2*).

Together, these results strongly suggest that this region is indeed part of the *MAT* locus of *R. toruloides*. Notably, one *R. toruloides* isolate (A399) (Gadanho et al., 2006), which was not included in the previous strain set because it exhibited ambiguous mating behavior (apparent sexual compatibility with several *MAT A1* and *MAT A2* strains), tested positive in both *MAT A1* and *MAT A2* diagnostic PCR assays, although amplification of *MAT A1* sequences was less efficient than observed for *MAT A1* strains.

3.3.4. Comparison with syntenic genomic regions from related red yeasts

Taking advantage of the recently released genome data of *Sp. roseus*, a search for genes with a high sequence identity with the catalytic domain of *RtSTE20* identified a gene likely to represent a *RtSTE20* homologue. Notably, this gene (named henceforth *SrSTE20*) is also flanked by two divergently transcribed putative pheromone precursor genes, which were named *SrRHA2* and *SrRHA3*, based on synteny with the *RHA2* and *RHA3* genes from *R. toruloides*. Unlike the latter, the two genes from *Sp. roseus* are completely identical not only in the coding sequence but also in the promoter and terminator regions. For *S. salmonicolor*, although genomic sequences are only available as Trace Archives, a search on this database revealed the presence of a gene very similar to the *STE20* genes of *Sp. roseus* and *R. toruloides*. This gene (*SsSTE20*) was used as an anchor to assemble a contiguous fragment that extended to approximately 2 kb downstream of its 3' end, where a putative pheromone precursor gene (named *SsRHA2*) was similarly found (Figure 3.6).

In silico analysis of the three PAK kinase orthologues (*RtSTE20*, *SrSTE20* and *SsSTE20*) revealed that they encode proteins that possess the two functional domains common to the entire PAK kinase family, namely, a highly conserved C-terminal catalytic domain and a CRIB domain (cdc42/ras interaction binding domain) (Figure 3.3). In addition, they have a third domain (PH-pleckstrin homology) (Figure 3.3 and 3.7) that is present in a subset of the PAK family members, such as the Ste20 kinases of *C. neoformans* and *C. gattii* (Wang et al., 2002) and the Cla4 kinase of *Ustilago maydis* (Leveleki et al., 2004). The former proteins were shown to be involved in mating or required for the onset of the filamentous phase and are present in the *MAT* loci of the two *Cryptococcus* sibling species. However, the *MAT*-specific alleles of *STE20* have diverged considerably more in *C. neoformans* than in *R. toruloides* (approximately 70% versus 90% sequence identity) (Figure 3.7). It is therefore unlikely that the two *R. toruloides* proteins exhibit functional dissimilarities comparable to those found for the *C. neoformans MAT*-specific kinases (Wang et al., 2002).

In *Sp. roseus*, besides the *SrRHA2* and *SrRHA3* genes, a third identical gene (named *SrRHA1*) was found downstream of a gene probably encoding a 30- to 40-kDa subunit of RNA polymerase III (*RNAPOL*) (Figure 3.6). Notably, the 3' end of the coding sequence of the *R. toruloides* homologue of this *RNAPOL* was found upstream of the *RHA1* gene in the genomic DNA fragment sequenced by Akada et al. (1989b). In *S. salmonicolor* a new search in the Trace Archives using the sequence of

the *SsRHA2* gene led to the identification of two additional genes potentially encoding pheromone precursors. While for one of the genes (*SsRHA3*) the longest possible contiguous fragment did not provide any clue as to its location, the second gene (*SsRHA1*) was also found downstream of a partial ORF encoding the 30- to 40-kDa subunit of RNA polymerase III (Figure 3.6). The *SsRHA1* and *SsRHA2* coding sequences are completely identical.

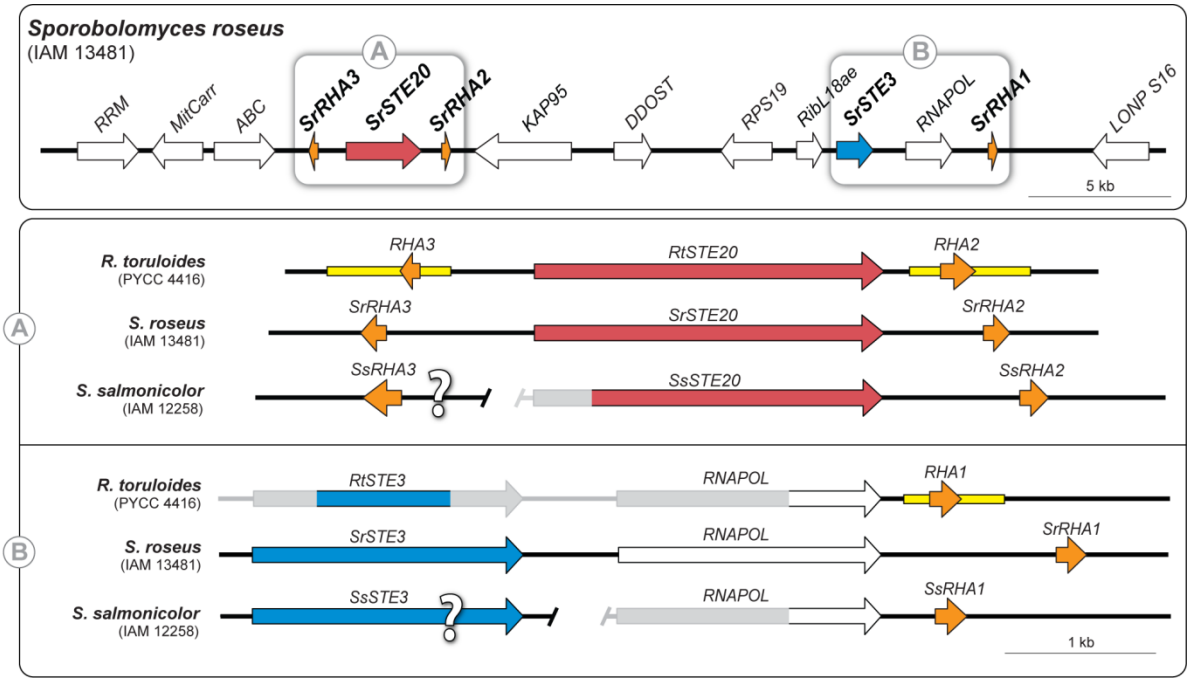

Figure 3.6. Structure of a putative *MAT* locus in *Sp. roseus* and comparison with homologous genomic regions of *R. toruloides* and *S. salmonicolor*. The top panel shows the gene content and organization in a 40-kb region around the putative pheromone precursor (box A) and pheromone receptor (box B) genes in *Sp. roseus*. The lower panels show the gene content and organization of the regions around the pheromone precursor (A) and the pheromone receptor (B) homologues in the three species. Orthologues are shown in the same colour and the coding regions are depicted by arrows, indicating the direction of transcription. The transcribed regions of *RHA* genes in *R. toruloides* identified by Akada et al. (1989b) are shown as yellow boxes. Unknown sequences are shown in light gray and the relative position and orientation of the genes marked "?" are also unknown. Additional genes identified in a 40-kb region of *Sp. roseus* (top panel) are depicted as white boxes and encode (from the left to the right): putative RNA binding protein of the RRM super family (*RRM*); putative mitochondrial carrier protein (*MitCarr*); putative ABC1 protein, mitochondrial precursor/ ubuiquinone biosynthesis protein coq 8 (*ABC*); putative Karyopherin β 1, importin (*KAP95*); dolichyl diphospho-digosaccharide protein glycosytransferase 48 KDa subunit (*DDOST*); ribosomal protein S19/S15 (*RPS19/15*); ribosomal L18ae protein (*RibL18ae*); putative DNA directed RNA polymerase III, 30/40 KDa subunit (*RNAPOL*); peptidase S16, Lon protease (*LONP S16*).

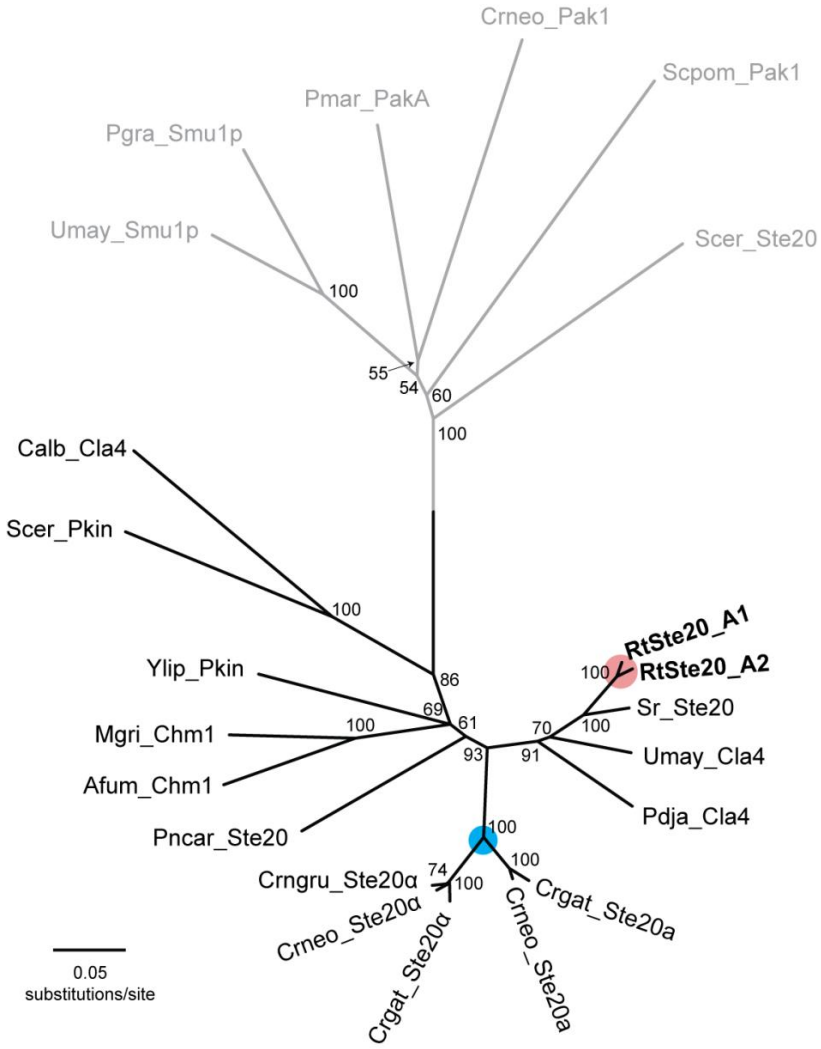

Figure 3.7. Phylogenetic tree showing the relationships between PAK kinases from different fungi, based on alignment of the respective protein sequences. Kinases lacking the PH domain are located in the gray branch of the tree. The evolutionary distances were computed using Protdist (JTT matrix-based method) (Whelan and Goldman, 2001). Neighbour-joining (Saitou and Nei, 1987) and MEGA4 (Tamura et al., 2007) were used to construct and visualize the phylogenetic tree. Numbers on branches are bootstrap values inferred from 10 000 replicates (values below 50% are not shown). Red and blue circles indicate the divergence point between mating type-specific sequences in both *Rhodosporidium toruloides* and *Cryptococcus neoformans*. Sequences used in the tree and their respective GenBank accession numbers are: *Ustilago maydis* (Umay_Cla4, AAT39367; Umay_Smu1p, Q4P5N0); *Pleurotus djamor* (Pdja_Cla4, AAS46755); *C. neoformans* (Cneo_Ste20α, XP_570118; Cneo_Ste20a, AAN75615); *C. neoformans* var. *grubii* (Crngru_Ste20α, AAN75173); *C. gattii* (Crgat_Ste20a, AAV28761; Crgat_Ste20α, AAV28796); *Pneumocytis carinii* (Pncar_Ste20, AAG38545); *Yarrowia lipolytica* (Ylip_Pkin, XP_502152); *Aspergillus fumigatus* (Afum_Chm1, XP_754042); *Candida albicans* (Calb_Cla4, O14427); *Magnaporthe grisea* (Mgri_Chm1, XP_369805); *Saccharomyces cerevisiae* (Scer_Pkin, CAA57879; Scer_Ste20p, NP_011856); *Penicillium marneffei* (Pmar_PakA, Q2VWQ3); *Schizosaccharomyces pombe* (Scpom_Pak1p, AAC49125); *R. toruloides* (RtSte20_A1, EU386160; RtSte20_A2, EU386161); *Sp. roseus* (Sr_Ste20, Scaffold 9, contig 11, JGI genome database); *Puccinia graminis* (Pgra_Smu1, PGTG_14316.2, supercontig 51: 338996-341873, Broad Institute of MIT and Harvard genome database).

The rhodotorucine *A* (*RHA*) genes were the first pheromone precursor genes whose structure was determined in basidiomycetes (Akada et al., 1989b). The same authors had previously determined the chemical structure of the mature pheromone as a lipopeptide (farnesyl-undecapetide), in which the C-terminal cysteine of the peptide moiety was shown to be covalently linked to a farnesyl group through a thioether bond (Kamiya et al., 1978a; Kamiya et al., 1978b). That was the first demonstration of this type of posttranslational modification, which was subsequently found in many other fungal pheromones, such as the a-factor of *Saccharomyces cerevisiae*, as well as in other eukaryotic proteins (Bölker and Kahmann, 1993; Caldwell et al., 1995). Proteins that are subjected to this posttranslational modification usually contain a C-terminal CAAX motif where C stands for cysteine, A for an aliphatic amino acid and X for any amino acid (often an alanine in fungal pheromones) (Caldwell et al., 1995). However, the C-terminal motif of the rhodotorucine *A* precursor is actually CTVA, where threonine does not conform to the general pattern of fungal lipopeptide pheromones, which have either valine, isoleucine or leucine as the middle amino acids (Figure 3.8) (Akada et al., 1989b; Caldwell et al., 1995; McClelland et al., 2002). Nevertheless, the former amino acid is also compatible with a potential farnesylation of the terminal cysteine of the peptide moiety (Caldwell et al., 1995) as seen, e.g., in four of the six pheromone precursors predicted in the smut fungus *Sporisorium reilianum*, which have a terminal CTIA motif (Schirawski et al., 2005).

The rhodotorucine *A* precursors are, nevertheless, unique among basidiomycete lipopeptide pheromones, since they contain multiple tandem copies of the peptide moiety (Akada et al., 1989b; Caldwell et al., 1995). In each repeat, a CAAX motif is present and is followed by a lysine residue (spacer) that has been proposed as a signal for proteolytic processing (Akada et al., 1989b; Bölker and Kahmann, 1993). In contrast, the spacer sequences of the predicted precursors of *Sp. roseus* and *S. salmonicolor* have both a lysine or arginine residue (basic amino acids) and an additional amino acid (threonine, serine or valine), which indicates that the recognition site for proteolysis to produce a mature pheromone, may be different (Figure 3.8). Another striking difference between the *RHA* genes of *R. toruloides* and those found in the *Sp. roseus* and *S. salmonicolor* species is that the C-terminal motifs of the latter species appear not to correspond to CAAX boxes. However, a lysine residue is present in the C-terminus of the *Sp. roseus* and *S. salmonicolor* precursors immediately after a putative CAAX box and may signal a proteolytic processing site that could produce a CAAX-bearing intermediate. Therefore, spacer peptides presumably attached to the C-terminus of the pheromone peptide may be also removed (Figure 3.8).

Interestingly, the putative rhodotorucine *a* precursor predicted from the *MAT A2*-specific *RHA2.A2* gene sequence appears to have also a CAAX-like C-terminal motif (CTIA instead of CTVA) but only two repeats of the peptide moiety, which are apparently preceded by a much longer N-terminus tail and separated by a longer peptide spacer (Figure 3.8). These differences suggest that the pheromone precursor genes of the two opposite mating types of *R. toruloides* may be processed by different mechanisms.

Figure 3.8. Predicted *MAT A1* and *MAT A2* pheromone precursor peptides. (a) Alignment of *MAT A1* pheromone precursor peptides from three red yeast species, as deduced from the sequence of the respective genes: *RHA* (*R. toruloides*), *SsRHA* (*S. salmonicolor*), and *SrRHA* (*Sp. roseus*). Sequences representing the mature pheromones are shadowed. Sequences resembling the CAAX prenylation motif are enclosed in boxes in the C-terminus of each precursor and at their internal repeats. Arrows indicate spacer peptides possibly involved in proteolytic processing. **(b)** Predicted *MAT A2* pheromone precursor of *R. toruloides* as deduced from the respective *RHA2.A2* gene is shown for comparison. Features are represented as in (a).

Akada et al. (1989b) suggested that those multiple repeats in *RHA* genes could have originated by an unequal crossing-over mechanism. They also affirmed that the production of multiple peptides per mole of precursor (as well as the presence of multiple *RHA* genes) may increase the production of the mating pheromone per cell and that this may be a prerequisite for the poorly diffusible small lipophilic peptides to function effectively. Interestingly, the mating reaction in *R. toruloides* appears to be triggered by the production of rhodotorucine *A* by *MAT A1* cells, according to the chain of events suggested by Abe et al. (1975). Therefore, the presence of three pheromone precursor genes in the three species examined, each encoding three to six copies of the mature peptide moiety, suggests that they have in common the production of abundant pheromone as an important feature of the mating process. This could mean that compatible partner recognition occurs mainly or exclusively through pheromone-receptor interactions. This is in contrast with other species, like *U. maydis*, where mate recognition requires a second level of compatibility determined by the multiallelic mating type *b* locus, which encodes the *bE* and *bW* homeodomain transcription factors.

3.3.5. Putative *MAT* locus in *Rhodosporidium* and related red yeast species

In *Sp. roseus* a long contiguous sequence around the *SrSTE20* gene is available. Therefore, in order to find out whether more *MAT*-related genes could be found in the vicinity of this *SrSTE20* core, a genomic region of approximately 40 kb was scrutinized for the presence of ORFs. Notably, a *STE3*-like gene (Hagen et al., 1986), potentially encoding a pheromone receptor, was found close to the *SrRHA1* gene (Figure 3.6, upper panel). Many other genes, apparently unrelated to mating but exhibiting high sequence identity with genes identified in other organisms, were also found in this region and are indicated in Figure 3.6. These results suggest that the entire region between the pheromone receptor core and the *SrSTE20* core may be part of a large *MAT* locus in this yeast. A *STE3*-like gene was also identified in *S. salmonicolor*, but the incompleteness of the genome sequence data did not allow any inference at this point about its location with respect to the other *S. salmonicolor* genes identified.

The two *STE3*-like genes identified in *Sp. roseus* and *S. salmonicolor* share regions of considerable sequence identity. A first pair of degenerate primers based on these regions failed to amplify a *STE3* fragment from *R. toruloides*. However, it was possible to obtain an amplicon when genomic DNA from *Rhodosporidium kratochvilovae*, a species closely related to *R. toruloides* (Sampaio et al., 2001), was used as a template in the PCR. Sequencing of this fragment and its alignment with the abovementioned *STE3*-like genes allowed the design of another set of primers that led to the amplification of the corresponding *STE3* fragment from *R. toruloides MAT A1* strain PYCC 4416, as confirmed by direct sequencing. Amplification of the latter fragment was possible only in *MAT A1* strains, as expected for a mating type-specific pheromone receptor gene. Furthermore, PCR with *RtSTE3* and *RHA1* gene-specific primers (Table 3.1) yielded an amplification product of

approximately 5 kb, consistent with a gene organization identical to that found in *Sp. roseus*. Sequencing of the ends of the amplified fragment finally confirmed the suspected synteny, including the presence of the *RNAPOL* homologue between the *STE3* and *RHA1* genes (Figure 3.6).

In the bipolar basidiomycetes *C. neoformans* and *Ustilago hordei*, the *MAT* locus is large (100 kb and 500 kb, respectively) and multigenic and encompasses both a pheromone/ receptor gene region and a region encoding homeodomain transcription factors (HD1/HD2) (Lengeler et al., 2002; Bakkeren et al., 2006). While genes encoding pheromone/ receptors were found in *Sp. roseus*, no *HD1/HD2* candidates were found upon inspection of the 40-kb genomic region surrounding the *SrSTE20* core. However, a very likely *HD1* candidate was found at a distant location in the genome (scaffold 7, contig 11). Blastp searches in fungal protein databases using part of the sequence of the encoded protein retrieved most of the basidiomycetous HD1 proteins previously characterized. No HD2 candidates with such characteristics were found, although additional genes potentially encoding homeodomain proteins were identified in the *Sp. roseus* genome. The gene encoding the putative *HD1* homologue was recently annotated in the *Sp. roseus* genome but, the current assembly of the sequence data does not allow the precise determination of the distance to the *STE20*/*RHA* region. However, even if they lie on the same chromosome, the distance of the *SrSTE20* core to this *HD1* candidate would be necessarily in excess of 600 kb, as judged from the positions of the genes within the scaffolds; the shortest distance to one of the ends of the respective scaffolds is 213 kb (*STE20*, scaffold 9) and 410 kb (*HD1*, scaffold 7), respectively. The region around this putative *HD1*-like gene was screened for the presence of a *HD2* homolog, since in most basidiomycetous *MAT* loci the *HD1* and *HD2* genes are closely linked and divergently transcribed. However, no *HD2* candidate gene was found, even when the homology searches using tBlastx were extended to the entire genome. Nevertheless, it should be noted that the presence of introns may considerably complicate this kind of search based on a short conserved region, as for example in the case of the *SXI2a* gene (HD2 homologue) of *C. neoformans* (Hull et al., 2005).

Together, these results suggest that the *MAT* locus extends beyond the region inspected in *Sp. roseus* or that the homeodomain transcription factors might not be linked to the *MAT* locus in this species, a situation corresponding to the *Aon MAT* status proposed by Fraser et al. (Fraser et al., 2007) as one of the possible modes of transition from tetrapolar to bipolar mating systems. Another instance of a bipolar basidiomycete whose *MAT* locus structure has been elucidated, the mushroom *Coprinellus disseminatus*, is an example of the reverse situation with respect to *MAT* locus content, since in this case mating type identity is conferred by a short locus containing one or two pairs of closely linked homeodomain transcription factor genes but the pheromone and pheromone receptor genes are not mating type specific (James et al., 2006) (see Figure 1.4b). Moreover, since none of the putative *Sp. roseus MAT* genes examined showed any sign of being degenerated or nonfunctional indicates that *Sp. roseus*, although deemed to be an asexual strain, may in fact correspond to a haploid mating type of a yet-unrecognized sexual species, a situation frequently encountered among basidiomycetous yeasts (Fell et al., 2001). Conversely, *S. salmonicolor* is a heterothallic species, and the sequenced strain (IAM 12258, CBS 483) has a known mating status

(*MAT A2*) (Statzell-Tallman and Fell, 1998). Putative pheromone receptor genes (*STE3* homologues) with a high degree of sequence identity were found in *R. toruloides MAT A1* strains and in the other two red yeast species. In *R. toruloides* and *Sp. roseus* the genes appear to be located upstream of the *RHA1* gene, with the same gene unrelated to mating (a subunit of RNA Pol III) in between. Therefore, also in this region, located in *Sp. roseus* at a distance of approximately 10 kb from the *SrSTE20* core, a strict synteny occurs between the three species (Figure 3.6). Since the pheromone receptor gene remains to be identified in *MAT A2* strains of *R. toruloides*, it is so far unknown whether the *MAT A2* locus is also syntenic in this region. Furthermore, the limited amount of information available for the opposite mating type (restricted to the *RtSTE20.A2* and *RHA2.A2* genes in *R. toruloides*) precludes any conclusion with respect to extended *MAT* synteny conservation between the two mating types. Finally, the strong sequence similarity observed between the Ste3 proteins and even between the putative mature pheromone regions in the three species suggests that the three *MAT* loci depicted in Figure 3.6 represent the "same mating type," analogous to *MAT A1* in *R. toruloides*.

3.4. Final remarks

In summary, this chapter presents the first insight into the *MAT* locus structure of Sporidiobolales (red yeasts), which was so far unexplored in this respect. While the basis of the work consisted of the identification of novel genomic regions in *Rhodosporidium toruloides*, the syntenic genomic regions of two additional red yeast species were also analyzed. Results strongly suggest that the studied group hold extended multigenic *MAT* loci, containing both mating-related and unrelated genes and constitutes the second instance of the presence of a PAK kinase gene within a basidiomycete *MAT* locus, therefore resembling the *MAT* locus of the distantly related bipolar *Cryptococcus* species. However, the apparent high level of synteny observed between the three red yeast species is in sharp contrast with the highly variable *MAT* gene order and orientation observed in the *Cryptococcus* species.

4. CHAPTER

The pseudo-bipolar *MAT* system of the red yeast *Sporidiobolus salmonicolor*

The work presented in this chapter was published in:

Coelho M. A., Sampaio J. P., and Gonçalves, P. 2010. A deviation from the bipolar-tetrapolar mating paradigm in an early diverged basidiomycete. PLoS Genet. 6:e1001052.

4.1. Introduction

Classical mating studies indicate a predominance of bipolar systems in the subphylum Pucciniomycotina both for rusts (Buller, 1950), other plant parasites like *Microbotryum violaceum* (Giraud et al., 2008) and for saprobic red yeasts of the genera *Rhodosporidium* and *Sporidiobolus* (Sampaio, 2011a; Sampaio, 2011c). However, detailed molecular analyses elucidating *MAT* gene content and function were very limited except for *M. violaceum*, in which pheromone receptor genes of both mating types were recently identified (Devier et al., 2009). Since molecular phylogenies and ultrastructure suggest that Pucciniomycotina diverged first among subphyla of Basidiomycota (Stajich et al., 2009), the elucidation of the genetic structure of the mating systems in this lineage is of major importance for understanding the origin and evolution of basidiomycete *MAT* loci. The characterization of the *MAT A1* pheromone receptor regions in several bipolar red yeast species, including the heterothallic red yeast *Sporidiobolus salmonicolor* (Chapter 3) contributed new information concerning this. However, it remained to be clarified if the homeodomain transcription factors are determining sexual identity and/or if they are part of the *MAT* locus in red yeasts.

This chapter is focused on the molecular characterization of the mating system in *S. salmonicolor*. The starting point of this work consisted in the identification, for the first time in the Pucciniomycotina, of the divergently transcribed genes encoding the HD1 and HD2 transcription factors which, in turn, allowed investigating how HD1/HD2 transcription factors and the pheromone receptor system interact to produce the bipolar mating behaviour observed in *S. salmonicolor*. Several lines of evidence presented here indicate that a new, non-bipolar non-tetrapolar, sexual mechanism exists in this species, in which occasional disruptions of the genetic cohesion of the bipolar *MAT* locus originate new mating types. A new model for the evolution of basidiomycete *MAT* loci is proposed to accommodate the novel findings.

4.2. Material and methods

4.2.1. Strains and mating type designations

The list of strains studied and relevant information associated to them is given in Table 4.1. The mating type designations of *S. salmonicolor* were re-assigned in order to match the "molecular mating type" identified by PCR detection of the pheromone receptor alleles (*STE3.A1* or *STE3.A2*).

4.2.2. Common settings for PCR amplification

Preparative PCR reactions were performed in a final volume of 50 μl with the following components (unless stated otherwise): 2 mM of $MgCl_2$, 0.20 mM of each of the four deoxynucleoside triphosphates (GE Healthcare), 1% DMSO, 0.8 μM of each primer, 100 ng of genomic DNA, and 1 U Taq DNA polymerase (Fermentas, Canada). Thermal cycling consisted of a 5-minute denaturation step at 95ºC, followed by 35 cycles of denaturation at 94ºC for 30 s, 30 s at the annealing temperature (variable), and extension at 72ºC (variable time). For annealing temperatures, extension times and primer sequences see Table 4.2. A final extension of 7 min at 72ºC was performed at the end of each reaction.

4.2.3. PCR amplification and sequencing of the *MAT A2* pheromone receptor

The pheromone receptor gene (pr-MatA2) from *Microbotryum violaceum* (GenBank: EF584741) was used to perform a Blastn search in the NCBI Trace Archive sequences of the genome project of *Rhodosporidium babjevae* strain WP1 (*MAT A2*). The complete *STE3.A2* sequence of *R. babjevae*, in addition to the flanking regions, were assembled (Trace sequences are given in section II.1, Appendix II) and subsequently scrutinized for the presence of genes exhibiting a higher degree of conservation across species than pheromone receptor genes. Sequences of two putative genes encoding an LSm-like protein (*LSm7*) and a ribosomal protein L6 (*RibL6*) were used to design a pair of degenerate primers (MC122 and MC123) to amplify and sequence the intervening region in *S. salmonicolor MAT A2* strain CBS 490^T. This region contained the *STE3.A2 S. salmonicolor* gene, which was sequenced by primer walking using primers MC126 and MC127 (Figure 4.1).

Table 4.1. Strains used in Chapter 4 and relevant information pertaining to them. *MAT A1* and *MAT A2* strains are indicated by 'A1' and 'A2' respectively; 'SF' indicates self-fertile strains and 'AS' asexual strains. The mating type designations (A1 or A2) were re-assigned to match the "molecular mating type" identified by PCR detection of the pheromone receptor alleles (*STE3.A1* or *STE3.A2*).

Sporidiobolus salmonicolor				
Strain	**Sexuality** (mating tests)	***STE3* alleles**	***HD1/HD2* alleles**	**Source**
CBS 1012	A2	*STE3.A2*	A2–16	Leaf of *Aristolochia* sp., hot-house of Botanic Laboratory, Delft, Netherlands
CBS 1013	A2	*STE3.A2*	A2–14	Air in dairy, USA
CBS 1039	A2	*STE3.A2*	A2–14	Air in dairy, USA
CBS 2635	A2	*STE3.A2*	A2–13	Infected skin of a man, Germany
CBS 2873	A2	*STE3.A2*	A2–15	Extract of oak bark, France
CBS 4030	A2	*STE3.A2*	A2–10	Soil, New Zealand
CBS 487	A2	*STE3.A2*	A2–10	Air, Japan (type strain of *Pseudomonilia rubiduncula* Okunuki)
CBS 490T	A2	*STE3.A2*	A2–10	Culture contaminant (type strain of *Sporidiobolus salmonicolor* Fell & Statzell Tallman)
CBS 495	A2	*STE3.A2*	A2–13	Culture contaminant, Japan
CBS 496	A2	*STE3.A2*	A2–10	Scorched skin of orange, Delft, Netherlands
CBS 5937	A2	*STE3.A2*	A2–14	Carious dentine, South Africa (type strain of *Aessosporon salmonicolor* van der Walt)
CBS 6322	A2	*STE3.A2*	A2–12	Rotting wood shavings, Chile
CBS 6470	A2	*STE3.A2*	A2–15	Air, Germany
CBS 6530	A2	*STE3.A2*	A2–15	Bog, locality unknown
CBS 6781	A2	*STE3.A2*	A2–15	Bituminous soil, oil field near Cumanà, Venezuela (type strain of *Sporidiobolus veronae* Balloni)
CBS 6832	A2	*STE3.A2*	A2–15	Cerebro-spinal fluid, India (type of *Sporobolomyces salmonicolor* Kluyver & van Niel var. fischeri Misra & Randhawa)
CBS 7260	A2	*STE3.A2*	A2–13	Slime flux of *Quercus pendunculata*, Russia (type strain of *Sporobolomyces philippovii* Krasil'nikov)
PYCC 5245	A2	*STE3.A2*	A2–13	Unknown substrate, Portugal
RJB 8219	A2	*STE3.A2*	A2–15	Isolated by R. J. Bandoni
RJB 950	A2	*STE3.A2*	A2–11	Isolated by R. J. Bandoni (deposited at CBS as *Sporobolomyces odorus* Derx)
ZP 392	A2	*STE3.A2*	A2–10	Fruit body of *Dacrymyces* sp., Germany
CBS 2647	A1	*STE3.A1*	A1–2	Gut of *Drosophila* sp., locality unknown
CBS 2648	A1	*STE3.A1*	A1–5	Exudate of *Kalopanax vicinifolium* var. *typicum*, isolated by K. Tubaki, Japan
CBS 4474	A1	*STE3.A1*	A1–5	Blistered skin of a man, Bonn, Germany
CBS 483	A1	*STE3.A1*	A1–2	Rusted leaf of *Citrus* sp., France (type strain of *Sporobolomyces odorus* Derx)
CBS 497	A1	*STE3.A1*	A1–4	Unknown (deposited at CBS as *Rhodomyces kochii*)
CBS 6529	A1	*STE3.A1*	A1–3	Culture contaminant, R.J. Bandoni
IHMT 2446/96	A1	*STE3.A1*	A1–6	Skin lesion, Portugal
NRRL Y-17498	A1	*STE3.A1*	A1–3	Ice core, Greenland

Table 4.1. Continued.

Strain	Sexuality (mating tests)	*STE3* alleles	*HD1/HD2* alleles	Source
ML 2241	A1	*STE3.A1*	A1–3	Isolated by J. W. Fell
PYCC 4558	A1	*STE3.A1*	A1–1	Polluted river water, Oeiras, Portugal
PYCC 4623	A1	*STE3.A1*	A1–3	Soil, USSR
RJB 948	A1	*STE3.A1*	A1–3	Isolated by R. J. Bandoni (deposited at CBS as *Sporobolomyces odorus* Derx)
ZP 648	A1	*STE3.A1*	A1–4	Leaf of *Ligustrum* sp., Monte de Caparica, Portugal
CBS 2641	AS	*STE3.A2*	A2–10	Air, Netherlands
CBS 4029	AS	*STE3.A2*	A2–15	Soil, New Zealand
Sporidiobolus johnsonii				
CBS 2634	A2	*STE3.A2*	A2–17	*Fragaria sp*., Japan (deposited at CBS as *Sporidiobolus johnsonii* Nyland)
CBS 2630	A1	*STE3.A1*	A1–9	Air, The Netherlands
CBS 2643	A1	*STE3.A1*	A1–9	Isolated by S. Windisch, Germany
CBS 4209	A1	*STE3.A1*	A1–8	Fruit body of *Exidia* sp., Japan (type strain of *Sporobolomyces coralliformis* Tubaki)
PYCC 4351	A1	*STE3.A1*	A1–9	Leaf of a tree, Portugal
CBS 1522	AS	*STE3.A1*	A1–9	Fodder yeast, Germany (type strain of *Sporobolomyces holsaticus* Windish)
CBS 7795	AS	*STE3.A1*	A1–7	Culture contaminant, isolated by W. I. Golubev, Russia
CBS 5470^{T}	SF	*STE3.A1* *STE3.A2*	A1/A2–18	Leaf of *Rubus idaeus* with dead pustule of *Phragmidium rubi-idaei*, USA (type strain of *Sporidiobolus johnsonii* Nyland)
CBS 8241	SF	*STE3.A1* *STE3.A2*	A1/A2–18	Unknown

4.2.4. PCR amplification and sequencing of the HD1/HD2 region

Using the sequence of the *HD1* gene of *Sporobolomyces roseus* (Coelho et al., 2008; Chapter 3), a Blastn search was performed in the NCBI Trace Archive database of *S. salmonicolor* and positive hits (listed in section II.1, Appendix II) were assembled into a complete *HD1* gene. The upstream flanking region of the *HD1* was assembled stepwise and inspected using AUGUSTUS software (Stanke et al., 2008), revealing the presence of a divergently transcribed *HD2* homologue. The deduced protein sequences of *HD1* and *HD2* of *Sp. roseus* and *S. salmonicolor* were aligned and the conserved regions were used to design specific primers (MC103 and MC104) to amplify and sequence the corresponding N-terminal and intergenic regions of the *HD1/HD2* genes in all *S. salmonicolor* and *S. johnsonii* strains (Figure 4.1, Table 4.2).

Table 4.2. List of primers and specific PCR conditions used in Chapter 4. Degenerate nucleotides follow the IUPAC nomenclature. Restriction enzymes used for the RFLP analysis are indicated. *'Ss'* and *'Sj'* strands for *S. salmonicolor* and *S. johnsonii*, respectively. All primers were obtained from MWG Biotech AG.

Primer Name	**Sequence** (5' – 3')	**Annealing / Extension**	**Remarks**	**Enzyme** (RFLPs)
MC122	TGGCGMGTYCCCGAACCYTACAAC	62.5ºC / 3 min	Amplify the complete *STE3.A2* allele in the type strain of *Ss* (CBS 490^T)	
MC123	AAAGGSTACGACCAGYTSCTCAAYCT			
MC126	AATRCCGCTCCGCTYTACTGGCA	62ºC / 90 s		
MC127	TGTTCGGGTCTACAGGTTGGAGAG			
MC103	AGGATGTCGAGGACTTCTTCGGTGA	62 ºC / 90 s	Amplify the HD1/HD2 region in strains of *Ss* and *Sj*	RsaI
MC104	CGGCGCGTGTTGGTGAACCAGGT			
MC053	CATCTTCTTCCTCCTCCTGCC	60ºC / 90 s	Screening the *STE3.A1* and *STE3.A2* alleles in strains of *Ss* and *Sj*	
MC054	GGGTAGGGAGGAAGCGAGAT			
MC126	AATRCCGCTCCGCTYTACTGGCA	62ºC / 90 s		
MC127	TGTTCGGGTCTACAGGTTGGAGAG			
MC111	CGCCGCTGGTTCCTTGRCAACTTCT	62ºC / 90 s	Amplify the C-terminal region of the *HD1* gene in selected *Ss* and *Sj* strains	
MC112	GCTGCCCATCTACGAGCAAGTCCC			
MC147	GARGACCTCGCSACCTACCC	60ºC / 30 s	Amplify part of the *PAN6* gene in *Ss*	AluI
MC148	AGGAKGAGMGCTTGYTGGATGT			
MC045	CGMGAAGACCTGAARCCGTT	56ºC / 90 s	Amplify part of the *STE20* gene in *Ss*	EcoRI
MC046	CATGWACTCCATSACGACCCA			
MC131	CGAYATCRCCRAAACAGCTGAGGAT	59ºC / 90 s	Amplify part of the *KAP95* gene in *Ss*	
MC132	TACAACGARCTCCARGTCAACATCT			
MC133	ACCTCAGCAAGTACGTYGACAAGAA	59ºC / 90 s	Amplify part of the *LSm7* gene in *Ss*	RsaI
MC134	AAAGGGTTTGCAATTTCCKCMGAC			
MC135	GAGAAGGAGCCKCARCCCAA	56ºC / 90 s	Amplify part of the *RibL18* gene in *Ss*	KpnI
MC136	ACATCAAGCAGCTCCTCACCC			
MC137	GACTTCCCRGGYCACTAYCCC	59ºC / 90 s	Amplify part of the *RNAPOL* gene in *Ss*	
MC138	CCTGTCCKCCKCGCATCTT			
MC141	CTACATGTTCAAGTACGACTCGAC	59ºC / 90 s	Amplify part of the *GPD* gene in *Ss*	StuI
MC142	GTACCACGAGAAGAGCTTGACGAA			
MC139	TCGATCCTCATGGAGCAYTTCATCT	59ºC / 90 s	Amplify part of the *MIP* gene in *Ss*	
MC140	CCAYTTGCCMACYTCTTCCATC			
MC157	GTCTTGACGACCGGGTTCAACGA	59ºC / 90 s	Amplify part of the *IsocL* gene in *Ss*	BamHI
MC158	GGCAWTCCKCCTCGGTAGTGGTA			
MC155	TCAACMTCCTCTCYGGCAAGTC	59ºC / 90 s	Amplify part of the *NGP1* gene in *Ss*	
MC156	GAGTTRTTGATGGAGGCGTGGAA			
MC149	CTGGACARGAGACGTACAATGC	59ºC / 90 s	Amplify part of the *AKOR2* gene in *Ss*	RsaI
MC150	TCTCGAGCRTGAARTCAAAGACGTT			
MC151	CTGAACCGMTACACCTTTGCC	59ºC / 1 min	Amplify part of the *RPB2* gene in *Ss*	
MC152	TGAATCTCGCAATGWGTCCAAG			
MC170	TATCTACGGCGCTGGTTGCTTGA	60 ºC / 90 s	Amplify part of the *sdhA* gene in *Ss*	
MC171	TCCAGATCATCGAGCGGTCCTTG			

Table 4.2. Continued.

Primer Name	Sequence (5' – 3')	Annealing / Extension	Remarks	Enzyme (RFLPs)
MC047	ACCAACCTYTGCGTYTCRGTCGA	63 ºC / 90 s	Amplify part of the *URA3* gene in *Ss*	
MC048	GATGACCTCGTCSGGYGTACKGTA			
MC172	GCCAGTCCCTCATCACCTCG	60ºC / 90 s	Amplify part of the *HXT1* gene in *Ss*	
MC173	TCCTTGAGCCTGAAGCAGTCG			
MC174	GACCAGCAGTTCTTGGCCGA	60ºC / 90 s	Amplify part of the *aldA* gene in *Ss*	
MC175	GAGCTTGACAAACTCGTCGTGAAC			
MC176	AGCGTTGCGTTCCTCCAGTC	60ºC / 90 s	Amplify part of the *PAL* gene in *Ss*	
MC177	GTTCTTGGCGACCTCGATCTG			
MC178	CTCATCATCGTCGACTCTGTCAC	60ºC / 60 s	Amplify part of the *DMC1* gene in *Ss*	Msp I
MC179	GGCGAGTCGACGATCTTGG			
MC180	GCCGTCTACCGAATCAACAAG	57ºC / 90 s	Amplify part of the *GEF1* gene in *Ss*	
MC181	CTGCATCTGGAGCTTGCATTT			
MC182	CCCATCCTGAATGTCAACTACAA	58ºC / 90 s	Amplify part of the *LACC* gene in *Ss*	
MC183	GGTGACAAGAACCGAGTAGCG			
MC186	CAAGTGCCTGTGCAAGAA	51ºC / 60 s	Amplify part of the *RAN1* gene in *Ss*	Hinf I
MC187	CATAAGCCATGTACGGTTT			

4.2.5. Correlation between mating behaviour and the presence of pheromone receptor genes

To confirm mating behavior and sexual compatibility, 2–4 day-old cultures were crossed on corn meal agar (Difco), incubated at room temperature for 1 week, and examined microscopically using phase-contrast optics for production of mycelium with clamp connections and teliospores. Diagnostic PCR with primers for *STE3.A1* (MC053 and MC054) and *STE3.A2* (MC126 and MC127) were carried out to identify the pheromone receptor genes.

4.2.6. Germination of teliospores and microscopic monitoring of meiosis

S. salmonicolor strains CBS 6832 (*MAT* A2–6) and ML 2241 (*MAT* A1–3) were mixed on corn meal agar (Difco) and incubated at room temperature (~22ºC) for 2 weeks, to allow for abundant production of teliospores. Small (~0.5 cm) agar blocks containing teliospores were soaked in sterile distilled water for 8–10 weeks at 4ºC. After this resting period, a suspension of teliospores was obtained by gently mashing the agar blocks with a small pestle. Teliospore germination was induced by transferring this suspension to 2% (w/v) low melting point (LMP)-agarose plates incubated at

room temperature for 2 weeks. To follow the germination of selected teliospores, small drops of the same suspension were transferred to 2% (w/v) LMP-agarose-coated-slides. The different stages of meiosis were observed microscopically by staining several germinating teliospores with Safranin O (Dhingra and Sinclair, 1995). Observations were made daily with 100× magnification using a Leica DMR microscope equipped with brightfield and differential interference contrast optics and microphotographs were recorded using a Leica DFC320 digital camera (Leica Microsystems GmbH, Wetzlar).

4.2.7. Micromanipulation of teliospores and segregation analysis

Teliospores resulting from the cross of strains CBS 6832 (*MAT* A2–6) and ML 2241 (*MAT* A1–3) were germinated as described above. Using a micromanipulator, teliospores in the initial stages of germination, i.e. with a non-septate basidium initial, were individually separated, transferred to MYP agar (Valério et al., 2008) and incubated at room temperature. The colony that formed after 1–2 days was re-streaked to obtain colonies derived from single cells. In three to six of these colonies, segregation of the two *MAT* regions was assessed by diagnostic PCR with specific primers for *STE3.A1* (MC053 and MC054), *STE3.A2* (MC126 and MC127) and *HD1/HD2* (MC103 and MC104). The *HD1/HD2* alleles were discriminated after amplification by digestion with restriction enzyme RsaI.

4.2.8. Sequence data and phylogenetic analyses

DNA and protein sequences were aligned with ClustalW 1.81 with minor manual corrections. For the phylogeny of *HD1/HD2* alleles, FindModel [a web implementation of ModelTest (Posada and Crandall, 1998)], with the Akaike information criterion (AIC) was used. The Tamura-Nei (TrN)+G model (Tamura and Nei, 1993; Yang, 1994) (shape parameters 0.98049 and 1.08652 in Figures 4.2 and 4.3 datasets, respectively) was employed in MEGA4 (Tamura et al., 2007) using the Neighbor-Joining algorithm (Saitou and Nei, 1987) and bootstrap values from 1000 replicates. For the phylogeny of the pheromone receptors (Figure 4.4) the substitution model of protein evolution was selected using ProtTest (Abascal et al., 2005) with AIC. The WAG+I+G (Reeves, 1992; Yang, 1994; Whelan and Goldman, 2001) model (shape parameter =2.99; proportion of invariable sites = 0.035) was employed in PhyML (Guindon and Gascuel, 2003) using maximum likelihood and bootstrap values from 100 replicates.

4.2.9. Estimation of the evolution rates of the *HD1* gene

To obtain the sequence of the 3' end of *HD1* gene the region encompassing the homeodomain and a conserved motif located 10 bp upstream of the STOP codon was amplified using primers MC111 and MC112. The 5' end of the *HD1* gene was amplified using primers MC103 and MC104 (Figure 4.1). Evolution rates were estimated by a Window Analysis of dN and dS, using the online interface of WINA 0.34 (Endo et al., 1996), in a sliding window (size = 20) along the alignment of nine *HD1* alleles.

4.2.10. *MAT* specificity and phylogenetic analysis of genes located at variable distances from the *HD1/HD2* and the pheromone receptor regions

Based on the available genomic information of the closely related species *Sporobolomyces roseus*, and assuming that synteny is maintained in *S. salmonicolor*, eight genes from the pheromone receptor region (scaffold 9) and four genes from the HD1/HD2 region (scaffold 7) were selected. NCBI *S. salmonicolor* Trace Archive sequences of these genes were obtained (section II.2, Appendix II) and assembled. Primers were designed for partial sequencing of these genes (Table 4.2). Protein-coding DNA sequences were deduced after removal of putative introns, either manually or using AUGUSTUS software, automatically aligned using ClustalW and manually edited according to the superimposed amino acid sequences. Synonymous substitutions (dS) values and the divergence percentage between mating type-specific alleles were calculated using DnaSP 5.0 (Librado and Rozas, 2009) for a set of 10 *S. salmonicolor* strains (Figures 4.11 and 4.12). Phylogenies were obtained in MEGA4 using Maximum Parsimony (MP), with the close-neighbor interchange algorithm for heuristic searches and bootstrap values from 1000 replicates.

4.2.11. Genotyping of the meiotic progeny

To detect differences among the meiotic progeny, nine genes located on four different scaffolds in the *Sp. roseus* genome (scaffolds 1, 2, 3 and 10; Figure 4.10) were selected. Sequences of these genes in *S. salmonicolor* were assembled from the NCBI Trace Archive (section II.3, Appendix II) and primers were designed to amplify partial sequences in parental strains CBS 6832 and ML 2241 (Table 4.2). Sequence polymorphisms were scored and used to track the parental origin of the alleles in all the meiotic progeny. This analysis was performed either by PCR and sequencing or by RFLP analysis (Table 4.2). The genes previously used to obtain mating type specific phylogenies were similarly assessed in this genotyping analysis.

4.3. Results and discussion

4.3.1. Identification of *MAT* genes in both mating types of *S. salmonicolor*

The *MAT A1* pheromone receptor and pheromone precursor genes were previously identified in the heterothallic bipolar red yeasts *Rhodosporidium toruloides* and *Sporidiobolus salmonicolor* and in *Sporobolomyces roseus* (Coelho et al., 2008) (see Chapter 3). These studies failed to uncover *HD1/HD2* transcription factor genes in the vicinity of the pheromone receptor in any of the species studied. However, an *HD1* homologue was found in the complete genome sequence of *Sp. roseus*, but at a considerable distance from the pheromone receptor gene (Coelho et al., 2008). Although no putative *HD2* homologues were identified at that time, re-analysis of the region flanking the *HD1* gene in *Sp. roseus* genome revealed that a divergently transcribed *HD2* homologue was indeed present at that genomic location. Its misidentification in the previous study was due to the presence of an intron in the short conserved region corresponding to the homeodomain motif. This new data was also confirmed in a recent revision paper by Morrow and Fraser (2009). To get more insight in the genetic structure of *MAT* loci in red yeasts, a set of 36 *S. salmonicolor* strains were chosen for characterization of their *MAT* genotype. This species was selected because of previous studies concerning its mating behavior (Valério et al., 2008) and of the availability of a 3× genome coverage in the form of Trace Archive sequences for one *MAT A1* strain, considerably facilitating the identification of novel genes.

First, it was needed to identify the pheromone receptor gene in the complementary mating type of *S. salmonicolor* (*MAT A2*). Degenerate primers based on the sequences available for homologous *MAT A2* receptors, including that of the closely related *R. babjevae* for which Trace Archive genomic sequences are available, failed to amplify the *S. salmonicolor* gene. An alternative approach was conceived, based on the previous observation of a high degree of synteny between the same mating type in different red yeast species (Coelho et al., 2008). For this, the region surrounding the pheromone receptor gene in *R. babjevae* was assembled stepwise using Trace Archive sequences and was subsequently scrutinized for the presence of genes exhibiting a higher degree of conservation across species than pheromone receptor genes. Two putative genes encoding an LSm-like protein (*LSm7*) and a ribosomal protein L6 (*RibL6*) were found upstream and downstream of *STE3.A2* in *R. babjevae*, respectively. Sequences homologous to these genes were readily found by Blastn search in *S. salmonicolor* Trace Archives but they were differently organized, since the sequenced strain (CBS 483) belongs to the opposite mating-type (*MAT A1*) (Figure 4.1). PCR primers based on the *S. salmonicolor LSm7* and *RibL6* genes allowed the amplification of the intervening *S. salmonicolor STE3.A2* gene (Figure 4.1) (GenBank: GU474641), attesting synteny conservation between *S. salmonicolor* and *R. babjevae* in this region within the same mating type. As expected, the predicted amino acid sequence of the *S. salmonicolor* receptor exhibited the highest identity with homologous sequences of *R. babjevae* (54%) and *M. violaceum* (47%).

The divergently transcribed genes encoding the HD1 and HD2 transcription factors were identified in the Trace Archive sequences of *S. salmonicolor* using the homologous sequences of *Sp. roseus*. Primers based on the homeodomain conserved region of these genes amplified a fragment encompassing the highly variable 5' regions of both genes as well as the intergenic region between the *HD1* and *HD2* genes (Figure 4.1). Sequence identity of the HD1 and HD2 proteins from *Sp. roseus* was, as expected, limited to the homeodomain region since all functional HD proteins characterized so far were found to have highly variable N-terminal domains, even when different *HD1/HD2* alleles from the same species are compared.

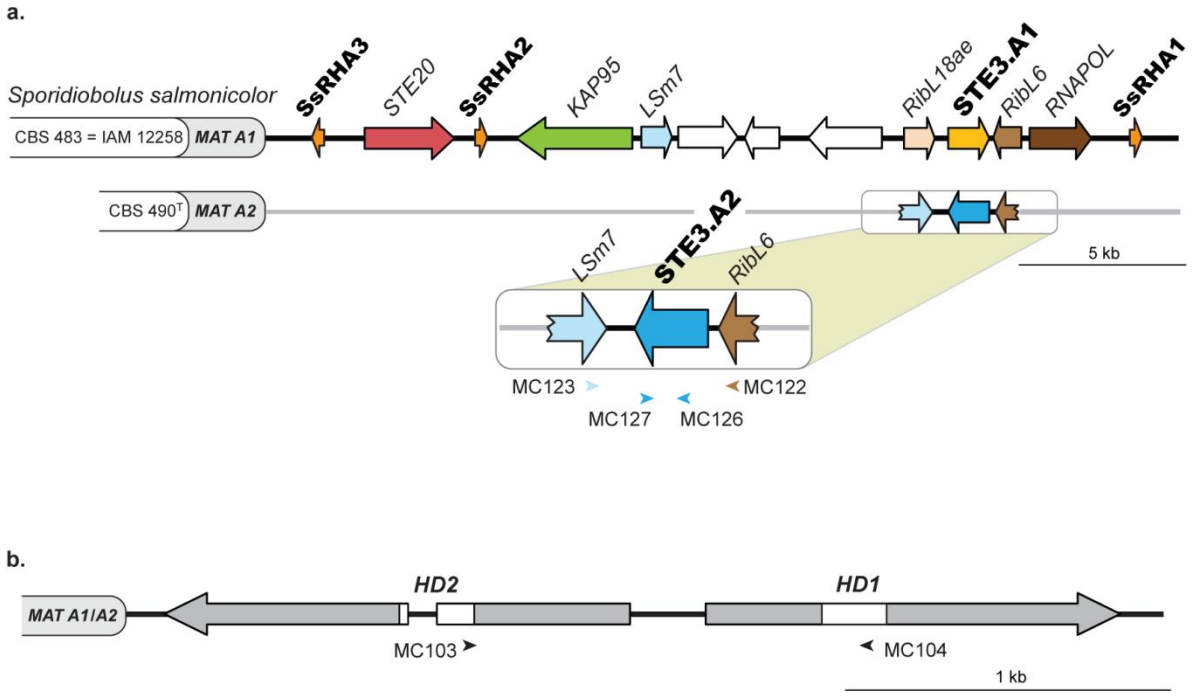

Figure 4.1. Organization of the genomic regions containing the pheromone receptor (*STE3*) and the *HD1/HD2* genes in *S. salmonicolor*. (a) Pheromone receptor region. Homologous genes are represented with the same colour in both mating types. The pheromone precursor genes (*SsRHA1*, *SsRHA2* and *SsRHA3*) are shown only in *MAT A1* because their position in *MAT A2* strains is currently unknown. Primers used to obtain the complete sequence of the *S. salmonicolor STE3.A2* gene are depicted by arrowheads. Synteny with the homologous genomic region of *Sp. roseus* was confirmed for all genes except those shown as empty arrows **(b)** The *HD1/HD2* region comprises two divergently transcribed homeodomain genes. The homeodomain motifs are indicated as white boxes and an intron present in the *HD2* gene is depicted. Black arrowheads indicate the position of the primers used to amplify the *HD1/HD2* region.

4.3.2. Allele number and mating type specificity of *HD1/HD2* and pheromone receptor genes in *S. salmonicolor*

Heterothallic red yeast species within the order Sporidiobolales were all described as having a bipolar behaviour in standard mating tests (Sampaio, 2011a; Sampaio, 2011c). Therefore, it is remarkable that in *Sp. roseus*, the distance between the pheromone receptor and HD1/HD2 regions, although still not accurately determined, was in any case larger than 600 kb (Coelho et al., 2008; Morrow and Fraser, 2009). Hence, if both classes of genes were part of a bipolar *MAT* locus with suppression of recombination over its entire length, it would be the largest such locus characterized so far in basidiomycetes. Such a region would nevertheless functionally resemble the *MAT* loci of *C. neoformans* (Lengeler et al., 2002; Fraser et al., 2004) and *U. hordei* (Lee et al., 1999; Bakkeren et al., 2006). Alternatively, one of the compatibility check points might no longer be required to determine sexual identity, as previously observed for two mushroom species (James et al., 2006; Yi et al., 2009). In light of this, it was important to investigate how HD1/HD2 transcription factors and the pheromone receptor system interacted to produce the bipolar mating behavior. Isolation of both classes of *MAT* genes in *S. salmonicolor* enabled the examination of this question for the first time. Specific PCR primers were used to assess the correlation between the presence of the alternate pheromone receptor genes (*STE3.A1* and *STE.A2*) and mating behavior in 36 natural isolates of *S. salmonicolor* (Table 4.1). Without exception, *STE3.A1* was present in *MAT A1* strains, whereas the *STE3.A2* receptor gene was found in *MAT A2* strains (Table 4.1). Hence, the pheromone receptor was clearly associated to mating behaviour. Next, PCR fragments encoding the 5' portions of the *HD1* and *HD2* genes were obtained from the 36 *S. salmonicolor* strains of both mating types and sequenced (GenBank: GU474649–GU474693). Surprisingly, instead of two *HD1/HD2* alleles, each strictly linked to one of the mating types, as might be expected for a bipolar species, 13 *HD1/HD2* alleles exhibiting substantial sequence divergence were uncovered (Figure 4.2). Such numbers of *HD1/HD2* alleles were previously observed only for tetrapolar species (Raper, 1966). In *U. maydis*, for example, 33 *HD1/HD2* alleles were identified (Fraser et al., 2007) which do not have a particular association to either one of the two pheromone receptors because the two *MAT* regions are located on different chromosomes (Bakkeren and Kronstad, 1994) and thus segregate independently at meiosis. On the contrary, in the bipolar species characterized so far, only two *HD1/HD2* alleles are present, each of which is linked to one of the two pheromone receptors forming two large bipolar loci (Lee et al., 1999; Lengeler et al., 2002). Recombination between the two genetically linked *MAT* regions is suppressed in these species as a result of extensive sequence divergence, gene inversions and the accumulation of repetitive elements reminiscent of sexual chromosomes in animals and some plants (Lee et al., 1999; Lengeler et al., 2002; Fraser and Heitman, 2004; Fraser and Heitman, 2005). In *S. salmonicolor*, each of the 13 alleles always appears associated with the same receptor, suggesting some form of genetic linkage between the two regions (Figure 4.2). However, each receptor is associated with seven (Ste3.A2) or six (Ste3.A1) different *HD1/HD2*

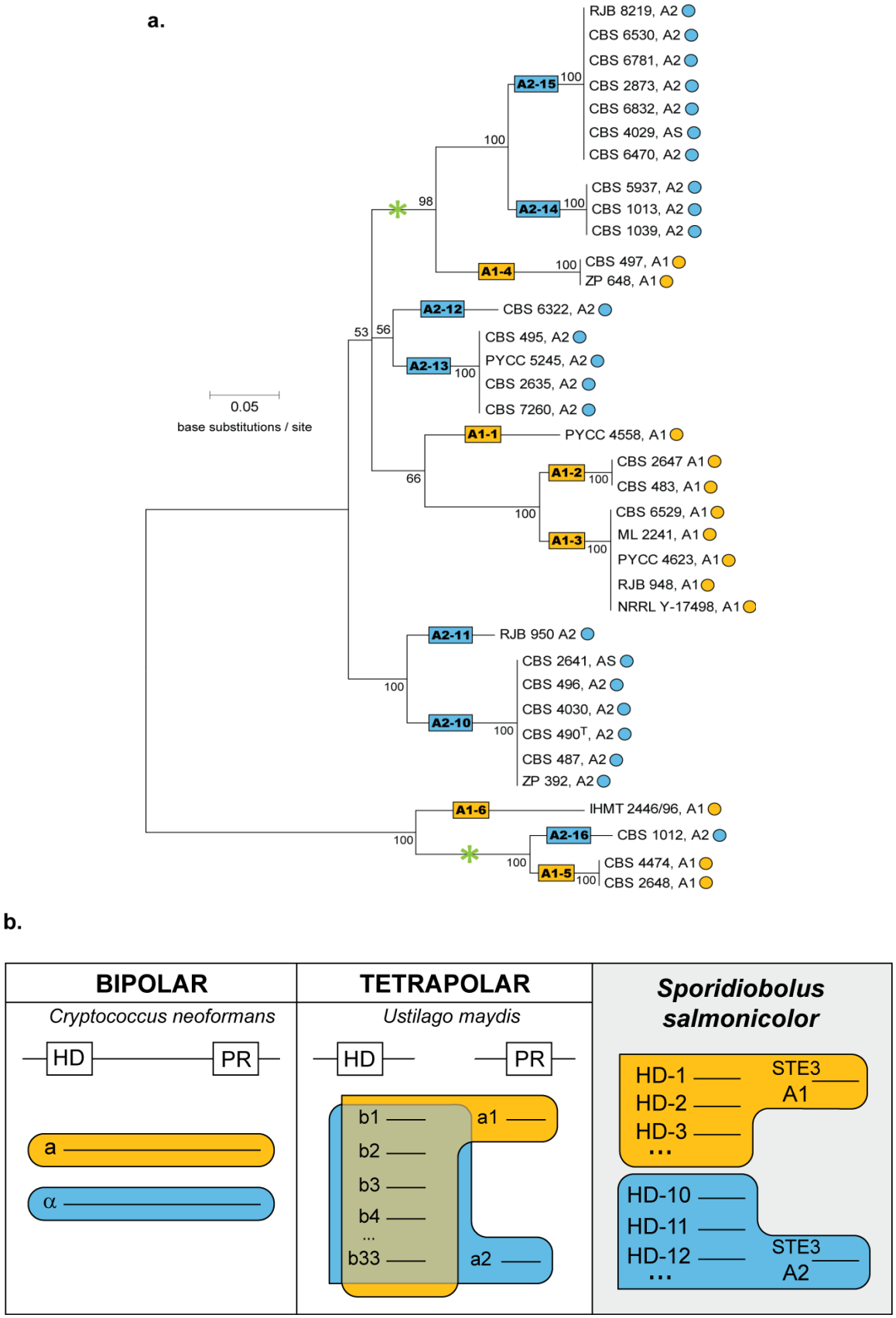

Figure 4.2. Diversity and phylogeny of *MAT* gene alleles in *S. salmonicolor*. See next page for description.

Figure 4.2. Diversity and phylogeny of *MAT* gene alleles in *S. salmonicolor*. (a) Phylogeny of *HD1/HD2* genes inferred by Neighbour-joining and using the Tamura-Nei (TrN)+G model (shape parameter = 0.98049). Mating types A1, A2, and asexual strains are designated as A1, A2 and AS, respectively. Boxes designate the various *HD1/HD2* alleles with numerals after the mating type designation. Circles after strain numbers depict the type of pheromone receptor gene (yellow, *STE3.A1*; blue, *STE3.A2*). Asterisks indicate the two instances where common ancestry of *HD1/HD2* alleles associated with opposite mating types is best supported. **(b)** Comparison of the number and distribution of *MAT* gene alleles in emblematic species representing the different mating systems (HD, homeodomain region; PR, pheromone receptor region; the continuous or discontinuous line between the HD and PR boxes denotes genetic linkage or its absence, respectively).

alleles which is in sharp contrast with other bipolar species (Figure 4.2). There is apparently no bias regarding the geographic distribution of *HD1/HD2* alleles, since strains carrying the same *HD1/HD2* allele were isolated from very diverse locations worldwide (Figure 4.2 and Table 4.1). To shed some light on the phylogenetic relationship between the *HD1/HD2* alleles, the alleles present in a very closely related species, *S. johnsonii*, were also examined. Several lines of evidence suggest that *S. salmonicolor* and *S. johnsonii* are not fully reproductively isolated and pre-zygotic barriers are absent, since crosses between sexually compatible strains of the two species normally yield dikaryotic mycelium and teliospores (Valério et al., 2008). Intriguingly, extant allele diversity seems to have been generated after the onset of this incipient speciation event, as *HD1/HD2* alleles from the two species are, in general, phylogenetically distinct (Figure 4.3). However, some incongruities between rDNA and *MAT* gene phylogenies can be observed (Figure 4.3), indicative of a certain degree of gene flow between the two species. For example, *S. salmonicolor* strain CBS 2634 has a *S. johnsonii* rather than a *S. salmonicolor* receptor allele and, in addition, a group of five *S. salmonicolor* strains carry *HD1/HD2* alleles (A1–5, A2–16, A1–6) (Figure 4.3) which seem to be phylogenetically more related to the *S. johnsonii* clade. The intraspecific common ancestry of *HD1/HD2* alleles contrasts with the situation observed for the pheromone receptor genes. The latter genes were found to exhibit a clear and ancient trans-specific polymorphism (Figure 4.4), as previously reported for *M. violaceum* (Devier et al., 2009).

Hence, the relationship between mating type and *HD1/HD2* allele number and distribution is unexpectedly complex in *S. salmonicolor*, and the possibility that the *HD1/HD2* transcription factors might have recently lost their linkage to *MAT* and are no longer involved in determining sexual compatibility in this species could not be readily discarded. To examine this, the evolution rates (dN/dS) of nine *HD1* alleles of *S. salmonicolor* and *S. johnsonii* were firstly determined (GenBank: GU474694–GU474702) considering that prior studies in the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*) brought to light exceptionally high evolution rates for the domains in HD1 proteins involved in self/non-self recognition (Badrane and May, 1999). Surprisingly, in *S. salmonicolor* even higher evolution rates for the homologous domains in the HD1 proteins were found (>1 in some segments, indicative of adaptive selection) (Figure 4.5), which constituted the first evidence that these transcription factors contribute to determine sexual compatibility together with the pheromone receptor system.

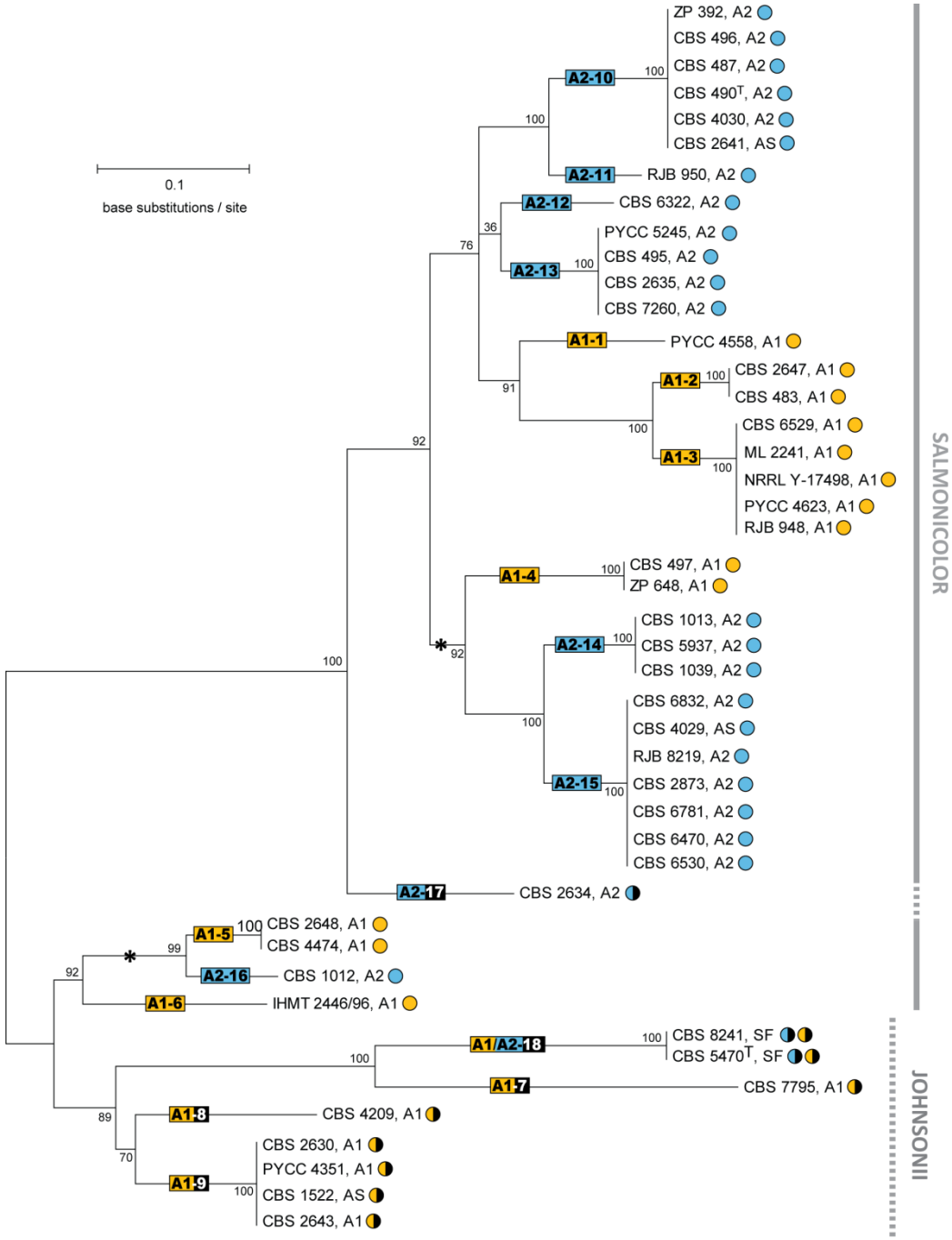

Figure 4.3. Phylogeny of *HD1/HD2* alleles in the species complex *S. salmonicolor*/*S. johnsonii*. Tree inferred by Neighbour-joining and using the Tamura-Nei (TrN)+G model (shape parameter = 1.08652). Mating types A1, A2, and asexual strains are designated as A1, A2 and AS, respectively. Boxes designate the various *HD1/HD2* alleles with numerals after mating type designation. Yellow boxes correspond to *HD1/HD2* alleles associated to the pheromone receptor gene *STE3.A1* and blue boxes correspond to those associated with *STE3.A2*. Circles depict the type of pheromone receptor gene (yellow, *STE3.A1*; blue, *STE3.A2*). Half-coloured circles or boxes are used for *S. johnsonii* alleles. Lateral bars delimit *S. salmonicolor* (solid bar) and *S. johnsonii* (dashed bar) as defined by rDNA phylogeny (Valério et al., 2008). The type strain of *S. johnsonii* (self-fertile) possesses both receptor genes *STE3.A1* and *STE3.A2*. Asterisks indicate the two instances where (post-speciation) common ancestry of *HD1/HD2* alleles associated with opposite mating types is best supported.

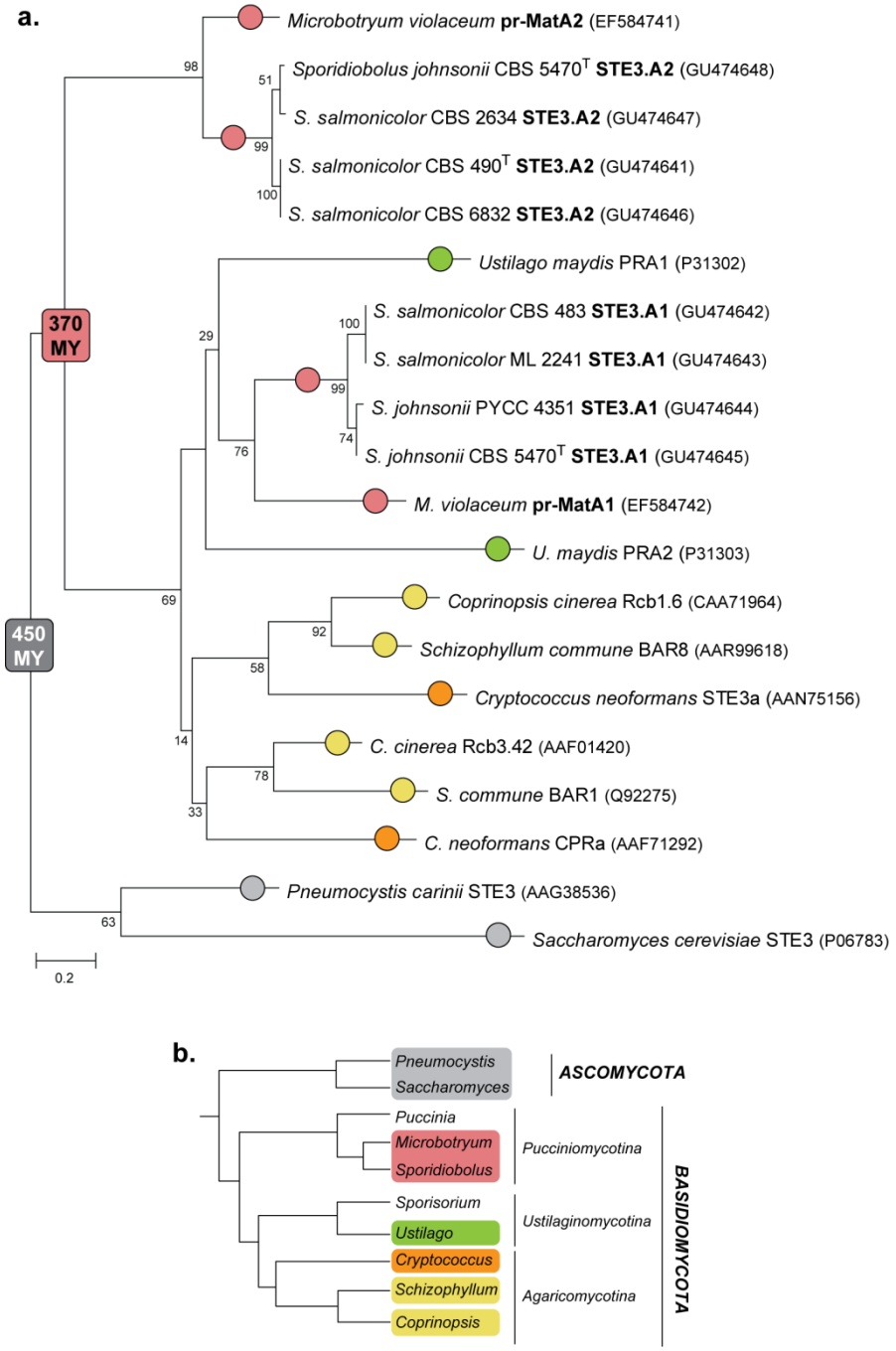

Figure 4.4. Phylogeny of pheromone receptors in the Basidiomycota. (a) Phylogeny and estimated time of divergence based on that of Devier et al. (2009). 450 MY, represents the divergence between ascomycetes and basidiomycetes and 370 MY, corresponds to the divergence between the pheromone receptor alleles in *Microbotryum violaceum*. The phylogeny was estimated by PhyML 3.0 using representative species of the major lineages in the Basidiomycota. Numbers on branches represent statistical support using 100 bootstraps replicates (see 4.2.8 for details). Coloured circles depict phylogenetic groups as defined in the rDNA phylogeny (Hibbett et al., 2007) shown in **(b)**. Sequence accession numbers are indicated in brackets after species names.

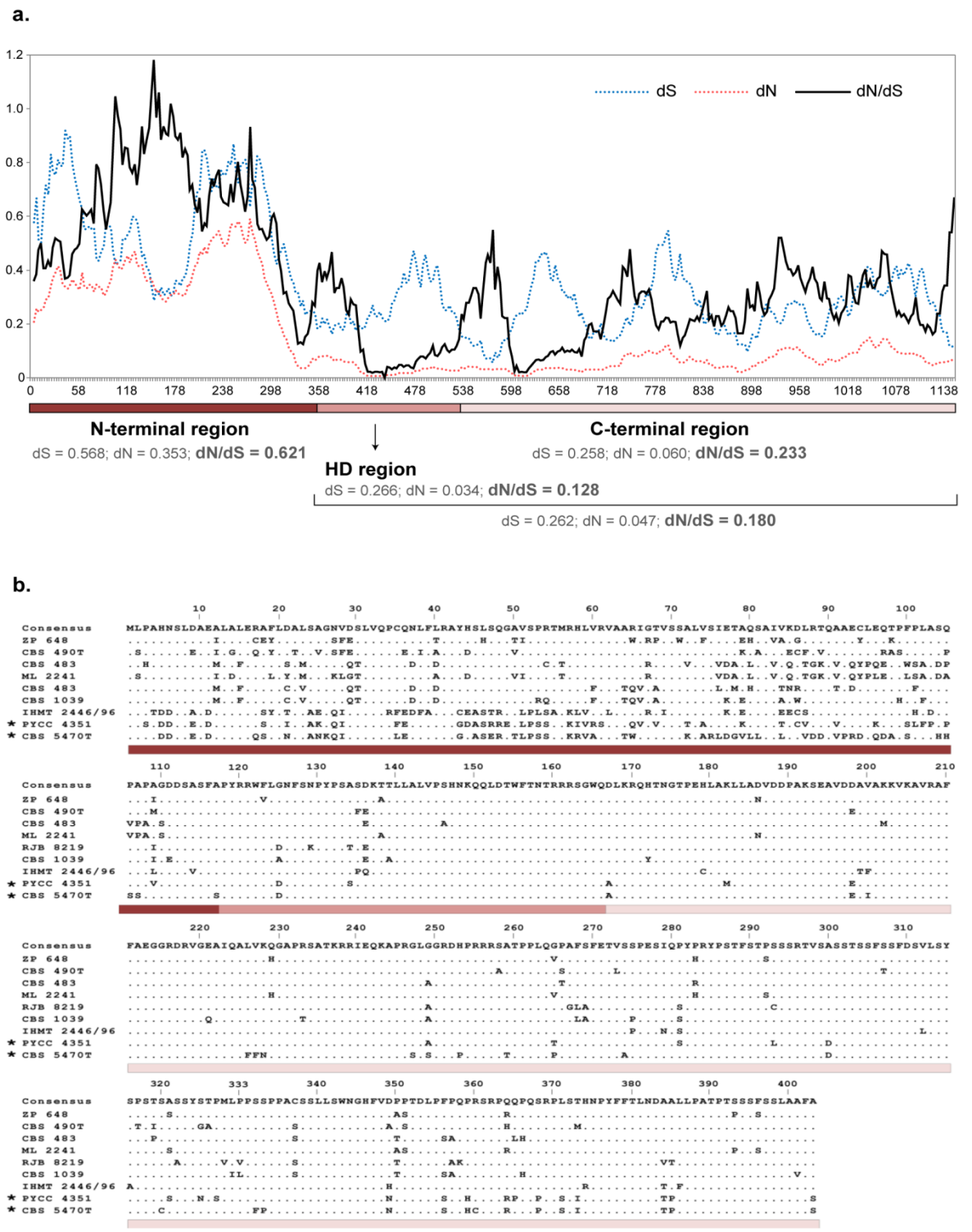

Figure 4.5. Evolution rate of the *HD1* gene in *S. salmonicolor*. (a) Synonymous (dS), non-synonymous (dN) and dN/dS average values for all pairwise comparisons between nine alleles of the *HD1* gene. Y-axis – evolution rates; X-axis – base pairs of *HD1* coding region. Average values are given for the highly variable N-terminal domain, which is involved in non-self recognition, and for the more conserved homeodomain (HD) and C-terminal domains. **(b)** Alignment of the nine deduced HD1 protein sequences used in (a). Coloured bars indicate the same three domains of the HD1 protein as in (a). Sequences marked with an asterisk belong to the sibling species *S. johnsonii*.

4.3.3. A pseudo-bipolar mating system

At this point, an explanation was needed for how the ‘‘one-to-many’’ ratio between pheromone receptor and *HD1/HD2* genes, characteristic of tetrapolar systems, resulted in a bipolar behaviour with only two mating types in *S. salmonicolor*. To this end, the possibility that the cohesion of the bipolar *MAT* locus in *S. salmonicolor* might be occasionally disrupted, giving rise to new receptor/HD allele combinations was investigated. Such recombination events would explain the common ancestry of extant A1- and A2-linked *HD1/HD2* alleles (Figure 4.2 and 4.3) and indeed the fact that all *HD1/HD2* alleles in the species seem to share a common ancestor more recent than the formation of the species itself (Figure 4.3). Therefore, bipolarity in *S. salmonicolor* could be a consequence of the scarcity of recombination events on the one hand and of the exceptionally high evolution rates of HD1/HD2 proteins on the other hand (Figure 4.5). The combination of these two factors would lead to the association of each extant *HD1/HD2* allele with only one of the alternate receptors, the hallmark of bipolarity. Alternatively, it could not be readily excluded that the observed apparent linkage could be due to population genetic phenomena related to the frequency of sexual reproduction and population size. To address the possibility of partial genetic linkage between the two *MAT* regions, the late stages of the sexual cycle were studied in detail (germination of teliospores with the formation of basidia and basidiospores) (Figure 4.6) in one cross between strains ML 2241 (*MAT A1*) and CBS 6832 (*MAT A2*). The progression through meiosis was microscopically monitored showing that although four nuclei could be observed after meiosis, only two basidiospores were formed, each arising from one of the two basidial compartments (Figure 4.6b). The two basidiospores were always found to be binucleated and they typically exhibited asynchronous germination (Figure 4.6). The molecular mating types of the progeny of eight meioses from the same cross were examined by micromanipulation of teliospores in the initial stages of germination. Several colonies isolated from each germinating teliospore were analysed and, in only one of the eight events examined (T1), were the two versions of the parental mating type genes recovered (Figure 4.7). The interpretation for this is that the asynchronous germination of the two basidiospores results in a very strongly biased composition of the colony towards descendants of the basidiospore that germinates first. Micromanipulation of basidiospores instead of the germinating teliospore would be more adequate to recover the germination products of both basidiospores with higher frequency, but this proved to be technically very difficult due to the fact that germination takes place inside the agar and the structures to be manipulated are friable. Furthermore, the various colonies examined for each teliospore are most likely mitotic clones, as they invariably shared the same parental allele of the *DMC1* gene (Figure 4.7) (GenBank: HM133872–HM133874). In six of the meiotic events examined, linkage between the pheromone receptor and the HD1/HD2 regions was maintained, as would be expected in a bipolar system (Figure 4.7). Germination of one additional teliospore yielded a strain which was apparently diploid since it carries pheromone receptor and *HD1/HD2* alleles from both parental strains (T8, Figure 4.7) and has a self-fertile phenotype with the formation of mycelium (apparently monokaryotic) and teliospores (Figure 4.8).

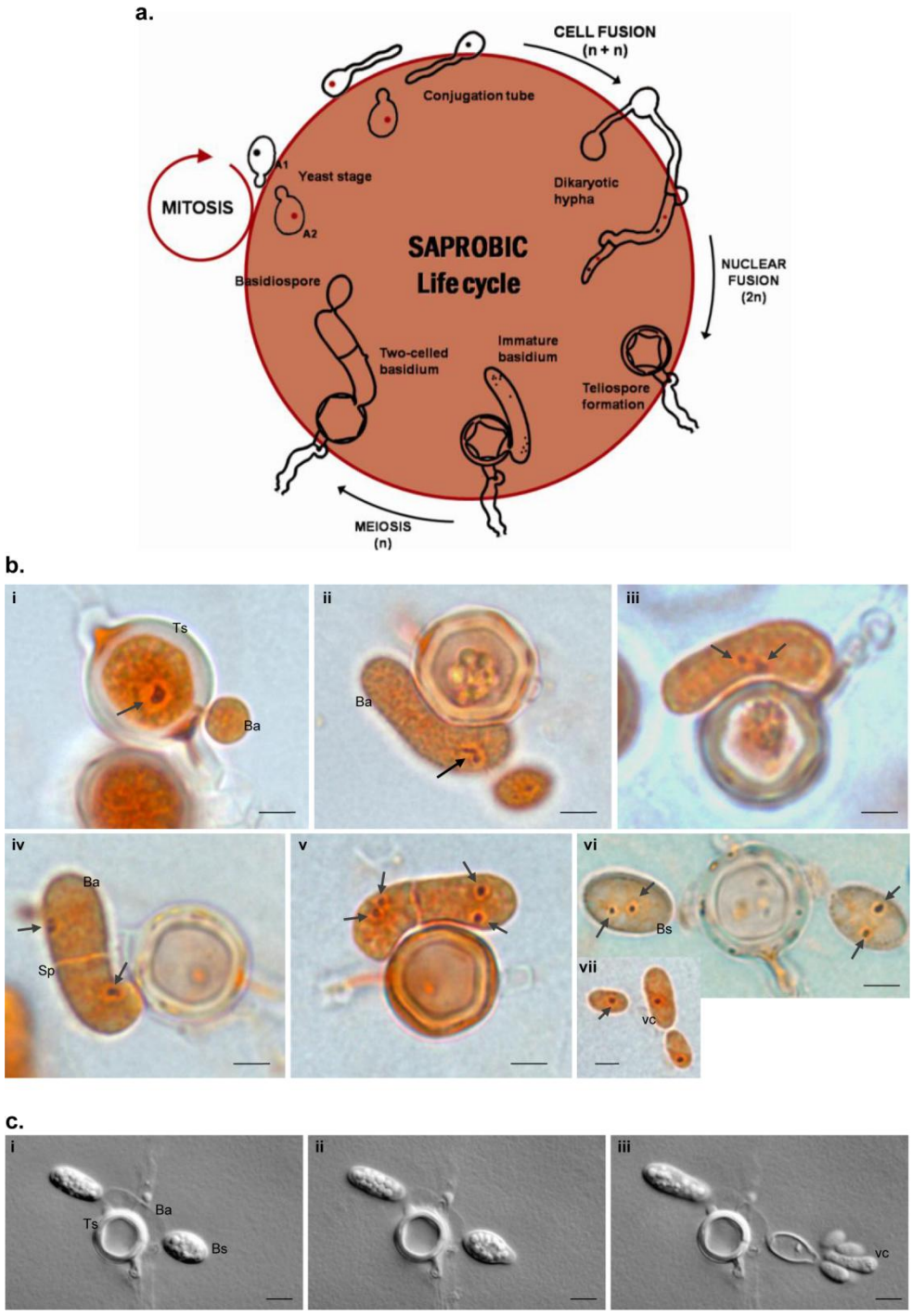

Figure 4.6. Life cycle of *Sporidiobolus salmonicolor*. (a) Diagram representing the saprobic life cycle of *S. salmonicolor*. **(b)** Different stages of teliospore germination and formation of basidia and basidiospores; nuclei were stained with safranin O and are indicated by arrows: (i) initial stage of teliospore [Ts] germination with diploid nucleus and basidium initial [Ba]; (ii) migration of the diploid nucleus to the immature basidium; (iii) first meiotic division; (iv) formation of septum [Sp] originating a two-celled basidium; (v) second meiotic division yielding four nuclei; (vi) formation of two large binucleated basidiospores [Bs]; (vii) haploid uninucleated vegetative yeast cells [vc] resulting from basidiospores germination. Bars = 2.5 μm. **(c)** Time course observation of asynchronous germination of the basidiospores (i–iii). Note that, contrary to the basidiospore on the left, the basidiospore on the right formed yeast cells. Bars = 5 μm.

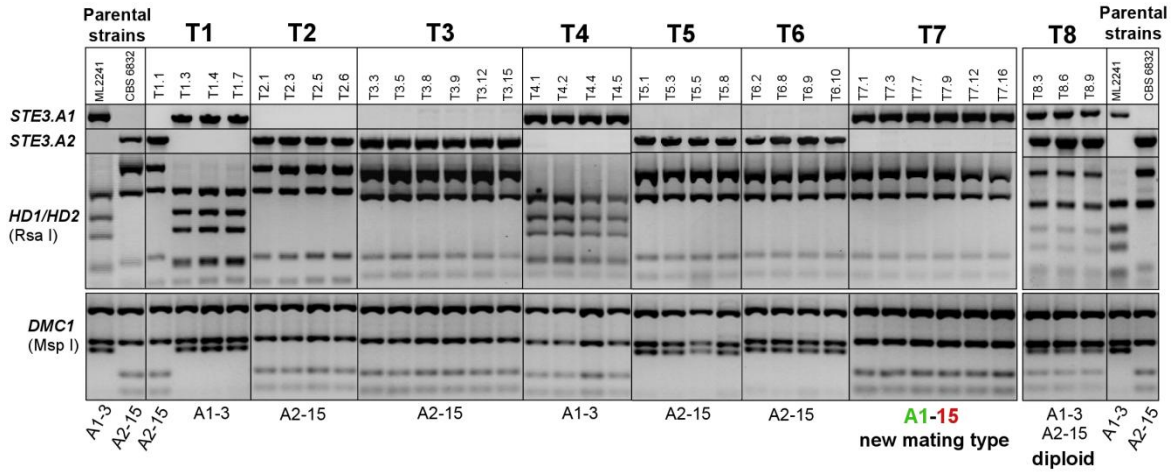

Figure 4.7. Segregation analysis of the *STE3* and *HD1/HD2* genes after meiosis. T1 to T8 refer to eight individually germinated teliospores obtained by micromanipulation from the cross between strains CBS 6832 (A2–15) and ML 2241 (A1–3). The result of the diagnostic PCRs for the presence of the alternate *STE3.A1* and *STE3.A2* genes and the identification of *HD1/HD2* amplicons by Rsa I digestion is shown. Screening of the parental origin of the *DMC1* allele by PCR and digestion with Msp I show clonality of the various colonies examined for each teliospore. For each germinated teliospore, three to six colonies were studied. With the exception of teliospore T1, only one mating type was recovered per meiosis presumably due to asynchronous germination of basidiospores (see Figure 4.6c). Teliospore T8 yielded a strain that is apparently diploid, and the new mating type was recovered from teliospore T7.

Self-fertile and asexual strains can be found among natural isolates of red yeasts of the Sporidiobolales (Sampaio, 2011a; Sampaio, 2011c) and the origin of the first may be related to the occasional formation of diploid yeast strains. In addition to the present observations, the formation of diploid strains was also previously reported in *R. toruloides* (Abe and Sasakuma, 1986). However in none of the species studied so far are these diploid states prevalent, leading to the presumption that although these strains do not exhibit obvious growth defects, their genetic makeup may carry some selective disadvantages. Moreover, preliminary observations suggest that teliospores formed by the *S. salmonicolor* diploid strain do not germinate, which argues against the co-existence of homothallic and heterothallic lifestyles in this species. Interestingly, the type strain of *S. johnsonii* (CBS 5470^{T}) may have originated from a diploid strain because it carries both pheromone receptor alleles. However, only one *HD1/HD2* gene pair is present which may indicate that a self-compatible *HD1/HD2* gene pair was generated by recombination between the two parental alleles (Figure 4.3). The type strain of *S. johnsonii* is homothallic (Valério et al., 2008; Sampaio, 2011c) but does not form basidia, differing in that respect from all the other strains of this species, which carry only the *MAT A1* receptor gene and are in general capable of mating with *S. salmonicolor MAT A2* strains.

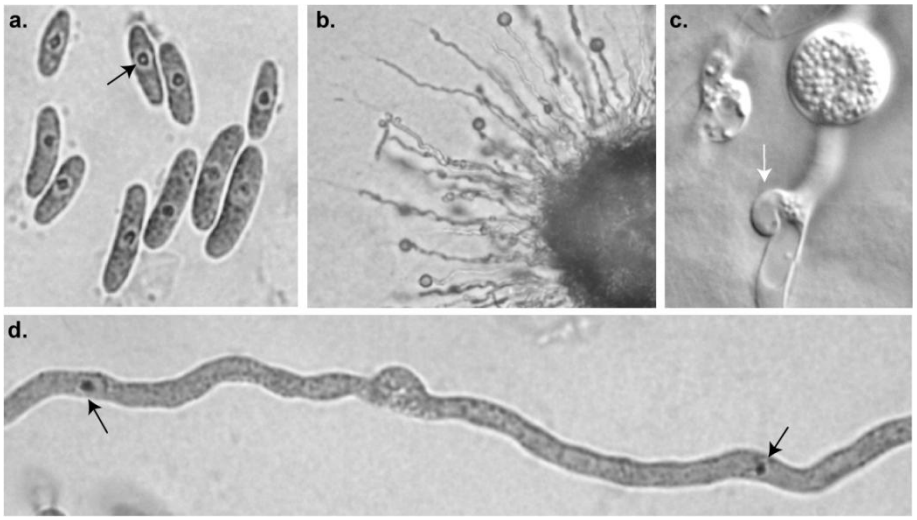

Figure 4.8. Self-fertile phenotype of the diploid strain T8. Nuclei were stained with safranin O and are indicated by black arrows. **(a)** Uninucleated, probably diploid, vegetative yeast cells resulting from basidiospores germination of a cross between strains CBS 6832 and ML 2241 (see Figure 4.7). **(b)** Single cell-derived colony with formation of mycelium and teliospores. **(c)** Mature teliospore and mycelium with incomplete clamp connection (white arrow). **(d)** Monokaryotic mycelium.

One germinated teliospore yielded the most striking result, since it produced an haploid strain in which recombination occurred, causing the *HD1/HD2* allele of parental strain CBS 6832 (*MAT A2*) to become associated with the Ste3.A1 receptor (Figure 4.7), a combination that is not found among natural isolates (Figures 4.2 and 4.3). The new mating type (strain T7) was unable to complete the sexual cycle when crossed with either parental strain (Figure 4.9). A cross with the parent sharing the same pheromone receptor gene resulted in a complete failure to switch to filamentous growth, whereas homozygosity at the *HD1/HD2* region precluded progression through the sexual cycle, but allowed some pseudohyphal growth, probably as a result of cell-cell pheromone signalling (Figure 4.9). This demonstrates unequivocally that both compatibility regions are required for sexual reproduction in *S. salmonicolor*. The new mating type did not show obvious defects in vegetative growth, exhibiting growth rates in synthetic and complete media similar to those of the parental strains. It also had normal sexual proficiency when crossed with strains carrying different alleles at both *MAT regions* (Figure 4.9), forming dikaryotic mycelium with teliospores that germinated normally. Recombination between the two classes of *MAT* genes seems to be possible but infrequent, a situation that, to our knowledge, has never been described before in basidiomycetes and can be regarded as intermediate between bipolar and tetrapolar systems. Therefore, this observation called for a closer examination of the meiotic events described above. In particular, it was important to establish how genetic markers unrelated to *MAT* segregated in these meiotic events, thereby getting some insight in the recombination frequency characteristic of this species in

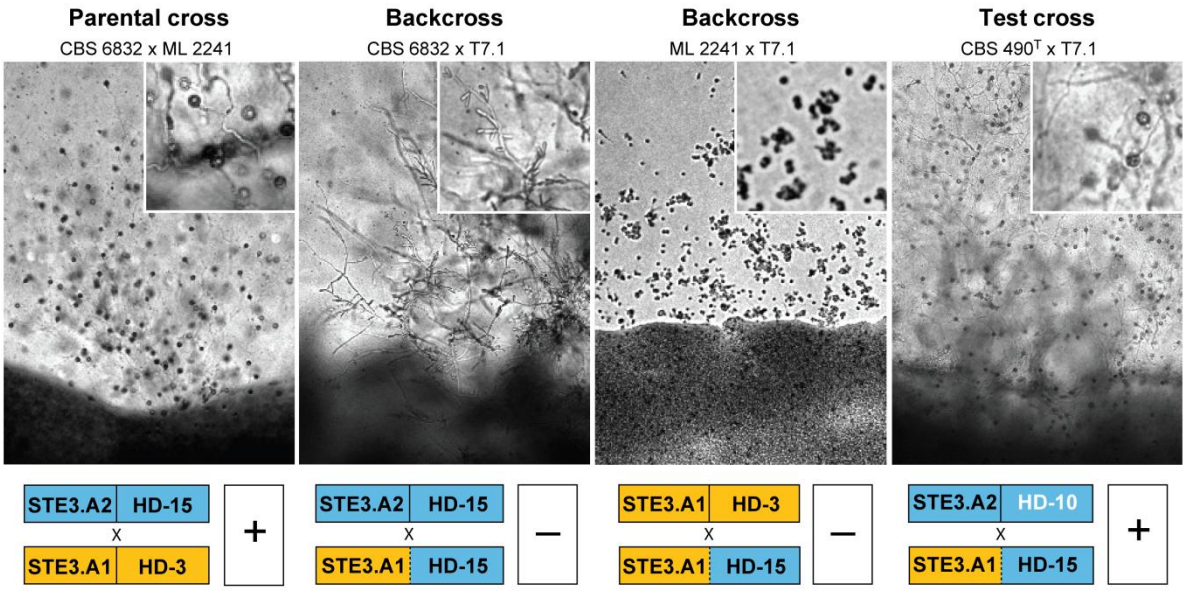

Figure 4.9. Sexual proficiency of T7.1, the new mating type of *S. salmonicolor*. For each cross, the top section indicates the type of cross and the strains involved. Micrographs of crosses (4 days of incubation at 25ºC on corn meal agar) are shown in the middle section (inserts depict details in higher magnification to show the presence or absence of teliospores). The bottom section shows the *STE3* and *HD* genotypes of each strain and the outcome of the cross evaluated by the presence (+) or absence (–) of mycelium with clamp connections and teliospores.

autosomal regions while simultaneously confirming that a normal meiosis had taken place in all the cases examined. It should be noted that a more precise determination of recombination frequencies and a statistically significant demonstration of linkage between genomic regions would require the examination of a larger number of meiotic products, but nevertheless examination of the eight available strains yields significant information concerning recombination inside and outside the *MAT* region.

To characterize the haploid progeny strains, nine genes located in four different *Sp. roseus* scaffolds were chosen (Figure 4.10). Pairs of genes located on the same scaffold were at distances ranging between 0.525 and 1.87 Mb from each other (Figure 4.10b). The selected genes were partially amplified in the two parental strains used in this cross (ML 2241 and CBS 6832) and sequence polymorphisms were scored which allowed the two parental alleles to be distinguished and their fate after meiosis to be tracked (GenBank: HM133857–HM133883). Sequencing of this set of genes in the seven haploid progeny strains confirmed that these markers segregate independently of *MAT*-specific genes (Figure 4.10). Markers located more than 1.2 Mb apart seem to be genetically unlinked, while the results suggested partial linkage for genes that are 525 to 920 kb apart, with an average of 32 kb/cM for these autosomal regions. The scaffolds containing the two *MAT* regions were subsequently subjected to a similar analysis. In scaffold 7, five genes, including the *HD1/HD2*

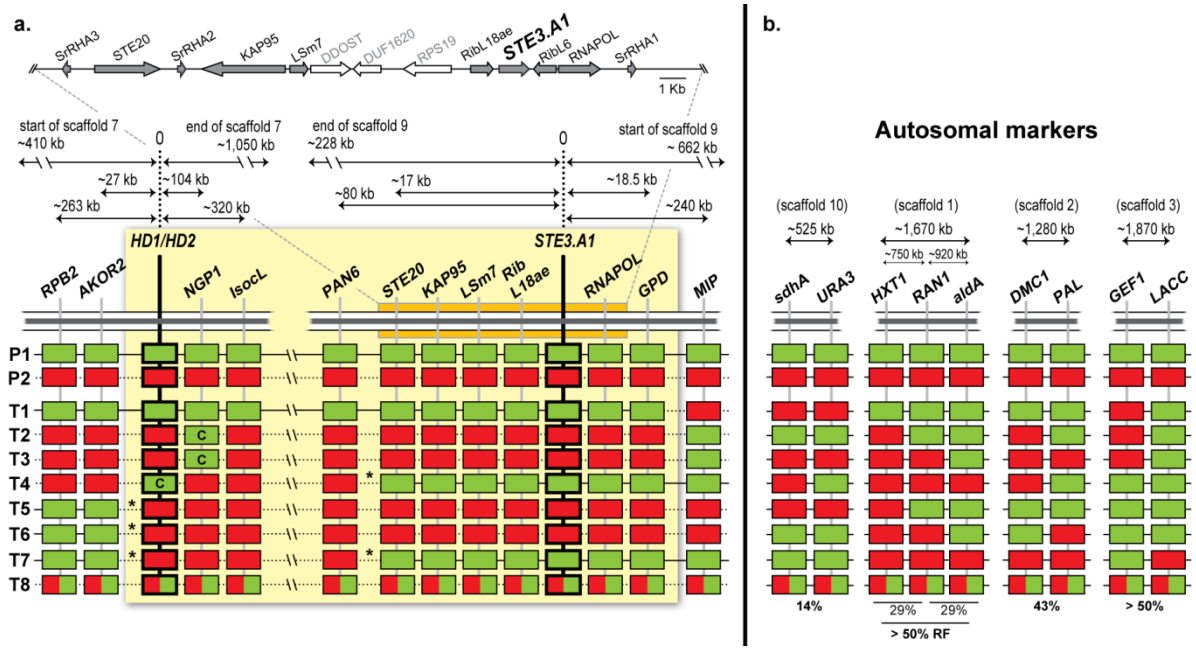

Figure 4.10. Analysis of meiotic progeny. (a) Genomic regions surrounding the *HD1/HD2* and pheromone receptor genes (*STE3*) in *S. salmonicolor MAT A1*. Genes in the immediate vicinity of *STE3.A1* are depicted as block arrows to denote the direction of transcription. The same genes are highlighted with an orange box. Synteny with the homologous genomic region of *Sp. roseus* was confirmed for all genes except those shown as empty arrows (depicted in the arrangement present in *Sp. roseus*). Outside the highlighted region, genes present in the same *Sp. roseus* scaffold at different distances (indicated by the arrows) from *STE3.A1* were selected. Four genes were similarly selected from the *HD1/HD2* scaffold, assuming synteny with *Sp. roseus*. For eight independent teliospores (T1–T8) the parental origin of the genes is indicated. Genes from the parental strain ML 2241 (P1, A1–3) are shown in green and those from strain CBS 6832 (P2, A2–15), in red. Asterisks mark sites were crossovers were detected. Possible gene conversion events are marked with ''C''. The two scaffolds are depicted in the most likely orientation. The region shadowed in light yellow encompasses the putative pseudo-bipolar *MAT* locus in *S. salmonicolor MAT A1*. **(b)** Segregation patterns and recombination frequencies (RF) of nine autosomal markers located in four different scaffolds. The position and relative distances between the genes are depicted as in *Sp. roseus* genome. T8 is a diploid strain since it contains both parental alleles for all genes tested.

pair were found to contain polymorphisms and were studied with respect to their parental origin in the seven haploid progeny strains (GenBank: GU474757–GU474780, HM133833–HM133856). In the scaffold harbouring the *STE3* gene, the parental origin of a total of eight additional genes exhibiting polymorphisms was determined (GenBank: GU474709–GU474756, HM133785–HM133832). The relative orientation of the two scaffolds could not be previously established (Coelho et al., 2008), but the results of the scrutiny of meiotic recombination in these regions suggest linkage between the *IsocL* gene in scaffold 7 and the *PAN6* gene in scaffold 9. This implies that the relative orientation of the two scaffolds is most likely as depicted in Figure 4.10. The analysis of the markers in these regions brought to light, in addition to sites where probably crossovers occurred, several

instances where very likely gene conversion took place instead, one of which curiously involves the *HD1/HD2* gene pair (Figure 4.10a). In line with this, it was found that the phylogeny of several genes located in the two scaffolds harbouring *MAT* genes also occasionally denotes signs of past gene conversion events when multiple strains of *S. salmonicolor* and *S. johnsonii* are examined (Figures 4.11 and 4.12). For example, the *LSm7* gene located close to *STE3*, exhibits an unusual mosaic phylogeny, which is species-specific in most of the gene but is *MAT* specific at the 3' end, proximal to *STE3* (Figure 4.11). This suggests that this gene may have integrated the *MAT* locus long ago as might be expected from its close proximity to a key *MAT*-specific gene, but that the 5'-end underwent conversion more recently, after incipient separation of the two species. Phylogenetic analysis of the *RibL18ae*, *RNAPOL*, *NGP1* and *AKOR2* genes also suggests gene conversion events in *S. salmonicolor* strain CBS 1012 (Figures 4.11 and 4.12). These observations support the idea that gene conversion is an important mechanism to maintain species-specific sequences at the *S. salmonicolor MAT locus*. This may, in turn, reflect the need to counteract the effect of recombination suppression which could lead to mating type-linked deleterious mutations. It is also important to note that sequence divergence between mating types of genes located in the vicinity of the pheromone/receptor *MAT* region is much lower in the red yeasts examined than, for example in *Cryptococcus neoformans* (Figures 4.11 and 4.12) (Fraser et al., 2004), although in the later species some events of gene conversion within the *MAT* loci have also been noticed (Fraser et al., 2004). Three crossovers were mapped to a region of 27 kb adjacent to the *HD1/HD2* gene pair (T5, T6 and T7; Figure 4.10), which seems to configure a hotspot for recombination (0.63 kb/cM) when compared with the average frequency of recombination observed for the autosomal regions examined (~32 kb/cM). Interestingly, hotspots have been mapped close to the borders of the *MAT* locus in *C. neoformans* (Hsueh et al., 2006) and were proposed to play an important role in the evolution of these specialized genomic regions. Three recombination events were also detected on the opposite side of the putative *MAT* locus in a region of ~220 kb (T1, T2 and T3; Figure 4.10) but in this case only one genetic marker (*MIP*) was studied on one side of this position, which makes it difficult to distinguish between gene conversion and crossover events. Finally, two crossovers were detected in the proximity of the *PAN6* gene, one of which was the cause of the only instance of recombination detected between the *STE3* and the *HD1/HD2* genes (T7, Figure 4.10). In the other case (T4, Figure 4.10), the *MAT A1* parental mating type prevailed, due to apparent gene conversion involving the *HD1/HD2* gene pair. In the 1.2 Mb region between the *HD1/HD2* and *PAN6* genes no crossovers seem to have occurred, suggesting that recombination frequency is lower than average or possibly suppressed in this region (Figure 4.10).

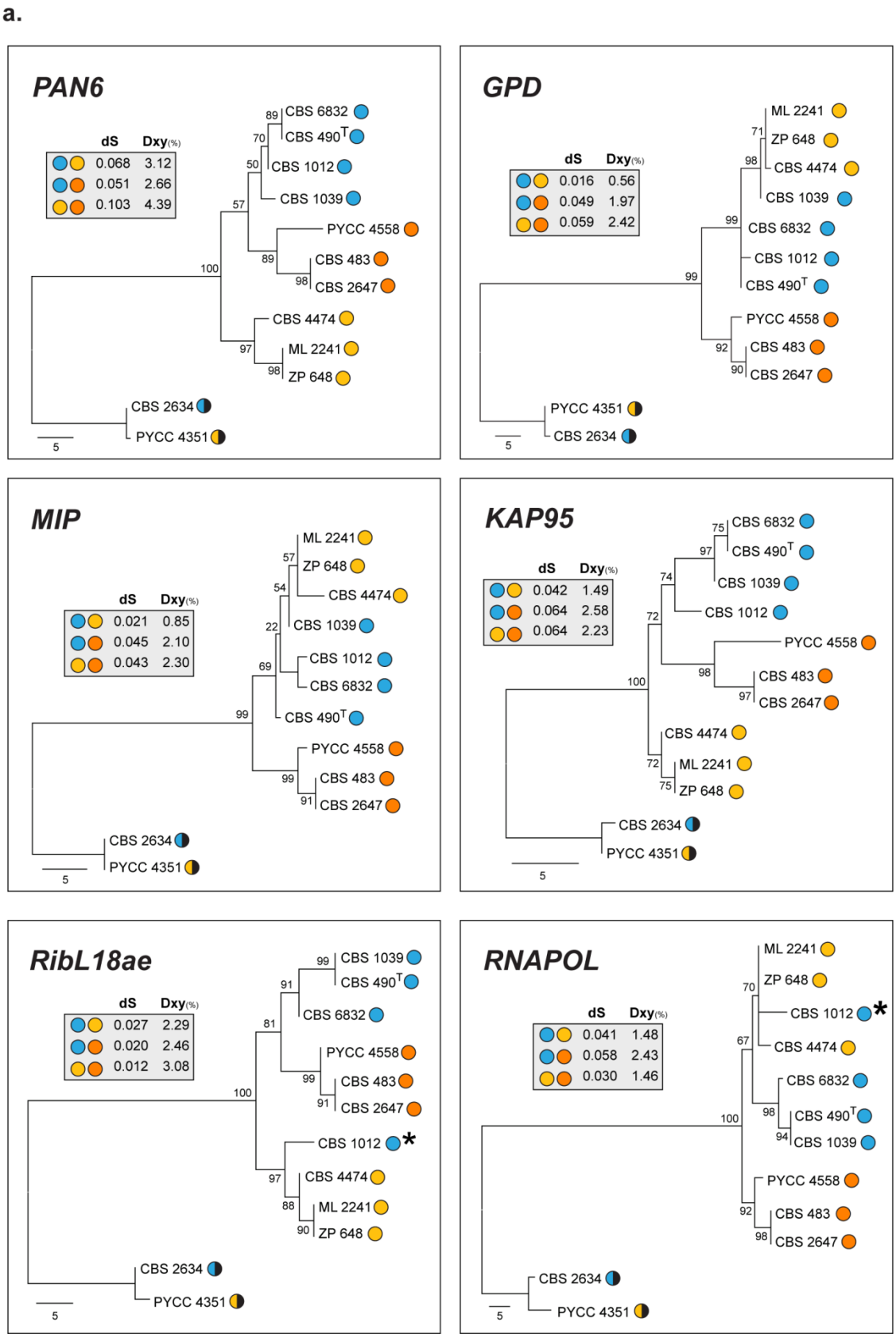

Figure 4.11. Phylogeny of several genes located in the scaffold harbouring the *STE3* region. See page 101 for description.

b.

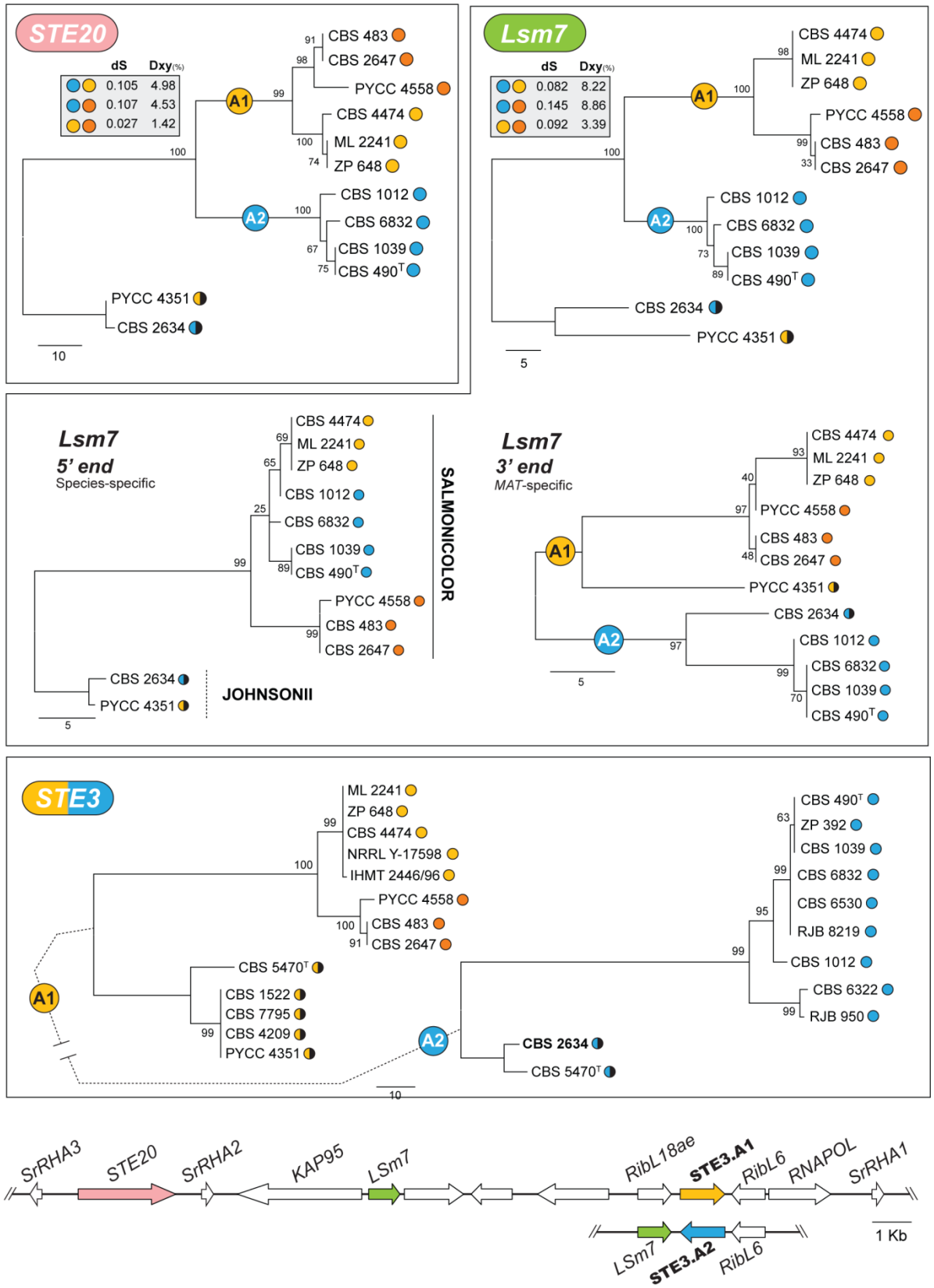

Figure 4.11. Continued. See next page for description.

Figure 4.11. Phylogeny of several genes located in the scaffold harbouring the *STE3* region. (a) The phylograms represent the single most likely tree for the selected genes. Trees were generated under the close-neighbour interchange algorithm in a heuristic search. Numbers on branches indicate statistical support calculated from 1000 bootstrap replicates. Possible gene conversion events are marked with an asterisk after the strain number. For the majority of the genes, sequences of strains of *MAT A2* (indicated by blue circles) are highly similar. In contrast, *MAT A1* strains exhibit more sequence divergence, suggesting the existence of two lineages within this mating type (strains indicated by yellow and orange circles). Half-coloured circles indicate *S. johnsonii* strains. **(b)** Trees depicting genes with sequence divergence clearly associated with mating type in *S. salmonicolor*. The position of the *LSm7* (green) relative to the *STE3* is shown for both mating types. Trees that used only the 5' or 3' end regions of *LSm7* gene are shown. Values for synonymous substitutions (dS) and divergence percentage (Dxy, %) are given for all genes (except for *STE3*) and in each pair comparisons.

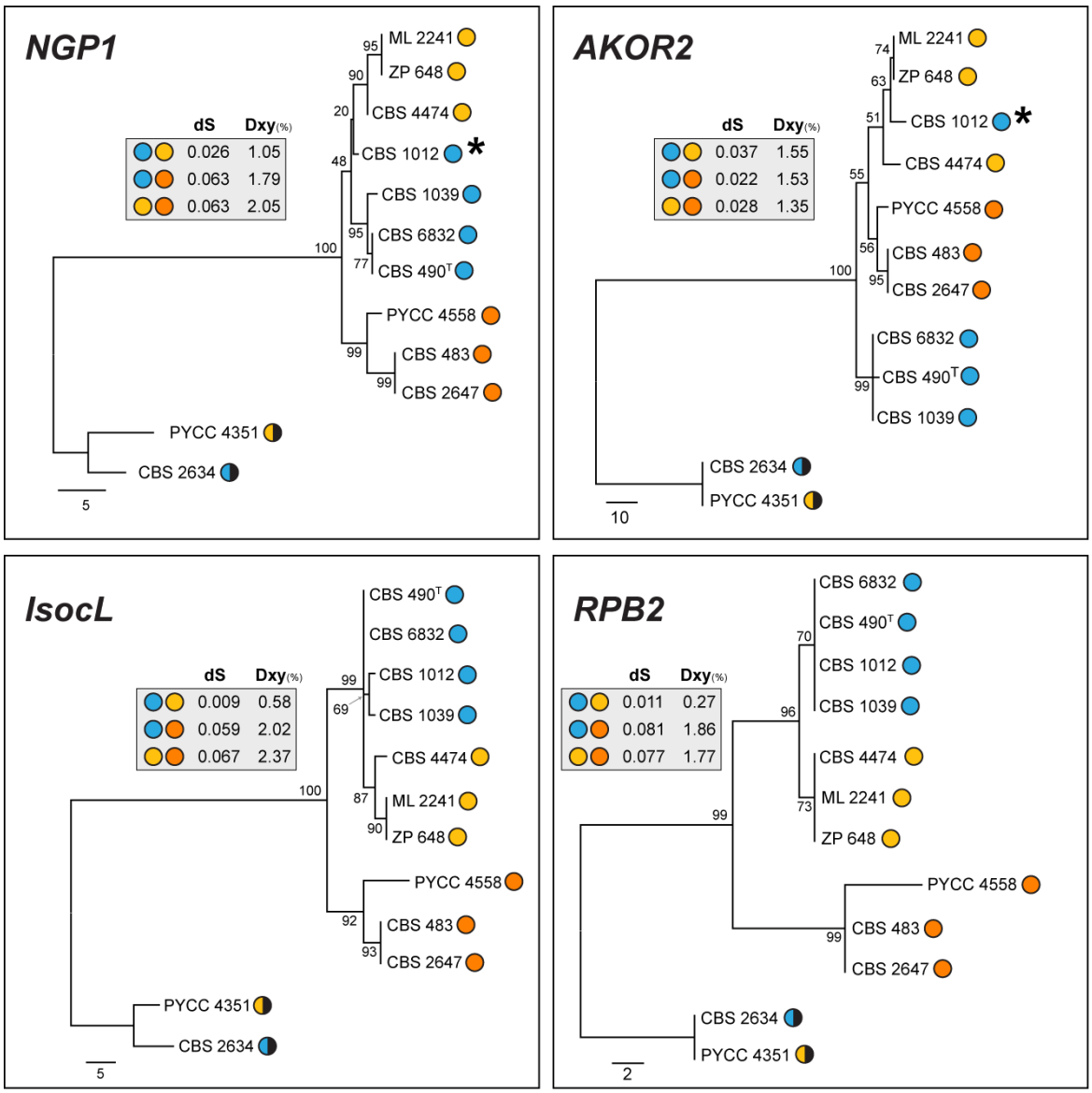

Figure 4.12. Phylogeny of genes located in the scaffold harbouring the *HD1/HD2* region. Features are represented as in Figure 4.11.

Hence, it is proposed that in *S. salmonicolor* a new type of fungal sex determining region operates, which is neither tetrapolar nor strictly bipolar. *S. salmonicolor* seems rather to have a large pseudo-bipolar *MAT* locus in which recombination in the region between the two *MAT* regions is infrequent but not suppressed, allowing for occasional disruptions of its genetic cohesion. The proposal of an intermediate system between bipolar and tetrapolar is based mainly on two kinds of observations, which are in line with each other. Firstly, the allelic distribution and number of *MAT* genes depicted in Figure 4.2 outlines a situation that is itself intermediate between bipolar and tetrapolar (multiple *HD1/HD2* alleles but only two mating types). Secondly, when trying to clarify this new finding, it was found that recombination may occur in the intervening region between the two *MAT* regions, which provides a likely and plausible explanation for the observed resemblance with tetrapolar systems in what concerns evolution and number of *HD1/HD2* alleles. On the other hand, the genomic region around the receptor/pheromone locus exhibiting synteny breaks, gene inversions and sequence divergence between the two mating types is significantly larger in *S. salmonicolor* than in tetrapolar species with similar numbers of *MAT* gene alleles, like *U. maydis* (Figure 4.1) (Fraser et al., 2004). It is envisaged that the offspring of the relatively rare events of meiotic recombination in the *MAT* locus of *S. salmonicolor* can be rescued from reproductive isolation by the existence of multiple *HD1/HD2* alleles associated to each of the two alternate receptors in natural populations. In such a pseudo-bipolar system, negative frequency-dependent selection could operate to preserve *HD1/HD2* allele diversity and explain the exceptionally high evolution rates of HD1 proteins.

4.3.4. A model for the evolution of *MAT* loci in basidiomycetes

Both the content and the organization of the basidiomycete *MAT* loci are unlike all other fungi, because compatibility is encoded in the *MAT* locus itself, while in other fungi the genes present in the *MAT* locus regulate expression of pheromones and pheromone receptors that are encoded elsewhere in the genome. The available data does not suffice to address definitively the question of how the basidiomycete *MAT* locus arose, but one hypothesis has been recently put forward that postulates that the pheromone receptor and/or the HD systems may have evolved in a self compatible manner (Fraser et al., 2007). The emergence of self-incompatible alleles by mutation or recombination could have subsequently laid the basis for the emergence of the tetrapolar system. There is currently insufficient evidence, also including the observations presented in this Chapter, to state decisively whether the ancestral basidiomycete mating system was initially bipolar or tetrapolar, at the onset of the involvement of both the HD and pheromone receptor systems in determining mating specificity. However, the existing observations concerning the mating system in *S. salmonicolor*, suggest that novel and useful insights may be gained in this respect by a wider phylogenetic sampling of mating systems in the Pucciniomycotina.

In line with this, the *HD1/HD2* allelic distribution was subsequently looked into *Rhodosporidium babjevae*, another bipolar red yeast species phylogenetically related to *S. salmonicolor*. Also in this case, multiple *HD1/HD2* alleles were associated with each receptor (Figure 4.13). Since this allelic distribution is the hallmark of the pseudo-bipolar system, it can be concluded that rather than being an oddity restricted to *S. salmonicolor*, this system is likely to operate in many of the yeast species of the order Sporidiobolales (Sampaio et al., 2003) and probably also in other members of the subphylum Pucciniomycotina that were thus far thought to be bipolar. The most parsimonious explanation for this is that a very similar system was present in the common ancestor of *Rhodosporidium* and *Sporidiobolus*. This implies that pseudo-bipolarity is not a short-lived transitional mating system, like those presumed to have marked the transition to the bipolar state in *C. neoformans* (Lengeler et al., 2002; Hsueh et al., 2008) and *U. hordei* (Bakkeren and Kronstad, 1994). Rather, it suggests that the pseudo-bipolar system may be evolving from an ancestral configuration in which the two classes of *MAT* genes were located on the same chromosome, but for which it is not possible to infer from currently available data whether or not the two *MAT* regions were genetically linked.

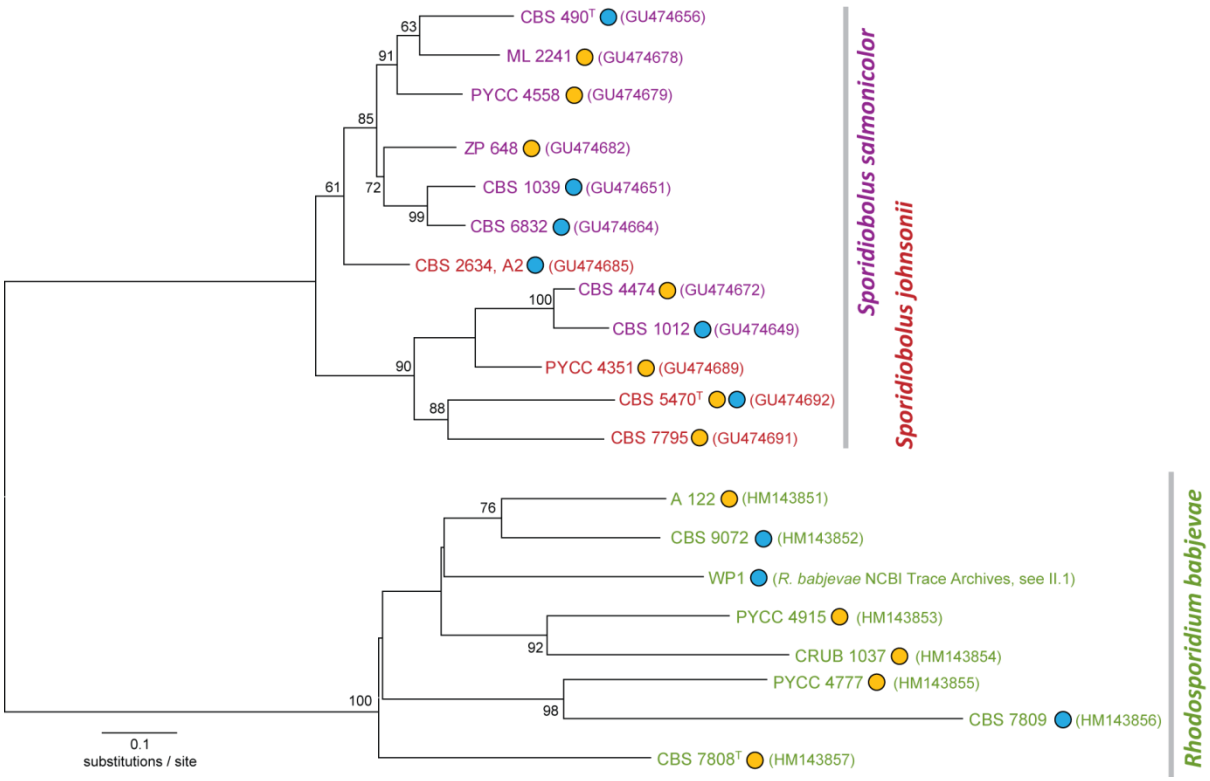

Figure 4.13. Phylogeny of *HD1/HD2* alleles in *Rhodosporidium babjevae*. The pheromone receptor gene (*STE3.A1*, yellow circles; *STE3.A2*, blue circles) and sequence accession numbers (in brackets) are indicated after strain number. The tree was reconstructed with Neighbour-joining and TrN+G model (shape parameter =1.68453). Bootstrap values (>50%) from 1000 replicates are shown. *S. salmonicolor* and *S. johnsonii* are included for comparison.

The model shown in Figure 4.14 depicts the possibility that the ancestral basidiomycete mating system may have been similar to the pseudo-bipolar system found in *S. salmonicolor*. This possibility is highlighted among others, because the Pucciniomycotina is the earliest derived basidiomycete lineage and seems therefore more likely that members of this group have retained mating systems akin to the common ancestral system. In this scenario, the Pucciniomycotina would have diverged before the stabilization of the tetrapolar system (with the two *MAT* regions on different chromosomes), which is presumed to be the ancestor of the mating systems in the other two lineages (Ustilaginomycotina and Agaricomycotina) (Fraser et al., 2007).
The model proposed in Figure 4.14 depicts this possibility and accommodates the possible occurrence of abrupt as well as gradual transitions to bipolar mating behaviour, the first but not the second being triggered by gross genomic rearrangements.

4.4. Outlook and final remarks

An intermediate, pseudo-bipolar mating system entails odds of inbreeding that are in between those observed for the bipolar and tetrapolar systems (50% and 25%, respectively) (Hsueh et al., 2008). This may provide for some species the right balance between genetic stability (imparted by inbreeding) and variation (favoured by outcrossing) and may, in turn, delay the attainment of a strictly bipolar state if a system is drifting away from a tetrapolar configuration without abrupt occurrences, such as large chromosomal rearrangements. Furthermore, the time course microscopic observation of meiotic events in *S. salmonicolor* indicates that basidiospores exhibit asynchronous germination (Figure 4.6c), decreasing the probability of a cross between siblings immediately after meiosis. This, together with the occasional meiotic recombination events, would generally decrease the odds of selfing when compared with a typical bipolar system, like that of the phylogenetically related plant pathogen *Microbotryum violaceum* (Giraud et al., 2008). Hence, the pseudo-bipolar mating system emerges as a remarkable novel context in which to explore how life-style, ecology and modes of reproduction interplay in the evolutionary history of eukaryotes.

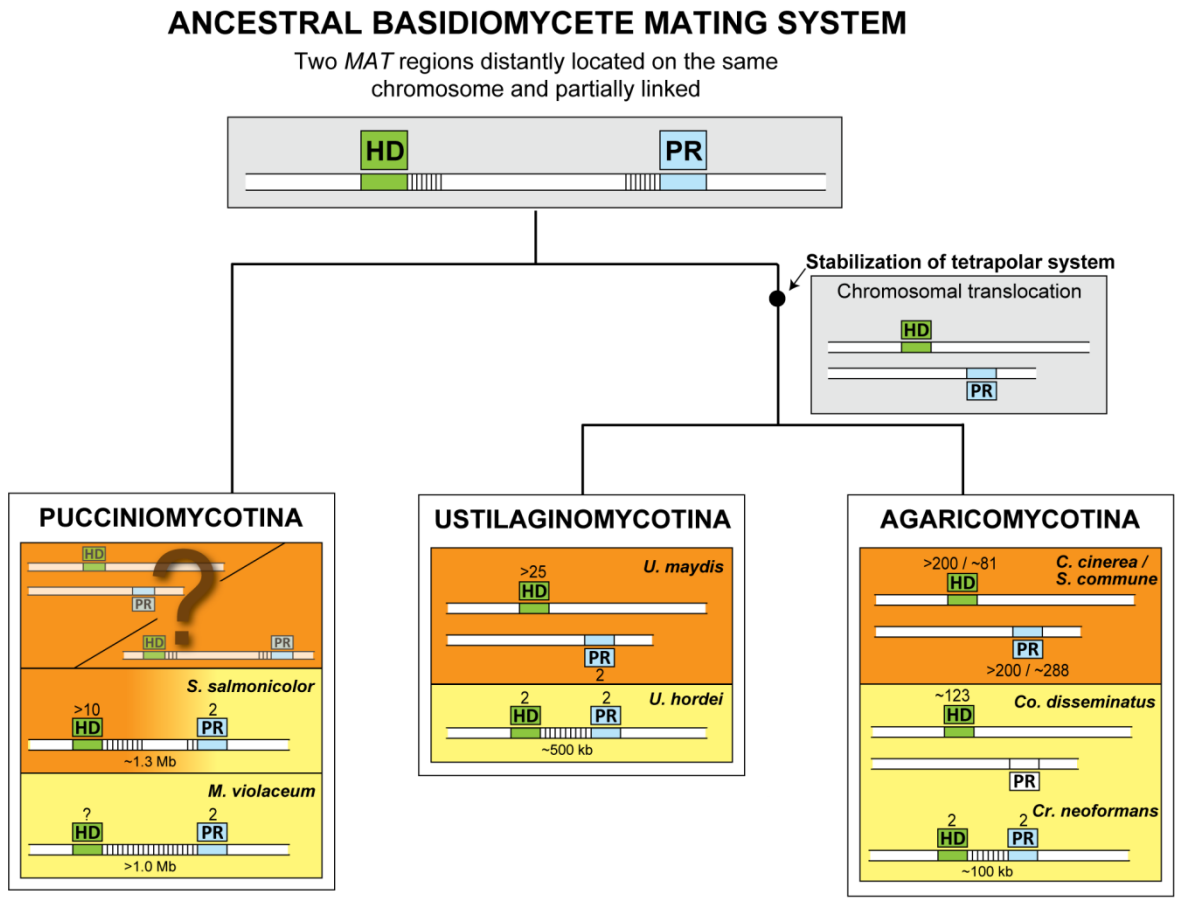

Figure 4.14. The pseudo-bipolar system and the evolution of *MAT* loci in Basidiomycota. The ancestral basidiomycete mating system may have been similar to the pseudo-bipolar system presently found in *S. salmonicolor*. The classic tetrapolar system emerged in the common ancestor of the Ustilaginomycotina and Agaricomycotina through a chromosomal translocation that placed the two *MAT* regions on different chromosomes. In *Ustilago hordei* and *Cryptococcus neoformans,* a second translocation may have triggered a precipitous transition to bipolarity, as previously proposed (Bakkeren and Kronstad, 1994; Hsueh et al., 2008). In the mushroom *Coprinellus disseminatus*, the pheromone receptor (white box) ceased to be associated to *MAT* leading to bipolarity (James et al., 2006). Mating systems in the Pucciniomycotina may have evolved to extant systems exhibiting partial (*S. salmonicolor*) or complete (*Microbotryum violaceum*) suppression of recombination between the two *MAT* regions. Tetrapolar systems in this lineage might be present in *Leucosporidium scottii* and the rusts (Uredinales) but have not been characterized at the molecular level yet. In these systems, *MAT* regions may be distantly located on the same chromosome or on different chromosomes. The number of *MAT* gene alleles identified in each species is shown next to the boxes representing the homeodomain (HD) and pheromone receptor (PR) regions. Stripes between the HD and PR regions denote reduced or suppressed recombination.

5. CHAPTER

Evidence for maintenance of sex determinants but not of sexual stages in red yeasts

Part of the work presented in this chapter was published in:

Coelho M. A., Gonçalves P., and Sampaio J.P. 2011. Evidence for maintenance of sex determinants but not of sexual stages in red yeasts, a group of early diverged basidiomycetes. BMC Evol Biol 11:249.

5.1. Introduction

In addition to the various modes of sexual reproduction observed in fungi, including heterothallic and homothallic mechanisms, asexual reproduction is also a conspicuous mode of fungal propagation. This suggests that for fungi the advantages of sex, namely the promotion of adaptive evolution and the more efficient elimination of deleterious mutations (Kondrashov, 1993), may in many instances be insufficient to counterbalance its disadvantages, normally associated with energy and time costs of sexuality (Maynard-Smith, 1978). The reasons why sexuality can be facultative in fungi seem to be related with the availability of mitotically formed spores and other modes of asexual propagation (Bell, 1982). However, even if the ability to reproduce mitotically in fast and efficient ways might act as an evolutionary promoter of asexuality (Aanen and Hoekstra, 2007), several lines of evidence indicate that asexuality is a derived character. Firstly, asexual species are clearly a minority [approximately 17% of all known fungal species are asexual (Hawksworth et al., 1995)]. Secondly, asexual species very often have phylogenetically close relatives capable of sexual reproduction (Kurtzman and Robnett, 1998; Begerow et al., 2000) (see 1.1.2).

A classical and presently much debated rule in fungal nomenclature imposes that asexual and sexual fungi are classified separately, in different species and genera (Hawksworth, 2011). For example, in the order Sporidiobolales sexual species are classified in the genera *Rhodosporidium* and *Sporidiobolus* whereas asexual forms are classified in *Rhodotorula* and *Sporobolomyces* (Sampaio et al., 2003). Therefore, the co-existence in the same phylogenetic group of a set of very closely related species of sexual and asexual taxa with similar ecological and physiological properties raises several questions: (i) are asexual species truly asexual having thus no traces of *MAT* genes in their genomes? (ii) alternatively, have asexual species formed recently and therefore still exhibit (traces of) sex related genes? (iii) is mating type imbalance possibly giving rise to asexual lineages? (iv) is asexuality contributing to speciation by giving rise to new species that derive from extant sexual stocks? These questions are addressed in this chapter by surveying the presence and integrity of *MAT* genes in an expanded set of red yeast species. In order to get some additional insight into the evolution of *MAT* loci in the Sporidiobolales, the gene content and organization around *MAT* genes was compared between opposite mating types and across different species.

5.2. Material and methods

5.2.1. Strains and mating tests

The list of strains studied and relevant information pertaining to them is given in section III.1 (Appendix III). Cultures were grown on MYP agar as described in section 3.2.1. To study sexual compatibility, pairs of 2-4 day-old cultures were mixed on corn meal agar (Difco), incubated at 18 – 22ºC for 1 week and examined microscopically using phase-contrast optics for production of mycelium with clamp connections and teliospores (globose and thick-walled resting structures that are the site of karyogamy). Strains were designated as 'self-fertile' (Sf) when they produced mycelia bearing teliospores without crossing with any other strains or, 'asexual' (As) if no sexual structures were observed in crossings with the available mating types.

5.2.2. PCR detection and sequencing of *MAT A1* and *MAT A2* pheromone receptor genes and the conserved surrounding regions

Diagnostic PCRs with degenerate primers for *STE3.A1* and *STE3.A2* alleles were carried out to directly identify the pheromone receptor genes and (re)assign the molecular mating types in most of the studied red yeast strains (see section III.1, Appendix III). For some selected strains, the amplification products were purified and sequenced with the same primers. However, this approach failed to yield amplification products for some of the species. In these cases, *STE3.A1* and *STE3.A2* alleles were detected by long range PCR using primers based on the available sequences of the more conserved genes flanking the *STE3.A1* in *Sporobolomyces roseus* IAM 13481 (genes *RibL18ae* and *RibL6* or *RNAPOL*) and *STE3.A2* in *R. babjevae* WP1 (genes *LSm7* and *RibL6* or *RNAPOL*) [see section III.2 and Figure III.1 (page 179), Appendix III]. The amplification products were subsequently sequenced by primer-walking. This latter approach was also used to confirm if synteny in the immediate vicinity of the pheromone receptor genes in *MAT A1* and *MAT A2* strains was maintained across an expanded set of red yeast species. For each yeast species, PCR reactions, cycling conditions, primer sequences and GenBank accession numbers are specified in sections III.1 an III.2 of Appendix III. Based on recent phylogenetic evidence (Valério et al., 2008), *Sporobolomyces roseus* IAM 13481 represents most likely a new undescribed species. Therefore due to its unsolved taxonomic placement, it will be henceforth referred as *Sporobolomyces* sp. IAM 13481. Likewise, *R. babjevae* WP1 strain was recently renamed as *Rhodotorula graminis* WP1 based on phylogenetic analyses (Xin et al., 2009).

5.2.3. PCR amplification and sequencing of the *HD1/HD2* region and the *RHA2* pheromone precursor gene

Using the *HD1/HD2* sequences of *S. salmonicolor* and *Sporobolomyces* sp. IAM 13481 (Coelho et al., 2010) (see Chapter 4), a Blastn search was performed in the NCBI Trace Archive database of *Rh. graminis* WP1 and the sequences corresponding to positive hits were assembled. The recently released genome of *Rh. graminis* WP1 (http://genome.jgi-psf.org/Rhoba1_1/Rhoba1_ 1.home.html) was employed to confirm the assembled region, which is located in scaffold 16 (base positions 15900-16300). In turn, the deduced HD1 and HD2 protein sequences of *Rh. graminis*, *S. salmonicolor* and *Sporobolomyces* sp. IAM 13481 were aligned and the conserved regions were used to design degenerate primers (MC118 and MC120) to amplify and sequence the corresponding N-terminal and intergenic regions of the *HD1/HD2* genes in several strains of *R. babjevae*, *Rh. graminis*, *Rh. glutinis* and *R. diobovatum* (section III.2, Appendix III). The *RHA2* gene was amplified in selected *S. salmonicolor* strains using primers (MC040 and MC073) that anchored on the *STE20* and *KAP95* flanking genes. The amplification products were subsequently sequenced by primer-walking. For PCR reactions, cycling conditions, primer sequences and GenBank accession numbers see sections III.1 and III.2 of Appendix III.

5.2.4. Sequence data and phylogenetic analyses

PCR amplification and sequencing of both D1/D2 and ITS regions was done as previously described (Gadanho and Sampaio, 2002). Sequences of both regions were concatenated and subsequently aligned using ClustalW 1.4 (Thompson et al., 1994) included in the BioEdit software (Hall, 1999). The phylogenetic tree was inferred by Maximum Likelihood (ML) with PhyML (Guindon and Gascuel, 2003). FindModel, a web implementation of ModelTest (Posada and Crandall, 1998; Guindon and Gascuel, 2003), was used with the Akaike information criterion (AIC) to select the model that best fit our data. The General Time Reversible model with a discrete gamma distribution was chosen (GTR+G; shape parameter = 0.1947). *Microbotryum lychnidis-dioicae* (GenBank: DQ366868/ AY877416) and *Leucosporidium scottii* (GenBank: AF0700419/ AF444495) were used to root the tree. For the *HD1/HD2* tree, nucleotide sequences were aligned with MUSCLE (Edgar, 2004) since it produced a better alignment for such highly divergent sequences. The phylogenetic tree was also inferred by ML based on the TN93+G+I model (shape = 2.0446; pinv = 0.0524). Protein sequences of the pheromone receptors (Ste3A1 and Ste3A2) were deduced from the DNA sequences after removal of putative introns, either manually or using AUGUSTUS software (Stanke et al., 2008). ProtTest (Abascal et al., 2005) with Akaike information criterion (AIC) was used to select the model that best fit our data and phylogenetic trees were inferred by ML. The MtREV+G+F model (shape parameter = 1.198) (Yang, 1993; Cao et al., 1994; Adachi and Hasegawa, 1996) was chosen for the

Ste3A1 tree and the WAG+G+F (shape parameter = 1.0496) (Yang, 1993; Cao et al., 1994; Whelan and Goldman, 2001) was selected for the Ste3A2 dataset. *Microbotryum lycnhidis-diocae* pr-MatA1 (GenBank: EF584742) and pr-MatA2 (GenBank: ABU62846) were used to root trees. Bootstrap values were calculated from 1000 replicates for all trees. GenBank accession numbers of the novel sequences are listed in section III.1 (Appendix III).

5.2.5. Estimation of the evolution rates of 5' end and homeodomain regions of the *HD1* gene

The region of the *HD1* gene corresponding to the 5'end and the initial part of the homeodomain motif was amplified and sequenced from several red yeast strains. Evolution rates were estimated by a Window Analysis of dN and dS, using the online interface of WINA 0.34 (Endo et al., 1996) in a sliding window (size = 3) along the alignment of each set of alleles. GenBank accession numbers and strains where *HD1* gene was amplified are listed in section III.1 (Appendix III).

5.3. Results and discussion

5.3.1. Updated phylogeny of the Sporidiobolales

A group of 43 red yeast species representing the range of phylogenetic diversity in the order Sporidiobolales was investigated (Hamamoto et al., 2011; Sampaio, 2011a; Sampaio, 2011c; Sampaio, 2011b). These species are classified in four genera, two sexual – *Rhodosporidium* and *Sporidiobolus* – and two asexual – *Rhodotorula* and *Sporobolomyces*. An updated molecular phylogeny of the Sporidiobolales based on the concatenated alignment of the complete ITS region (ITS1, 5.8S and ITS2) and the D1/D2 domain of the LSU rRNA is shown in Figure 5.1. Three statistically well-supported clades were observed. Although such organization was somewhat anticipated in previous studies (e.g. Sampaio et al., 2003; Sampaio, 2011a), the incorporation of more taxa, including species not yet formally described, and the combination of ITS and D1/D2 data, improved the resolution of the present analysis. Whereas clade A included only species of *Rhodosporidium* and *Rhodotorula*, which are characterized by the absence of forcibly discharged mitospores (ballistoconidia), clade C contained only species of *Sporidiobolus* and *Sporobolomyces* that are able to form ballistoconidia. At variance with clades A and C, clade B included representatives of the four genera, which indicates that the traditional classification separating genera based on the ability to form ballistoconidia is not supported at the molecular level. It is also evident from Figure 5.1 that in all three clades, sexual and asexual species are intermingled.

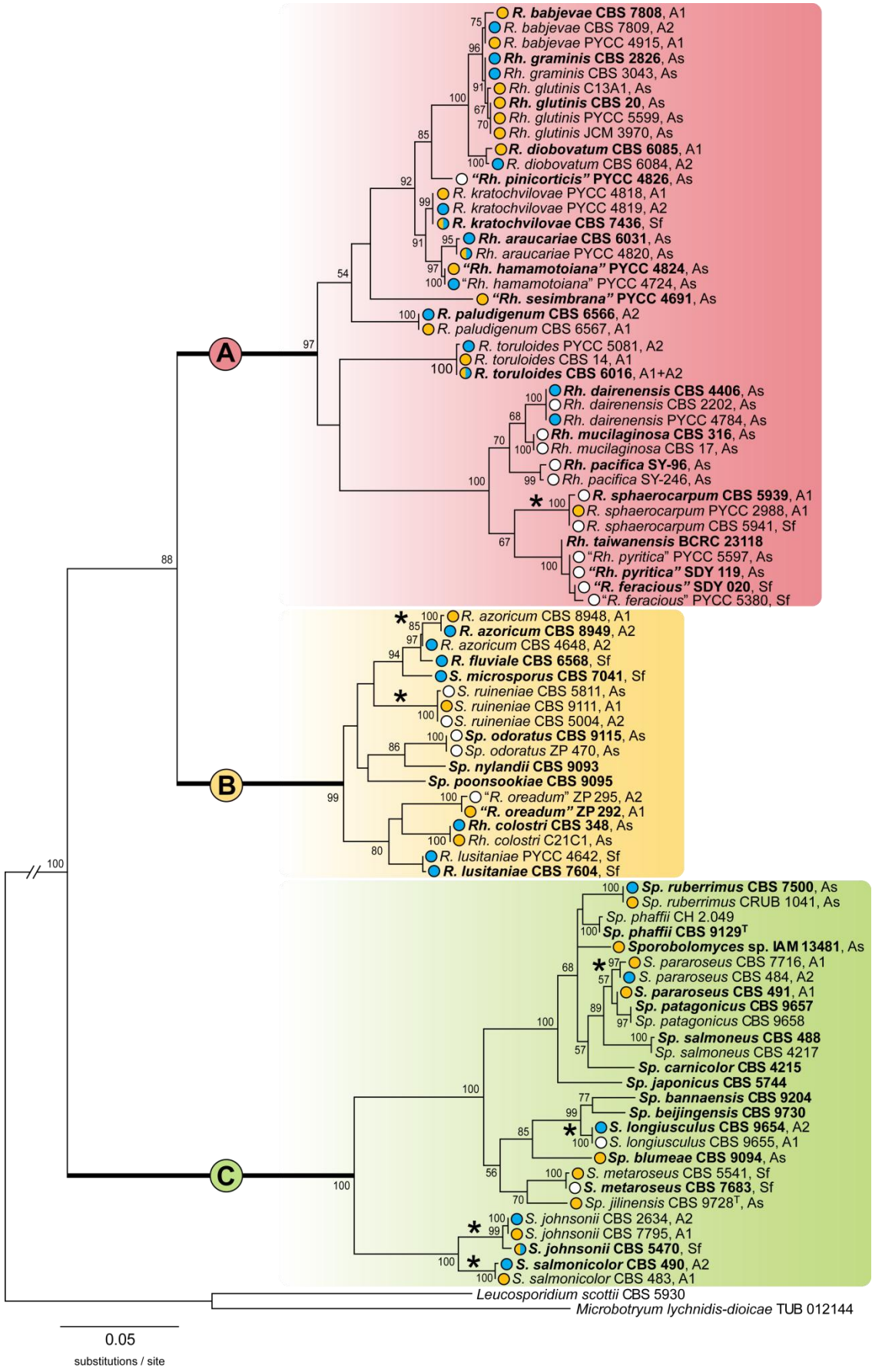
R. babjevae CBS 7808, A1
R. babjevae CBS 7809, A2
R. babjevae PYCC 4915, A1
Rh. graminis CBS 2826, As
Rh. graminis CBS 3043, As
Rh. glutinis C13A1, As
Rh. glutinis CBS 20, As
Rh. glutinis PYCC 5599, As
Rh. glutinis JCM 3970, As
R. diobovatum CBS 6085, A1
R. diobovatum CBS 6084, A2
"Rh. pinicorticis" PYCC 4826, As
R. kratochvilovae PYCC 4818, A1
R. kratochvilovae PYCC 4819, A2
R. kratochvilovae CBS 7436, Sf
Rh. araucariae CBS 6031, As
Rh. araucariae PYCC 4820, As
"Rh. hamamotoiana" PYCC 4824, As
"Rh. hamamotoiana" PYCC 4724, As
"Rh. sesimbrana" PYCC 4691, As
R. paludigenum CBS 6566, A2
R. paludigenum CBS 6567, A1
R. toruloides PYCC 5081, A2
R. toruloides CBS 14, A1
R. toruloides CBS 6016, A1+A2
Rh. dairenensis CBS 4406, As
Rh. dairenensis CBS 2202, As
Rh. dairenensis PYCC 4784, As
Rh. mucilaginosa CBS 316, As
Rh. mucilaginosa CBS 17, As
Rh. pacifica SY-96, As
Rh. pacifica SY-246, As
R. sphaerocarpum CBS 5939, A1
R. sphaerocarpum PYCC 2988, A1
R. sphaerocarpum CBS 5941, Sf
Rh. taiwanensis BCRC 23118
"Rh. pyritica" PYCC 5597, As
"Rh. pyritica" SDY 119, As
"R. feracious" SDY 020, Sf
"R. feracious" PYCC 5380, Sf
A
R. azoricum CBS 8948, A1
R. azoricum CBS 8949, A2
R. azoricum CBS 4648, A2
R. fluviale CBS 6568, Sf
S. microsporus CBS 7041, Sf
S. ruineniae CBS 5811, As
S. ruineniae CBS 9111, A1
S. ruineniae CBS 5004, A2
Sp. odoratus CBS 9115, As
Sp. odoratus ZP 470, As
Sp. nylandii CBS 9093
Sp. poonsookiae CBS 9095
"R. oreadum" ZP 295, A2
"R. oreadum" ZP 292, A1
Rh. colostri CBS 348, As
Rh. colostri C21C1, As
R. lusitaniae PYCC 4642, Sf
R. lusitaniae CBS 7604, Sf
B
Sp. ruberrimus CBS 7500, As
Sp. ruberrimus CRUB 1041, As
Sp. phaffii CH 2.049
Sp. phaffii CBS 9129T
Sporobolomyces sp. IAM 13481, As
S. pararoseus CBS 7716, A1
S. pararoseus CBS 484, A2
S. pararoseus CBS 491, A1
Sp. patagonicus CBS 9657
Sp. patagonicus CBS 9658
Sp. salmoneus CBS 488
Sp. salmoneus CBS 4217
Sp. carnicolor CBS 4215
Sp. japonicus CBS 5744
Sp. bannaensis CBS 9204
Sp. beijingensis CBS 9730
S. longiusculus CBS 9654, A2
S. longiusculus CBS 9655, A1
Sp. blumeae CBS 9094, As
S. metaroseus CBS 5541, Sf
S. metaroseus CBS 7683, Sf
Sp. jilinensis CBS 9728T, As
S. johnsonii CBS 2634, A2
S. johnsonii CBS 7795, A1
S. johnsonii CBS 5470, Sf
S. salmonicolor CBS 490, A2
S. salmonicolor CBS 483, A1
C
Leucosporidium scottii CBS 5930
Microbotryum lychnidis-dioicae TUB 012144
0.05
substitutions / site

Figure 5.1. (previous page) **Molecular phylogeny of Sporidiobolales based on a concatenated alignment of the ITS region (ITS1, 5.8S and ITS2) and the D1/D2 domain of the LSU rRNA.** Mating types (*MAT A1* and *MAT A2*), asexual and self-fertile strains are designated as A1, A2, As and Sf, respectively. Circles before each strain depict the type of pheromone receptor gene: yellow, *STE3.A1*; blue, *STE3.A2*; white, not detected; half-coloured circles, simultaneous detection of *STE3.A1* and *STE3.A2*. The type strain of each species is highlighted in boldface. A, B and C are three statistically well-supported clades referred in the text. Asterisks indicate instances where the original mating type designation was altered to match the molecular identity of the strains. Species not yet formally described are indicated by inverted commas. The tree was inferred by Maximum Likelihood (ML) using the GTR+G model (shape = 0.1947) and rooted with sequences of *Microbotryum lychnidis-dioicae* and *Leucosporidium scottii*. Bootstrap values (>50%) from 1000 replicates are shown. Abbreviations: *R.*, *Rhodosporidium*; *Rh.*, *Rhodotorula*, *S.*, *Sporidiobolus* and *Sp.*, *Sporobolomyces*.

5.3.2. Pheromone receptor genes as molecular markers of mating type identity

Multiple *HD1/HD2* alleles were found in red yeast species, and phylogenetic analyses suggested that allele diversification occurred only after each species was formed (Coelho et al., 2010). In addition, *HD1/HD2* genes of different red yeast species were found to be highly divergent precluding their utilization for mating type-screening purposes over a wide phylogenetic range. On the contrary, phylogenetic analyses of the pheromone receptor alleles (*STE3.A1* and *STE3.A2*) of both *M. violaceum* and *S. salmonicolor*, revealed ancient trans-specific polymorphisms (Devier et al., 2009; Coelho et al., 2010) as expected for genes maintained under long-term balancing selection (Muirhead et al., 2002). Since only two pheromone receptor alleles were present in each species examined so far, and the presence of each allele correlated in all cases with mating behaviour (Coelho et al., 2008; Coelho et al., 2010) (see Chapters 3 and 4), pheromone receptor genes can be advantageously used as markers for mating type at the molecular level across a broad range of red yeast species.

All heterothallic red yeast species within the order Sporidiobolales were originally described as having a bipolar mating behaviour. Such studies involved only standard crossing tests and strains were therefore arbitrarily assigned to either *MAT A1* or *MAT A2* within each species. Pheromone receptor genes were first identified in *R. toruloides* and were named *STE3.A1* and *STE3.A2* in accordance with the mating type of the strains in which they were initially characterized (Coelho et al., 2008) (see Chapter 3). After the identification of the sequences of the pheromone receptor genes in *S. salmonicolor* and *S. johnsonii*, it became clear that as a result of the original arbitrary assignment of mating types, each *STE3* allele was associated with strains of the opposite mating type, [e.g. *STE3.A1* was found in *S. salmonicolor* and *S. johnsonii MAT A2* strains, (Coelho et al., 2010)]. To avoid future confusion, the mating type designations of these two species were previously renamed to match the molecular data (Coelho et al., 2010). In this chapter, the current mating type designation of the heterothallic species included in the Sporidiobolales, were investigated at the molecular level. To that end, a PCR-based approach with degenerate primers was employed that

specifically amplified *STE3.A1* or *STE3.A2* alleles. In approximately half of the heterothallic species analysed (*R. sphaerocarpum*, *R. azoricum*, *S. ruineniae*, *S. pararoseus* and *S. longiusculus*) the incongruences previously observed for *S. salmonicolor* were also detected, and in these cases the mating type designations were altered so as to match the molecular identity of the strains. It should be noted that in *R. sphaerocarpum* and *S. ruineniae* only the *STE3.A1* allele could be identified (Figure 5.2). Nevertheless, the fact that the *STE3.A1* allele was only found in strains that mated as *MAT A1* suggested that the *STE3.A2* allele is most probably present in *MAT A2* strains, but the set of primers that were used failed to amplify it. For *R. babjevae*, 25 strains were investigated (see section III.1 of Appendix III). Previous mating studies had identified five strains as *MAT A1*, four strains as *MAT A2* and two strains as self-fertile (Gadanho and Sampaio, 2002; Sampaio, 2011a). The remaining 14 strains were found to be asexual. The molecular characterization of pheromone receptor alleles for the sexual strains confirmed the results obtained with standard crossing tests. It also showed that one of the self-fertile strains (VKM Y-2911) contained both receptor alleles and that 11 asexual strains belonged to *MAT A1* and three to *MAT A2*. Both mating types were present in strains isolated in Russia (six strains), Argentina (five strains) and Japan (two strains). However, the six strains isolated in Portugal belonged solely to *MAT A1*. Using tester strains with good mating abilities and five asexual strains for which the molecular determination of mating type had been obtained, several crosses were repeated and confirmed that these strains do not mate in culture.

5.3.3. Pheromone receptor genes in asexual species

Subsequently, the presence of pheromone receptor homologues (*STE3.A1* and *STE3.A2*) was investigated in 16 asexual species classified in the genera *Rhodotorula* and *Sporobolomyces* (Figure 5.1 and section III.1 of Appendix III). Amplification of *STE3.A1* and/or *STE3.A2* alleles was successful in 11 species among the 16 examined (Figure 5.1). When multiple strains were available for a species, both receptor alleles were normally detected, thus demonstrating the presence of the two mating types in natural populations of these presumably asexual species. However, an interesting exception was found for the sibling species *Rh. glutinis* and *Rh. graminis*. Whereas all six strains of *Rh. glutinis* possessed the *STE3.A1* allele, all four strains of *Rh. graminis* exhibited the *STE3.A2* allele (Figure 5.2). These two species are also closely related to the heterothallic species *R. babjevae*. In contrast to *R. babjevae*, only a limited number of strains of *Rh. graminis* and *Rh. glutinis* have been isolated worldwide (Sampaio, 2011b), all of which were included in this study. For these three species, inter-species mating had been tested, but was not observed (Gadanho and Sampaio, 2002). It is tempting to speculate that asexuality in *Rh. graminis* and *Rh. glutinis* might be related to a strong but relatively recent mating type imbalance. Nevertheless, a larger sampling of both species would be required to accurately determine the mating type distribution and frequency and to elucidate if sex occurs in nature by looking for footprints of recombination, as done previously for other asexual fungi (Taylor et al., 1999).

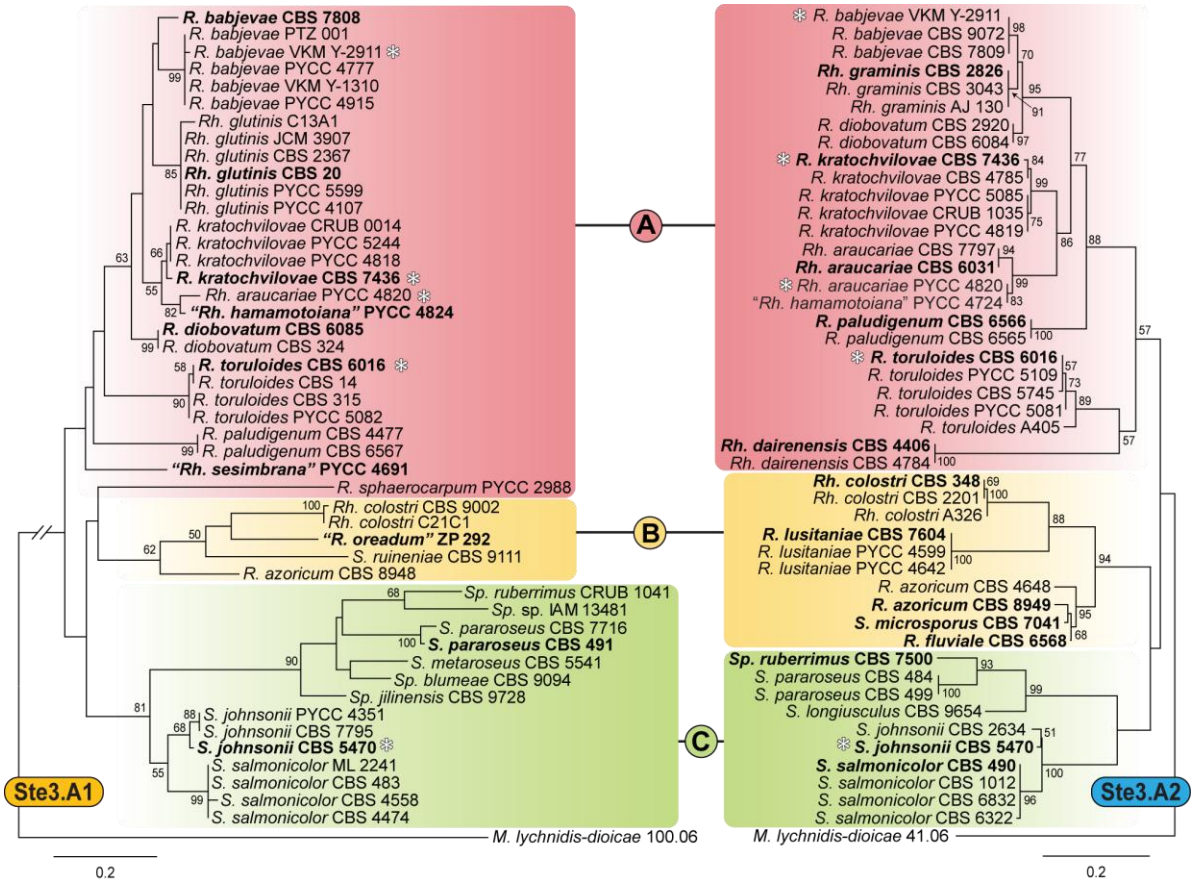

Figure 5.2. Phylogeny of the pheromone receptors of several red yeast species. Phylogeny based on the amino acid sequence of the *MAT A1* and *MAT A2* pheromone receptors (Ste3.A1 and Ste3.A2, respectively). Groups A, B and C are the same as in Figure 5.1. Strains marked with asterisks have both pheromone receptors. *MAT A1* and *A2* receptor sequences from *Microbotryum lychnidis-dioicae* were used as outgroups. Trees inferred by ML using MtREV+G+F (shape = 1.198; Ste3.A1) and WAG+G+F (shape = 1.049; Ste3.A2) models. Bootstrap values (>50%) from 1000 replicates are shown. See Figure 5.1 for additional details.

Next, the integrity of pheromone receptor genes was investigated in asexual species. To this end, the complete genes were sequenced. Interestingly, none of them showed hallmarks of degradation and a protein sequence could be readily obtained by conceptual translation in all cases, suggesting that these genes encode functional proteins. Moreover, the pheromone receptor genes in the recently released genomes of the presumptively asexual species *Sporobolomyces* sp. IAM 13481 (*STE3.A1*) and *Rh. graminis* WP1 (*STE3.A2*) (formerly designated as *R. babjevae*; see 5.2.2.), available at the Joint Genome Institute (JGI) were also inspected. Again, intact and complete sequences of the pheromone receptor genes were identified. Therefore, these results are consistent with other studies that have shown that in asexual ascomycetes mating type genes are present and apparently functional (O'Donnell et al., 2004; López-Villavicencio et al., 2010).

In contrast to these observations, for five asexual species (*Rh. mucilaginosa*, *Rh. pacifica*, *"Rh. pinicorticis"*, *"Rh. pyritica"* and *Sp. odoratus*) no pheromone receptor alleles were found (Figure 5.1

and section III.1 of Appendix III). With the exception of *"Rh. pinicorticis"* and *Sp. odoratus*, the remaining species are closely related to each other and belong to a lineage of clade A that includes also *R. sphaerocarpum* and *Rh. dairenensis*. The phylogenetic analysis of *STE3.A1* from *R. sphaerocarpum* and *STE3.A2* from *Rh. dairenensis*, suggested that the amino acid sequences of these pheromone receptor genes are unusually divergent (Figures 5.1 and 5.2). Therefore, it is likely that the lack of amplification in the remaining species may be due to mismatches between the primer and target sequences. A similar explanation might apply to the amplification of *R. sphaerocarpum STE3.A1* in only one out of six *MAT A1* strains investigated.

For *Rh. araucariae* PYCC 4820, two pheromone receptor alleles were detected (Figures 5.1 and 5.2). However, the nucleotide sequence of the *STE3.A2* allele in this strain differed from homologous sequences of conspecific strains of the same mating type (CBS 6031 and CBS 7797; 87% identity). Surprisingly, the *STE3.A2* allele of *Rh. araucariae* PYCC 4820 was found to be 100% identical to the *STE3.A2* gene of *"Rh. hamamotoiana"* PYCC 4724. Interestingly, and contrary to the other strains of *Rh. araucariae*, cells of strain PYCC 4820 formed a slender projection that resembled a conjugation tube. However, actual conjugation between cells of the same strain or with cells of the other strains of *Rh. araucariae* was not observed. It seems plausible that the phenotype of PYCC 4820 is due to the presence of both *MAT* alleles (including the pheromone receptors), as a result of a past hybridization event between the two species. However, incompatibilities at other loci, like for example the homeodomain locus, are perhaps impairing progression through the sexual cycle, as no dikaryotic mycelium or teliospores were observed.

5.3.4. Pheromone receptor genes in self-fertile species

In homothallic species, a single strain can complete the life cycle without the need for a complementary mating partner. Although several mechanisms of homothallism have been put forward in fungi (Lin and Heitman, 2007), the most common explanation for this sexual behaviour is that strains of these species have both *MAT* alleles in a single genome and can therefore form sexual structures alone (see 1.2.3.1). Here, the presence of both pheromone receptor alleles was explored in three species from clade B, which contains only self-fertile strains (*R. fluviale*, *R. lusitaniae* and *S. microsporus*), but only *STE3.A2* alleles were found. Although the absence of amplification of *STE3.A1* alleles due to mismatches between primer and target sequences could not be ruled out, this explanation is less likely in this case, since *STE3.A1* alleles were identified and sequenced in closely related species of clade B, like *R. azoricum* and *Rh. colostri*. Therefore, it is possible that for these three species self-fertility could be due to compatibility at the homeodomain locus despite homozygosity at the pheromone receptor locus. If confirmed, such a situation would parallel that found in the corn smut pathogen *Ustilago maydis*, in which diploid strains that are homozygous at the pheromone receptor locus but heterozygous at the homeodomain locus are able to produce teliospores that germinate and undergo meiosis in a way that is indistinguishable from

that of the heterozygous diploids at both *MAT* loci (Banuett and Herskowitz, 1989). Nevertheless, whereas these diploid strains of *U. maydis* only produce incipient hyphae because maintenance of filamentous growth also requires heterozygosity at the pheromone receptor locus (Banuett and Herskowitz, 1989), for the self-fertile yeast species studied extensive hyphal growth was found. Some heterothallic species of red yeasts also include a few self-fertile strains (viz. *R. kratochvilovae* CBS 7436, *R. babjevae* VKM Y-2911 and *S. johnsonii* CBS 5470). In most of those cases, it was found that both pheromone receptor alleles were present (section III.1, Appendix III) suggesting that self-fertility in red yeasts can be attained by different molecular mechanisms.

5.3.5. Phylogeny of the pheromone receptors

Using partial sequences of *STE3.A1* and *STE3.A2* alleles, the evolutionary history of the two pheromone receptors was compared, whose trans-specific phylogenies indicate that their divergence dates back to the origin of sexuality in basidiomycetes (Devier et al., 2009; Coelho et al., 2010; Devier et al., 2010). Phylogenies based on the protein sequences of these genes were constructed separately for each receptor (Ste3.A1 and Ste3.A2) in order to maximize the quality of the alignments (Figure 5.2). In spite of a general concordance between rDNA and pheromone receptor trees, three conflicts were detected in the Ste3.A1 tree. These cases correspond to the placement of *R. diobovatum*, *R. paludigenum* and *R. sphaerocarpum* outside their respective groups, by comparison with rDNA and Ste3.A2 trees (Figures 5.1 and 5.2). However, it should be noted that bootstrap values in the Ste3.A1 tree are lower than those obtained for the Ste3.A2 tree and that therefore these conflicts might be due to poor resolution in the Ste3.A1 tree.

5.3.6. Structure of the *MAT A1* and *MAT A2* genomic regions encompassing the alternate pheromone receptors

As previously observed in a restricted number of red yeast species (Chapters 3 and 4), genomic regions around the pheromone receptor locus exhibited synteny breaks, gene inversions and sequence divergence between the two mating types as a result of apparent partial suppression of recombination. On the other hand, synteny was well conserved across species when *MAT* regions of the same mating type were compared (Coelho et al., 2008; Coelho et al., 2010) (see Chapters 3 and 4). Therefore, it was relevant to assess whether or not these observations also held true for the expanded set of red yeast species under scrutiny in this Chapter. Taking advantage of whole-genome data of *Sporobolomyces* sp. IAM 13481 (*MAT A1*) and *Rh. graminis* WP1 (*MAT A2*), the synteny around the pheromone receptor genes in red yeast species belonging to clades A, B and C was assessed (Figure 5.1 and 5.3). PCR primers anchored on the 3' end region of the genes

flanking the *STE3.A1* (*RibL18ae* and *RibL6*/ *RNAPOL*) and *STE3.A2* (*LSm7* and *RibL6*/ *RNAPOL*) yielded almost always an amplification product between 2.5 to 3.0 kb, consistent with a gene organization identical to that found in the genomes of *Sporobolomyces* sp. IAM 13481 and *Rh. graminis* WP1 (Figure 5.4). Sequencing of the complete intervening regions confirmed the suspected synteny within the same mating type across the majority of the species (Figure 5.3). However, for some species, like *MAT A1* strains of *R. paludigenum*, gene order seemed to have been disrupted since no amplification products were obtained even when using different combinations of primers for the more conserved flanking genes. Another interesting exception was *S. ruineniae* CBS 5811, where a *MAT A2* pheromone precursor homologue similar to the *RHA2.A2* of *R. toruloides* (Coelho et al., 2008) was found between *RibL18ae* and *RibL6* instead of the *STE3.A1* gene usually present at this location (Figure 5.3). This indicates that the *STE3.A1* gene has been relocated to a different and presently unknown genomic location in this species.

Next, a larger genomic region of the *MAT A1* and *MAT A2* loci was examined. To do so, 50 kb regions encompassing the *STE3.A1* gene in *Sporobolomyces* sp. IAM 13481 and the *STE3.A2* gene in *Rh. graminis* WP1 were compared (Figure 5.4). In *Rh. graminis* WP1, two identical putative pheromone precursor homologues (Rg*RHA1.A2* and Rg*RHA2.A2*) were identified and were similar but not syntenic to the *RHA2.A2* gene of *R. toruloides MAT A2* strains (Figure 5.4) (Coelho et al., 2008). The regions upstream of Rg*RHA2.A1* and Rg*RHA2.A2* were identical. A shorter sequence of approximately 280 bp was also identical and was detected downstream of *RNAPOL* in an inverted orientation (Figure 5.4; see asterisks). This suggests that an additional copy of the pheromone precursor gene may have existed at this genomic location similarly to that found in the *MAT A1*. Notably, *MAT A1* pheromone genes in *S. salmonicolor* and *R. toruloides* were found to be embedded in conserved inverted repeats (Coelho et al., 2008), a situation earlier documented in the basidiomycete human pathogen *Cryptococcus neoformans* (Fraser et al., 2004). These repeats are likely maintained by intra-allelic gene conversion to ensure maintenance of gene function in face of the absence of meiotic recombination characteristic of these regions (Rozen et al., 2003; Chen et al., 2007).

Additional comparison between both *MAT* genomic regions showed a similar gene content despite the phylogenetic distance between *Sporobolomyces* sp. and *Rh. graminis* (Figures 5.1 and 5.4) and that apparently functional copies of each allele have been retained in both *MAT* loci. This is in contrast to genes encoded in sex chromosomes of diploid organisms, for which degeneration and loss of one functional copy frequently occurs (Charlesworth et al., 2005). Maintenance of both alleles in fungal sex-determining regions, including the *MAT* loci of red yeasts, is most probably related to the fact that fungi commonly occur as haploids in the environment, and thus gene degeneration or loss is very likely to be detrimental. However, gene order seems to have been deeply altered during mating type divergence and many gene blocks were found in inverted positions, probably restraining recombination between opposite *MAT* regions (Figure 5.4). On the contrary, within the same mating type, the analysed regions were found to be highly syntenic even between more distantly related species (Figure 5.4).

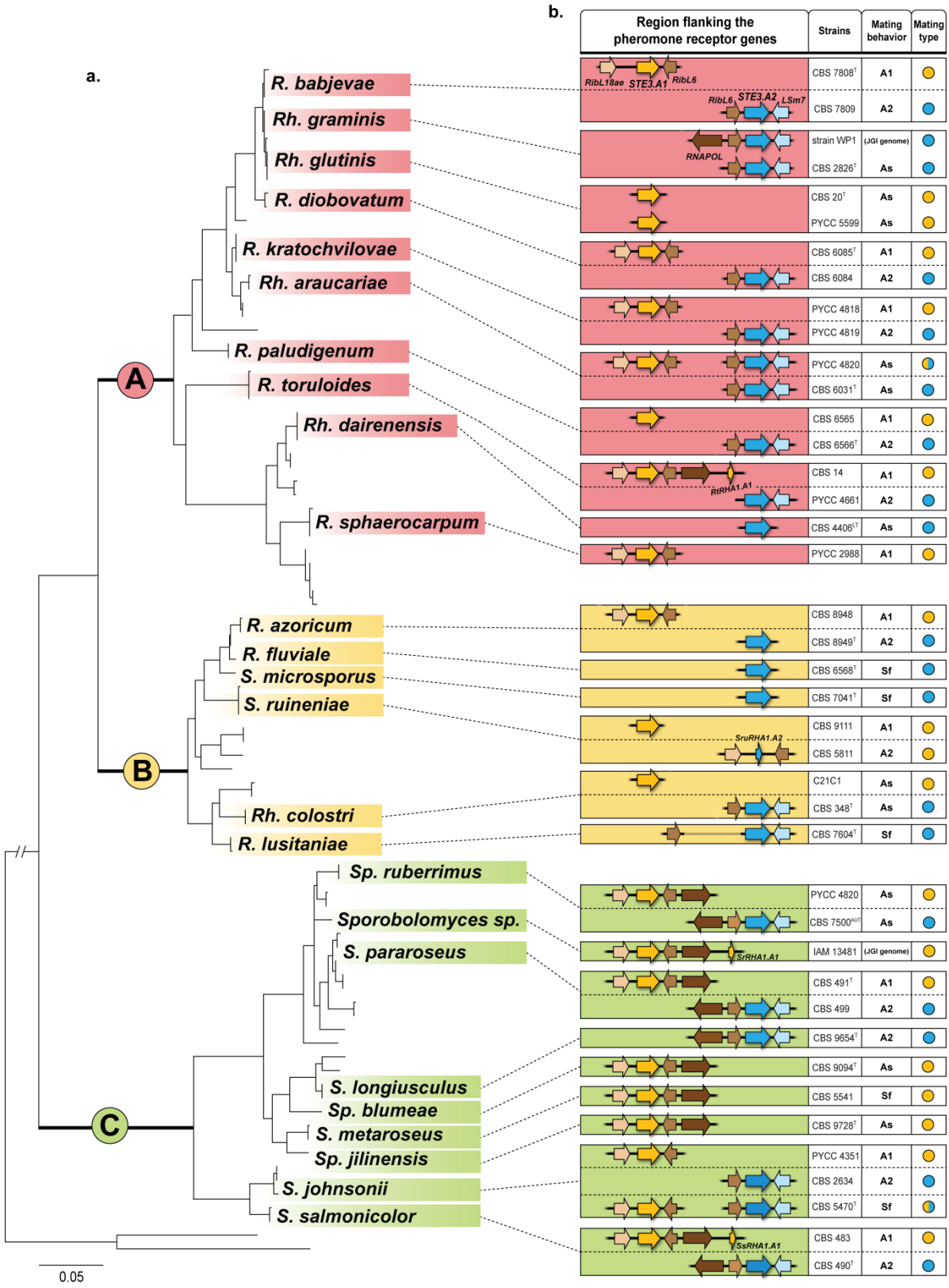
a.
b.
Region flanking the pheromone receptor genes
Strains
Mating behavior
Mating type
R. babjevae
Rh. graminis
Rh. glutinis
R. diobovatum
R. kratochvilovae
Rh. araucariae
R. paludigenum
R. toruloides
Rh. dairenensis
R. sphaerocarpum
R. azoricum
R. fluviale
S. microsporus
S. ruineniae
Rh. colostri
R. lusitaniae
Sp. ruberrimus
Sporobolomyces sp.
S. pararoseus
S. longiusculus
Sp. blumeae
S. metaroseus
Sp. jilinensis
S. johnsonii
S. salmonicolor
A
B
C
RibL18ae
STE3.A1
RibL6
STE3.A2
LSm7
RNAPOL
RtRHA1.A1
SruRHA1.A2
SrRHA1.A1
SsRHA1.A1
CBS 7808T
CBS 7809
strain WP1
CBS 2826T
CBS 20T
PYCC 5599
CBS 6085T
CBS 6084
PYCC 4818
PYCC 4819
PYCC 4820
CBS 6031T
CBS 6565
CBS 6566T
CBS 14
PYCC 4661
CBS 4406LT
PYCC 2988
CBS 8948
CBS 8949T
CBS 6568T
CBS 7041T
CBS 9111
CBS 5811
C21C1
CBS 348T
CBS 7604T
PYCC 4820
CBS 7500NUT
IAM 13481
CBS 491T
CBS 499
CBS 9654T
CBS 9094T
CBS 5541
CBS 9728T
PYCC 4351
CBS 2634
CBS 5470T
CBS 483
CBS 490T
A1
A2
As
Sf
(JGI genome)
0.05

Figure 5.3. (previous page) **Synteny of the genomic regions flanking the alternate pheromone receptors in *MAT A1* and *MAT A2* strains of several red yeasts species. (a)** Simplified illustration of the tree represented in Figure 5.1, indicating the phylogenetic placement of the species where gene organization in the vicinity of the pheromone receptor genes was determined. Clades A, B and C are the same as in Figure 5.1 and 5.2. **(b)** The mating behaviour (A1, A2, As and Sf stands for mating type A1, A2, asexual and self-fertile, respectively), the molecular mating type (*STE3.A1*, yellow circles; *STE3.A2*, dark blue circles; *STE3.A1* and *STE3.A2*, half-coloured circles) and the obtained genomic regions flanking the pheromone receptor alleles (*STE3.A1* and *STE3.A2*) are shown for each strain. Orthologues are shown in the same colour. In *Rhodosporidium lusitaniae*, the intervening region between the *STE3.A2* and the *RibL6* genes (faint line) was not sequenced. Abbreviations of generic names are as in Figure 5.1 (for other features see Figure 5.4).

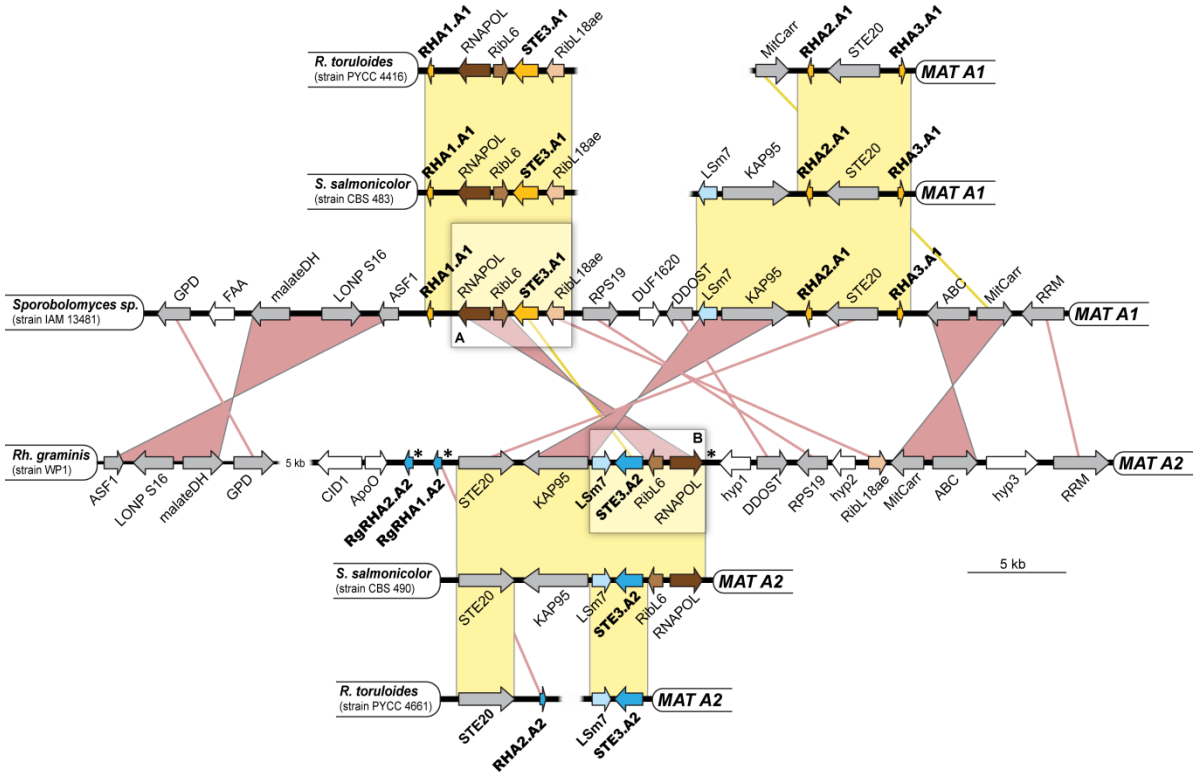

Figure 5.4. Synteny between *MAT A1* and *MAT A2* pheromone receptor loci in different red yeast species. The structure of the *MAT A1* and *MAT A2* pheromone receptor loci (~50 kb) is shown for *Sporobolomyces* sp. IAM 13481 and *Rh. graminis* WP1 and compared with shorter homologous regions of *S. salmonicolor* and *R. toruloides*. *MAT A1* and *MAT A2*-specific genes are depicted in yellow and dark blue arrows, respectively, showing the direction of transcription. Genes enclosed in box A are flanking the *STE3.A1* gene in most members of Sporidiobolales and those in box B are usually flanking the *STE3.A2* gene. Orthologous genes in boxes A and B have the same colour. Genes shared by *Sporobolomyces* sp. and *Rh. graminis* are shown in gray, while others are shown in white. Yellow lines or bars connect, respectively, genes or syntenic blocks that are in the same orientation, while pink lines or bars indicate inversions. Asterisks in the *MAT* region of *Rh. graminis* indicate three identical repeats of ~280 bp (the repeat located downstream of the *RNAPOL* gene is inverted in relation to the other two).

5.3.7. Diversity and evolution of the homeodomain transcription factors

As previously observed, the *HD1/HD2* alleles of some red yeasts are much more divergent than the correspondent pheromone receptor genes (Coelho et al., 2010) (see Chapter 4), and therefore less suitable for comparisons involving more distantly related species. Nevertheless, it was important to ascertain if the genes encoding these transcriptional regulatory factors appear to be functional within a more expanded group of red yeasts including sexual and asexual strains and species, and to evaluate their pattern of evolution. To this end, partial sequences of *HD1/HD2* alleles were obtained for four additional species (six in the total) and their phylogeny is shown in Figure 5.5. In clade A, formed by *S. johnsonii* and *S. salmonicolor*, the topology of the *HD1/HD2* tree is not congruent with the rDNA (Figure 5.1) and Ste3 (Figure 5.2) trees because *S. johnsonii* CBS 2634 harbours an *HD1/HD2 S. salmonicolor*-type allele and the converse situation occurs for *S. salmonicolor* strains CBS 1012, CBS 4474 and IHMT 2446/96 (Figure 5.5). Such discrepancies suggest that the two species are not fully isolated genetically and that mating with exchange of *HD1/HD2* alleles is still possible. In clade B, the lack of strict species-specificity of *HD1/HD2* alleles is also evident for the complex *R. babjevae – Rh. glutinis – Rh. graminis*, but not for *R. diobovatum*. In this clade, instances of interspecies hybridizations seem to resolve into viable and often sexually competent haploid individuals that retain evidence of gene flow between the species. The observation that pheromone precursor genes seem to encode very similar or even identical peptides in closely related species, possibly precluding discrimination between species prior to cell fusion is also in line with this. For example, only synonymous substitutions were detected when analyzing the *RHA2.A1* genes encoding the tandem pheromone peptides from different strains of *S. salmonicolor* and *S. johnsonii* (Figure 5.6). The outcome is the production of identical peptide pheromones, a circumstance that is likely to facilitate interspecies mating.

Judging from available data, a geographic structure for *HD1/HD2* allele distribution seems to be lacking since strains of *Rh. glutinis* from different regions have the same *HD1/HD2* alleles and in *Rh. graminis* different alleles were recovered from the same region (Figure 5.5 and section III.1 of Appendix III). For most of the species studied, each strain had a unique *HD1/HD2* allele although in *Rh. glutinis* four strains shared the same *HD1/HD2* allele. For two strains of *R. babjevae* (VKM Y-1310 and CBS 9072) it seems that a recent recombination event between the two *MAT* regions gave rise to strains with different pheromone receptor alleles that share the same *HD1/HD2* allele (Figure 5.5), consistent with the pseudo-bipolar system proposed in Chapter 4 (Coelho et al., 2010). Although the phylogenies and diversity of homeodomain alleles in *S. salmonicolor* and *S. johnsonii* suggested that such events were likely to occur albeit at low frequency, they had been thus far detected only in laboratory crosses (Coelho et al., 2010) (see Chapter 4). Allele divergence as inferred by branch length seems to be lower for the species in clade A, solely composed of sexually competent strains, than in clade B, where asexual strains predominate (Figure 5.5).

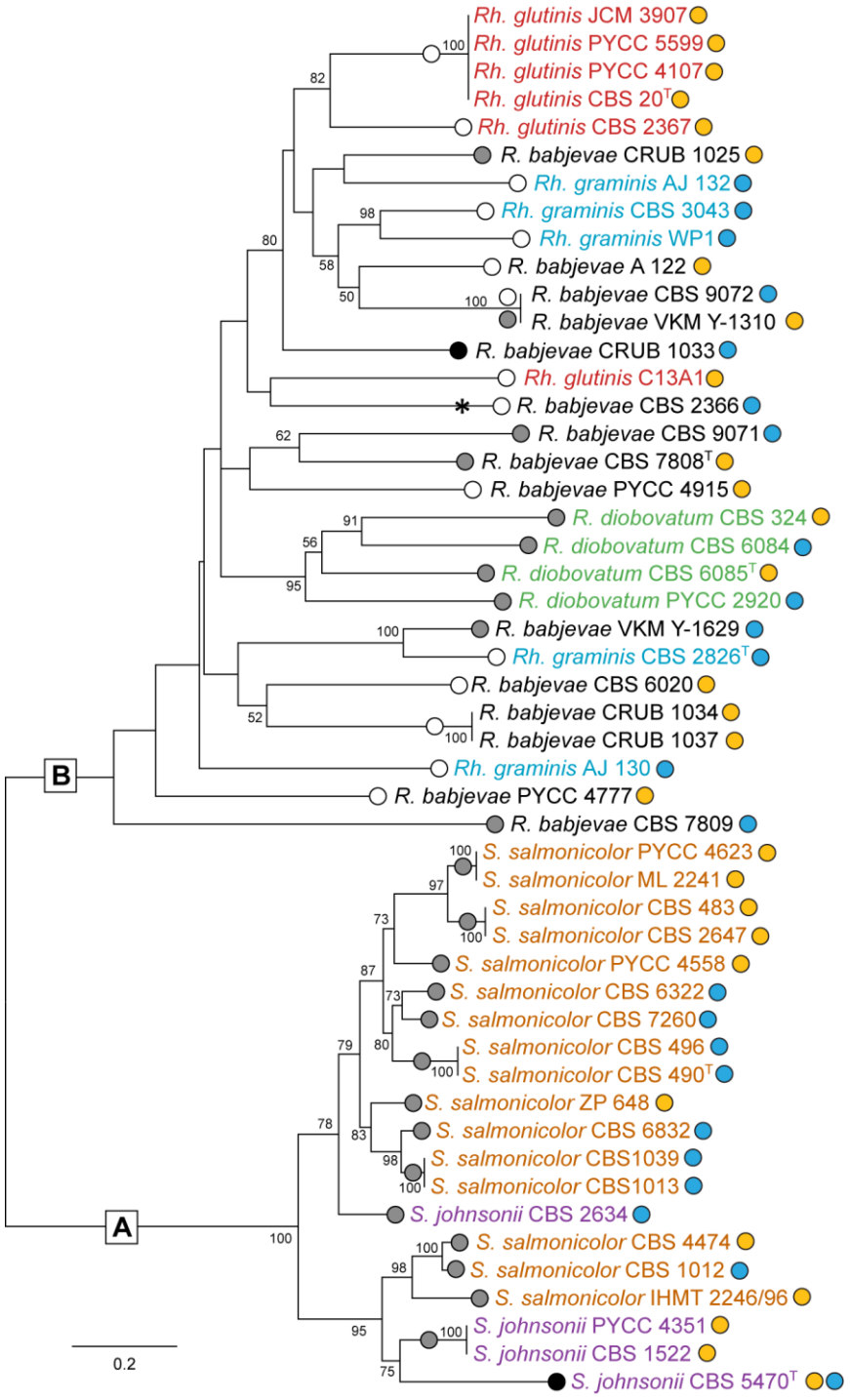

Figure 5.5. Diversity and phylogeny of *HD1/HD2* alleles in several red yeast species. Sexual, asexual and self-fertile strains are indicated by gray, white and black circles, respectively, before strain numbers. Circles after each strain depict the type of pheromone receptor genes (yellow, *STE3.A1*; blue, *STE3.A2*). Asterisk indicates a STOP codon in the *HD2* allele of strain CBS 2366. Branch lengths are given in number of substitutions per site. Bootstrap values (>50%) from 1000 replicates are shown. Clade A – *S. salmonicolor* (orange) and *S. johnsonii* (purple); clade B – *Rh. glutinis* (red), *Rh. graminis* (blue), *R. babjevae* (black), *R. diobovatum* (green). Unrooted tree inferred by ML using the TN93+G+I model (shape = 2.0446; pinv = 0.0524).

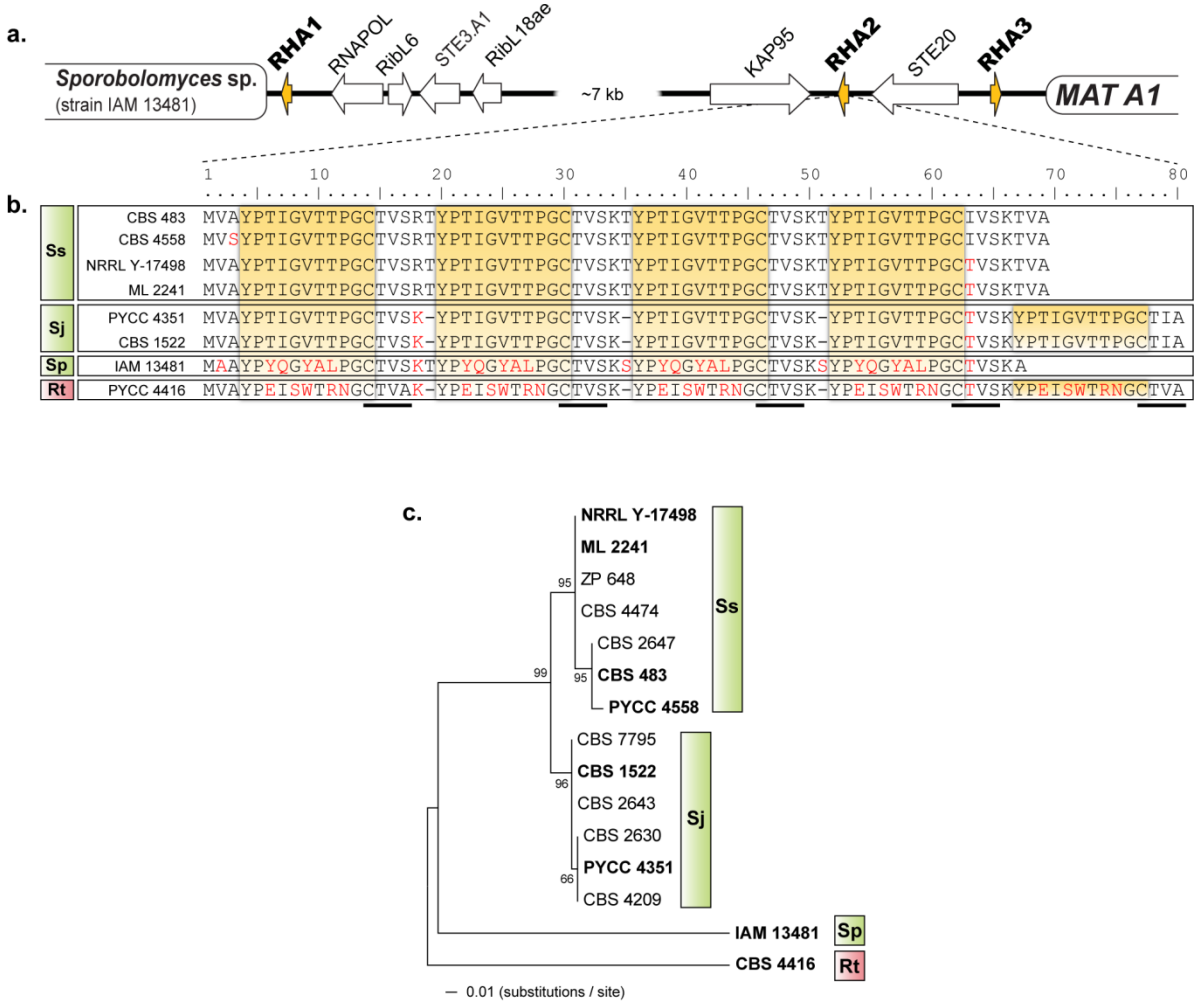

Figure 5.6. Pheromone precursors of different red yeast species. (a) The organization of the genomic region encompassing the *MAT A1* pheromone precursor genes in *Sporobolomyces* sp. IAM 13481 is shown on top. Coding regions of the pheromone precursor genes (*RHA1*, *RHA2* and *RHA3*) are depicted by yellow arrows, indicating the direction of transcription. **(b)** Alignment of the Rha2 pheromone precursor of different *MAT A1* red yeast species/ strains (Ss, *Sporidiobolus salmonicolor*; Sj, *Sporidiobolus johnsonii*; Sp, *Sporobolomyces* sp. IAM 13481; Rt, *Rhodosporidium toruloides*). Amino acids differing from the *S. salmonicolor* strain CBS 483 are shown in red. Sequence repeats proposed to represent the peptide moiety of the mature pheromone are shadowed and those resembling the CAAX motif are underlined. **(c)** Phylogenetic tree showing the relationships between *RHA2* genes from the indicated red yeast species, based on the alignment of their coding sequences. Groups are the same as in (b). Sequences of strains depicted in boldface are shown in (b). The tree was inferred using Maximum Parsimony. Bootstrap values from 1000 replicates are shown in the tree nodes.

Then, the dN/dS ratio of the N-terminal domain of *HD1* alleles, which is known to define mating specificity, as well as the initial part of the homeodomain motif, was evaluated. This comparison involved sexual and asexual species and sexual and asexual strains within a species (*R. babjevae*) and is shown in Figure 5.7. Stretches of codons under diversifying selection (dN/dS > 1) were found within the regions determining *MAT* specificity for the three groups of sexual strains (*S. salmonicolor, R. diobovatum* and *R. babjevae*), in accordance with previous observations (Coelho et al., 2010). Moreover, in these strains the homeodomain motif exhibited low dN/dS ratios indicative of purifying selection, whereas asexual strains had unusually high dN/dS ratios (>1 in strains of *Rh. glutinis* and asexual strains of *R. babjevae*). Finally, with the exception of *Rh. glutinis*, asexual strains did not exhibit evidence of diversifying selection in the *MAT* specificity domain (Figure 5.7). Since in functional HD1/HD2 transcription factors the homeodomain motif is highly conserved and evolving under purifying selection (Badrane and May, 1999), our observations might be indicative of relaxing selection as a consequence of asexuality both in *Rh. glutinis* and in the asexual subset of strains of *R. babjevae*. In line with this, a STOP codon was found in the *HD2* gene of *Rh. graminis* CBS 2366. Nevertheless, the average dN/dS along the N-terminal region in sexual and asexual red yeasts was <1, as also observed in other systems such as the *S*-loci of plants (Clark and Kao, 1991; Hinata et al., 1995). It seems possible that sequence divergence of the *HD1/HD2* genes examined here is governed by both diversifying and relaxing selection, depending on the maintenance or loss of sexual competence by individual strains.

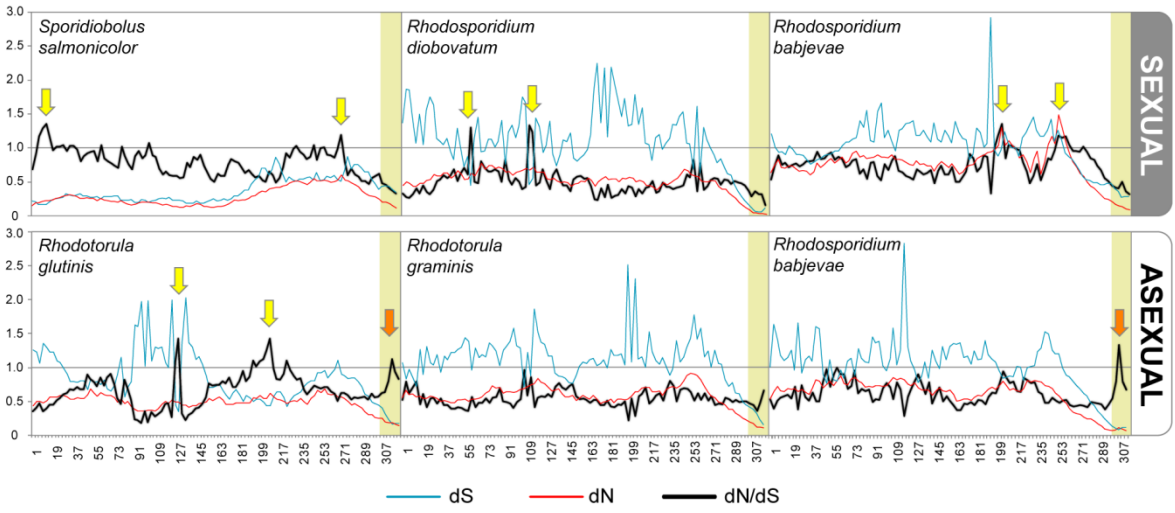

Figure 5.7. Evolution rate of the 5' end and homeodomain regions of the *HD1* gene in sexual and asexual red yeast strains. In each graph, synonymous (dS), non-synonymous (dN) and dN/dS values are shown along the analysed *HD1* coding region (Y-axis – evolution rate; X-axis – base pairs) and for all pairwise comparison between selected alleles of the *HD1* genes (see 5.2.5). Sites in the region corresponding to the variable domain of the HD1 protein where dN/dS is higher than 1.0 are indicated by yellow arrows, while those occurring in the homeodomain region (shadowed section) are indicated by orange arrows. Graphs on the top refer to alleles from sexual strains whereas those on the bottom refer to asexual strains.

These results further suggest that loss of sex might not occur uniformly across the Sporidiobolales. Different molecular evolution patterns may be related with distinct selective constraints acting on different species. *S. salmonicolor* and *S. johnsonii* exemplify a case of strictly sexual red yeasts whose *HD1/HD2* alleles exhibit lower dS values than the species in clade B, possibly due to some form of constraint on synonymous substitutions like codon usage bias (Whittle et al., 2011a; Whittle et al., 2011b) (Figures 5.5 and 5.7).

5.4. Final remarks

After examining 216 red yeast strains belonging to 32 species of the Sporidiobolales, apparently functional pheromone receptor homologues were found for most strains, irrespective of them being sexual or asexual. In addition, in all but one case, apparently intact *HD1/HD2* homologues were detected in the subset of strains that were investigated. Moreover, all asexual species were found to have phylogenetic close relatives that mate well in laboratory conditions, which further suggest that asexual red yeasts may have lost the ability to cross, although this has not (yet) resulted in the degeneration or loss of *MAT* genes.

Taken together, the results presented in this chapter suggest a fuzzy separation between sexual and asexual red yeast species, illustrated by the existence of asexual strains within sexual species and by the maintenance of *MAT* regions in asexual species. It looks like that the extant variations of the dominant heterothallic sexual reproductive mechanisms in red yeasts could promote the emergence of asexual lineages. For example, *Rh. glutinis* and *Rh. graminis* seem to derive from populations of a heterothallic *R. babjevae*-like ancestor that have independently lost the ability to mate. Notably, sexual reproduction in the latter species seems to be lost easily, as many of the *R. babjevae* strains here investigated were unable to mate.

Several authors have argued that most asexual fungi are evolutionary dead ends awaiting extinction. This fits well with the observations found for red yeasts since no evidence was found among extant species for ancient events of loss of sexuality. Instead, abundant substantiation for frequent and recent such events was uncovered, leading in some cases to the isolation of lineages and eventually to speciation but never to the generation of exclusively asexual phylogenetic lineages that persist for a long time.

6. CHAPTER

Concluding remarks and Future directions

Fungi have proved to constitute an exceptional system for understanding the genetic basis and evolution of how organisms intercommunicate and engage in sex. The reasons for this are three-fold: (i) all fungal lineages harbour a wide diversity of modes of sexual reproduction and lifestyles; (ii) due to their relatively small genome size, fungi are good candidates for genome sequencing projects and consequently, the complete sequence of several fungal genomes are presently available and (iii) their phylogenetic proximity with the lineage that includes animals (Metazoa) enables the study of broad principles that are shared by the two groups. Nevertheless, and in spite of numerous studies in several groups of fungi that have brought to light how sexual identity is determined at the molecular level, for other (equally important) fungal lineages this information is still lacking. To get more insight into this, two groups of basidiomycete fungi were studied in the course of this work: the red yeasts in the Pucciniomycotina and the *Ustilago*/*Pseudozyma* complex in the Ustilaginomycotina. While the first group was found to include a new form of sexual compatibility mechanism, the latter group allowed exploring some of the associations between the sexual and asexual modes of reproduction in connection with the parasitic or saprobic lifestyles.

6.1. Lifestyle and mode of reproduction in *Ustilago*/*Pseudozyma*

In Chapter 2, the sexual cycle of *Pseudozyma prolifica*, a species previously considered asexual, was disclosed after crossing the type strain of this species with compatible strains of *Ustilago maydis* that resulted in disease in maize plants. Therefore, *P. prolifica* has to be viewed just the same as *U.* maydis. The consequences of this finding are several. First, it confirms that *U. maydis* has an active saprobic stage that can be found in natural substrates other than maize plants (the studied strain of *P. prolifica* was found on debris of *Scirpus*). Secondly, from a taxonomic point of view, *P. prolifica* is redundant and should be merged with *U. maydis*. This finding highlighted the possibility that other species of *Pseudozyma* could represent the haploid (asexual) stage of pathogenic and sexual smuts for which no sexual stage has yet been identified. This relationship was investigated for several *Pseudozyma* species and for most of them, no evidence of loss of sexual reproduction was observed since they retained apparently functional homologues of *PRF1* (a key regulator of sexual reproduction in *U. maydis*). However, for *Pseudozyma aphidis* and other members of the same clade, which is mainly composed of asexual species frequently found in habitats different from those of their plant pathogen relatives, a different situation may be emerging. For this group, the molecular analysis of Prf1 revealed two novelties: an insertion of 20 amino acids and abnormally high level of expression of *PRF1* exhibited by the type strain of *Pseudozyma aphidis*. Although it is clear that further studies in *P. aphidis* and in other species of this group are needed to understand which factors may have contributed to the conservation of genes intimately related with sexual development and pathogenicity in the apparent absence of a parasitic and sexual lifestyle, a tentative hypothesis is that the shift to a different life strategy precipitated *PRF1* into a

new functional role. Studies addressing the presence and functionality of mating type genes (e.g. pheromone/ receptor and homeodomain transcription factor genes) will be needed in the future to characterize the mating system and mating capability of these species. In this context, it will be interesting to test the functionality of *MAT*-related genes of *Pseudozyma* by heterologous expression in *U. maydis* deletion mutants. Likewise, whole-genome data of some *Pseudozyma* species would provide an important tool to test the evolutionary hypothesis here proposed. It would also permit the search for genes that are considered key markers for necrotrophic fungi [e.g. polysaccharide hydrolases and pectin esterases (Dean et al., 2005)], which use such enzymes to degrade living and dead plant material. A deviation from the normal set of these genes in asexual species could further suggest a shift from a parasitic to a saprobic lifestyle. The same line of thought could be applied when investigating the presence of secreted protein effectors, which have been recently characterized as virulence-associated genes that enable biotrophic development in both *U. maydis* and *Sporisorium reilianum* (Kämper et al., 2006; Schirawski et al., 2010). Interestingly, although these proteins seem to evolve rapidly (Schirawski et al., 2010), they may be identified in closely related species by looking for genomic regions that present higher sequence variability than expected, but that are simultaneously flanked by more conserved regions across species (Schirawski et al., 2010).

Taken together, the results presented in this thesis suggest a complex and mostly unknown relationship between sexuality and lifestyle in the *Ustilago*/ *Pseudozyma* system. Is sex maintained in this group, primarily because of its intimate association with plant pathogenicity and secondly because of its intrinsic benefits? If so, a lifestyle change would explain the finding of apparently asexual lineages like *P. aphidis* for which a sexual counterpart is not known. Therefore, a comparative genomics approach using a wide-range of sexual and asexual species of this group, would probably provide a more genuine picture of the evolution and fate of the gene complexes involved in sexual reproduction and pathogenicity, as well as on their regulation.

6.2. Red yeasts as a novel model for studying the evolution of sex-determining regions in fungi

Although *U. maydis* serves as a remarkable model to understand how sexual reproduction and pathogenicity are interconnected, findings concerning this organism cannot be directly extrapolated to other important, but distantly related groups of fungal pathogens like the anther smuts and the rust fungi. Because of this, in the present study red yeasts of the genera *Rhodosporidium* and *Sporidiobolus* were used as a model to investigate the molecular basis of sexual reproduction in the same phylogenetic lineage of the abovementioned fungal pathogens (Pucciniomycotina), and also to understand fungal sex-determining systems in a broad evolutionary perspective.

In Chapter 3, *MAT*-specific genes encoding pheromone and receptors were first determined in *Rhodosporidium toruloides* along with other *MAT*-specific genes (viz. *STE20* gene) presumably involved in the steps that occur after pheromone stimulation. Their identification and relative organization was facilitated by the availability of whole- or partial-genome information of related red yeast species and contributed to the first insight into the *MAT* locus structure in Pucciniomycotina. Genome sequences also provided a new framework to study the other component of the basidiomycete *MAT* system encoding homeodomain transcription factors. The identification and sequencing in *S. salmonicolor* of the homeodomain transcription factor genes allowed a more detailed analysis of the *MAT* configuration of this species, ultimately revealing a pseudo-bipolar *MAT* system that deviates from the classical bipolar-tetrapolar mating paradigm (Chapter 4). Furthermore, it was also found that the putative *MAT* region is large, multigenic and that synteny is quite conserved within the same mating type throughout the diversity of red yeasts species in Sporidiobolales, but not between mating types (Chapters 4 and 5). The identification of the *MAT* genes in an expanded set of sexual and asexual red yeast (Chapter 5) also brought to light that asexual species maintained *MAT*-specific genes and that the separation between sexual and asexual species in this group is not strictly defined, as illustrated by the existence of asexual strains within sexual species (viz. *R. babjevae*). Therefore, asexuality in this group might promote the dispersion of these yeasts in the environment leading to wide geographic and ecological distributions and, eventually, to the establishment of phenotypically asexual lineages that are currently recognized as asexual species in the genera *Rhodotorula* and *Sporobolomyces*.

Taken together, these results call for an adjustment of the model of *MAT* evolution in the Basidiomycota originally postulated by Raper (1966). In the revised model proposed in Chapter 4, the two classes of *MAT* genes (encoding pheromone/ receptors and homeodomain transcription factors) are portrayed on the same chromosome in the ancestral configuration and most likely genetically unlinked (Figure 4.14). According to this model, the mating system of *S. salmonicolor* may be the result of a gradual transition to a bipolar mating behavior without genomic rearrangements that could place both *MAT* regions in different chromosomes. This is in sharp contrast with the abrupt collapse of the tetrapolar system, which took place in the well documented cases of *Cryptococcus neoformans* and *Ustilago hordei* (Bakkeren and Kronstad, 1994; Lengeler et al., 2002). Furthermore, when analyzing the sequence divergence of genes obtained from both *MAT* regions and mating types of *S. salmonicolor*, little divergence was observed indicating that other molecular mechanisms, like gene conversion, may be acting to counteract the detrimental effects of suppression of recombination. It is also possible that genomic rearrangements occurring within the two *MAT* regions are fairly recent and no significant divergence between alleles has yet been attained despite a drastic and inevitable reduction of recombination.

The proposed evolutionary path of *MAT* systems in Pucciniomycotina, consisting in a gradual transition from tetrapolarity to bipolarity, is the first significant addition to the model envisaged by Raper in 1960s. Therefore, the new findings in *S. salmonicolor* of an intermediate *MAT* system between tetrapolar and bipolar, represent substantial progress towards the understating of the

evolution of sex in fungi and are worthy of being pursued in the future using more detailed and cost-effective approaches. These would include whole-genome sequencing of at least two strains of opposite mating types of the same red yeast species and of different species, using next-generation sequencing (NGS) technologies (Metzker, 2010), which can now be used in *de novo* sequencing projects (Zerbino and Birney, 2008). This approach will enable a more conclusive determination of the extent of the *MAT*-specific regions in these species.

In order to further challenge the model proposed here, it will be interesting to investigate the *MAT* locus structure outside of the order Sporidiobolales, particularly in the saprobic yeast *Leucosporidium scottii* for which a multiallelic tetrapolar mating system has been proposed (Fell and Statzell-Tallman, 1982), but not yet analysed at the molecular level. It will also be important to compare the configuration of the *MAT* system of *L. scottii* and that of the red yeasts with the *MAT* configuration of a third and closely related species, *Microbotryum violaceum*, for which dimorphic "sex chromosomes" have been preliminarily characterized (Hood, 2002; Votintseva and Filatov, 2009; Votintseva and Filatov, 2010) and genomic data is presently available. Together, these approaches will allow refining the model of evolution of *MAT* loci in basidiomycetes.

As a final requirement for red yeasts to become a suitable model system, it would be necessary the development of molecular tools (e.g. transformation system and mutant strains) so as to facilitate the evaluation of gene function. In a very recent study (Ianiri et al., 2011), different transformation procedures were used with success in a red yeast to generate both random insertional mutants and targeted gene deletions by homologous recombination using uracil auxotrophic mutants. Hence, red yeasts emerge now as a remarkable model for studying the evolution of *MAT* systems in fungi and to understand the function of novel genes in the Pucciniomycotina.

REFERENCES

1. **Aanen DK, Hoekstra RF** (2007) Why sex is good: on fungi and beyond. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 527-534.

2. **Abascal F, Zardoya R, Posada D** (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104-2105.

3. **Abe K, Kusaka I, Fukui S** (1975) Morphological change in the early stages of the mating process of *Rhodosporidium toruloides*. J Bacteriol 122:710-718.

4. **Abe K, Sasakuma T** (1986) Identification of a diploid self-sporulating cycle in the basidiomycetous yeast *Rhodosporidium toruloides* J Gen Microbiol 132:1459-1465.

5. **Adachi J, Hasegawa M** (1996) Model of amino acid substitution in proteins encoded by mitochondrial DNA. J Mol Evol 42:459-468.

6. **Akada R, Kai J, Yamashita I, Miyakawa T, Fukui S** (1989a) Genomic organization of multiple genes coding for rhodotorucine *A* lipopeptide mating pheromone of the basidiomycetous yeast *Rhodosporidium toruloides* Arch Microbiol 152:484-487.

7. **Akada R, Minomi K, Kai J, Yamashita I, Miyakawa T, Fukui S** (1989b) Multiple genes coding for precursors of rhodotorucine *A*, a farnesyl peptide mating pheromone of the basidiomycetous yeast *Rhodosporidium toruloides*. Mol Cell Biol 9:3491-3498.

8. **Alby K, Schaefer D, Bennett RJ** (2009) Homothallic and heterothallic mating in the opportunistic pathogen *Candida albicans*. Nature 460:890-893.

9. **Anderson JB, Kohn LM** (1995) Clonality in soilborne, plant-pathogenic fungi. Annu Rev Phytopathol 33:369-391.

10. **Arutchelvi JI, Bhaduri S, Uppara PV, Doble M** (2008) Mannosylerythritol lipids: a review. J Ind Microbiol Biotechnol 35:1559-1570.

11. **Astell CR, Ahlstrom-Jonasson L, Smith M, Tatchell K, Nasmyth KA, Hall BD** (1981) The sequence of the DNAs coding for the mating-type loci of *Saccharomyces cerevisiae*. Cell 27:15-23.

12. **Badrane H, May G** (1999) The divergence-homogenization duality in the evolution of the *b1* mating type gene of *Coprinus cinereus*. Mol Biol Evol 16:975-986.

13. **Bakkeren G, Kronstad JW** (1994) Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. Proc Natl Acad Sci U S A 91:7085-7089.

14. **Bakkeren G, Jiang G, Warren RL, Butterfield Y, Shin H et al.** (2006) Mating factor linkage and genome evolution in basidiomycetous pathogens of cereals. Fungal Genet Biol 43:655-666.

15. **Bakkeren G, Kämper J, Schirawski J** (2008) Sex in smut fungi: Structure, function and evolution of mating-type complexes. Fungal Genet Biol 45 Suppl 1:S15-21.

16. **Banham AH, Asante-Owusu RN, Göttgens B, Thompson S, Kingsnorth CS, Mellor E, Casselton LA** (1995) An N-terminal dimerization domain permits homeodomain proteins to choose compatible partners and Initiate sexual development in the mushroom *Coprinus cinereus*. Plant Cell 7:773-783.

17. **Banuett F, Herskowitz I** (1989) Different *a* alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. Proc Natl Acad Sci U S A 86:5878-5882.

18. **Banuett F** (1995) Genetics of *Ustilago maydis*, a fungal pathogen that induces tumors in maize. Annu Rev Genet 29:179-208.

19. **Banuett F** (2007) History of the mating types in *Ustilago maydis*. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 351-375.

20. **Barsoum E, Martinez P, Astrom SU** (2010) α3, a transposable element that promotes host sexual reproduction. Genes Dev 24:33-44.

21. **Barton NH, Charlesworth B** (1998) Why sex and recombination? Science 281:1986-1990.

22. **Basse CW, Kolb S, Kahmann R** (2002) A maize-specifically expressed gene cluster in *Ustilago maydis*. Mol Microbiol 43:75-93.

23. **Basse CW, Farfsing JW** (2006) Promoters and their regulation in *Ustilago maydis* and other phytopathogenic fungi. FEMS Microbiol Lett 254:208-216.

24. **Begerow D, Bauer R, Boekhout T** (2000) Phylogenetic placements of ustilaginomycetous anamorphs as deduced from nuclear LSU rDNA sequences. Mycol Res 104:53-60.

25. **Bell G** (1982) The masterpiece of nature: the evolution and genetics of sexuality. CroomHelm, London, United Kingdom.

26. **Billiard S, López-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T** (2011) Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol Rev 86:421-442.

27. **Boekhout T, Fell JW** (1998) *Pseudozyma* Bandoni emend. Boekhout and comparison with the yeast state of *Ustilago maydis* (De Candolle) Corda. In: Kurtzman CP, Fell JW (eds) The Yeasts: a taxonomic study, The Netherlands, pp 790-797.

28. **Bölker M, Urban M, Kahmann R** (1992) The *a* mating type locus of *U. maydis* specifies cell signaling components. Cell 68:441-450.

29. **Bölker M, Kahmann R** (1993) Sexual pheromones and mating responses in fungi. Plant Cell 5:1461-1469.

30. **Bortfeld M, Auffarth K, Kahmann R, Basse CW** (2004) The *Ustilago maydis a2* mating-type locus genes *lga2* and *rga2* compromise pathogenicity in the absence of the mitochondrial p32 family protein Mrb1. Plant Cell 16:2233-2248.

31. **Brachmann A, König J, Julius C, Feldbrügge M** (2004) A reverse genetic approach for generating gene replacement mutants in *Ustilago maydis*. Mol Genet Genomics 272:216-226.

32. **Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R** (2009) *Ustilago maydis* as a Pathogen. Annu Rev Phytopathol 47:423-445.

33. **Brown AJ, Casselton LA** (2001) Mating in mushrooms: increasing the chances but prolonging the affair. Trends Genet 17:393-400.

34. **Buller AHR** (1950) Researches on Fungi: The sexual process in the Uredinales. University of Toronto Press, Toronto, Canada.

35. **Butler G, Kenny C, Fagan A, Kurischko C, Gaillardin C, Wolfe KH** (2004) Evolution of the *MAT* locus and its *Ho* endonuclease in yeast species. Proc Natl Acad Sci U S A 101:1632-1637.

36. **Butler G** (2007) The evolution of *MAT*: the ascomycetes. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 3-18.

37. **Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S et al.** (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459:657-662.

38. **Butler G** (2010) Fungal sex and pathogenesis. Clin Microbiol Rev 23:140-159.

39. **Buzzini P, Martini A** (2000) Biodiversity of killer activity in yeasts isolated from the Brazilian rain forest. Can J Microbiol 46:607-611.

40. **Caldwell GA, Naider F, Becker JM** (1995) Fungal lipopeptide mating pheromones: a model system for the study of protein prenylation. Microbiol Rev 59:406-422.

41. **Cao Y, Adachi J, Janke A, Pääbo S, Hasegawa M** (1994) Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. J Mol Evol 39:519-527.

42. **Casselton LA, Olesnicky NS** (1998) Molecular genetics of mating recognition in basidiomycete fungi. Microbiol Mol Biol Rev 62:55-70.

43. **Casselton LA** (2002) Mate recognition in fungi. Heredity 88:142-147.

44. **Casselton LA, Kües U** (2007) The origin of multiple mating types in the model mushrooms *Coprinopsis cinerea* and *Schizophyllum commune* In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 283-300.

45. **Charlesworth D, Charlesworth B, Marais G** (2005) Steps in the evolution of heteromorphic sex chromosomes. Heredity 95:118-128.

46. **Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP** (2007) Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8:762-775.

47. **Clark AG, Kao TH** (1991) Excess nonsynonymous substitution of shared polymorphic sites among self-incompatibility alleles of Solanaceae. Proc Natl Acad Sci U S A 88:9823-9827.

48. **Coelho MA, Rosa A, Rodrigues N, Fonseca Á, Gonçalves P** (2008) Identification of mating type genes in the bipolar basidiomycetous yeast *Rhodosporidium toruloides*: first insight into the *MAT* locus structure of the Sporidiobolales. Eukaryot Cell 7:1053-1061.

49. **Coelho MA, Sampaio JP, Gonçalves P** (2010) A deviation from the bipolar-tetrapolar mating paradigm in an early diverged basidiomycete. PLoS Genet 6:e1001052.

50. **Coppin E, Debuchy R, Arnaise S, Picard M** (1997) Mating types and sexual development in filamentous ascomycetes. Microbiol Mol Biol Rev 61:411-428.

51. **Dacks J, Roger AJ** (1999) The first sexual lineage and the relevance of facultative sex. J Mol Evol 48:779-783.

52. **Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK et al.** (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434:980-986.

53. **Devier B, Aguileta G, Hood ME, Giraud T** (2009) Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. Genetics 181:209-223.

54. **Devier B, Aguileta G, Hood ME, Giraud T** (2010) Using phylogenies of pheromone receptor genes in the *Microbotryum violaceum* species complex to investigate possible speciation by hybridization. Mycologia 102:689-696.

55. **Dhingra OD, Sinclair JB** (1995) Basic Plant Pathology Methods. CRC Press, Boca Raton, Florida.

56. **Doehlemann G, Wahl R, Horst RJ, Voll LM, Usadel B et al.** (2008) Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*. Plant J 56:181-195.

57. **Edgar RC** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.

58. **Endo T, Ikeo K, Gojobori T** (1996) Large-scale search for genes on which positive selection may operate. Mol Biol Evol 13:685-690.

59. **Fedler M, Luh KS, Stelter K, Nieto-Jacobo F, Basse CW** (2009) The *a2* mating-type locus genes *lga2* and *rga2* direct uniparental mitochondrial DNA (mtDNA) inheritance and constrain mtDNA recombination during sexual development of *Ustilago maydis*. Genetics 181:847-860.

60. **Fell JW, Statzell-Tallman A** (1982) Multiple allelic incompatibility factors among bifactorial strains of the yeast *Leucosporidium* (*Candida*) *scottii*. Curr Microbiol 7:213-215.

61. **Fell JW, Statzell-Tallman A** (1998) *Rhodosporidium* Banno. In: Kurtzman CP, Fell JW (eds) The Yeasts, a taxonomic study, 4th edn. Elsevier, Amsterdam, The Netherlands, pp 678-692.

62. **Fell JW, Boekhout T, Fonseca Á, Sampaio JP** (2001) Basidiomycetous yeasts. In: McLaughlin DJ, McLaughlin EJ, Lemke PA (eds) The Mycota, VII, Part B, Systematics and evolution. Springer-Verlag, Berlin, pp 3-35.

63. **Ferreira AV, An Z, Metzenberg RL, Glass NL** (1998) Characterization of *mat A-2*, *mat A-3* and *ΔmatA* mating-type mutants of *Neurospora crassa*. Genetics 148:1069-1079.

64. **Findley K, Rodriguez-Carres M, Metin B, Kroiss J, Fonseca Á, Vilgalys R, Heitman J** (2009) Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the Tremellales. Eukaryot Cell 8:353-361.

65. **Fisher MC, Hanage WP, de Hoog S, Johnson E, Smith MD, White NJ, Vanittanakom N** (2005) Low effective dispersal of asexual genotypes in heterogeneous landscapes by the endemic pathogen *Penicillium marneffei*. PLoS Pathog 1:e20.

66. **Fitzpatrick DA, Logue ME, Stajich JE, Butler G** (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol 6:99.

67. **Fowler TJ, DeSimone SM, Mitton MF, Kurjan J, Raper CA** (1999) Multiple sex pheromones and receptors of a mushroom-producing fungus elicit mating in yeast. Mol Biol Cell 10:2559-2572.

68. **Fowler TJ, Mitton MF, Rees EI, Raper CA** (2004) Crossing the boundary between the *Bα* and *Bβ* mating-type loci in *Schizophyllum commune*. Fungal Genet Biol 41:89-101.

69. **Fraser JA, Heitman J** (2003) Fungal mating-type loci. Curr Biol 13:R792-795.

70. **Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS, Heitman J** (2004) Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS Biol 2:e384.

71. **Fraser JA, Heitman J** (2004) Evolution of fungal sex chromosomes. Mol Microbiol 51:299-306.

72. **Fraser JA, Heitman J** (2005) Chromosomal sex-determining regions in animals, plants and fungi. Curr Opin Genet Dev 15:645-651.

73. **Fraser JA, Hsueh YP, Findley KM, Heitman J** (2007) Evolution of the mating-type locus: the basidiomycetes. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 19-34.

74. **Gadanho M, Sampaio JP** (2002) Polyphasic taxonomy of the basidiomycetous yeast genus *Rhodotorula*: *Rh. glutinis* sensu stricto and *Rh. dairenensis* comb. nov. FEMS Yeast Res 2:47-58.

75. **Gadanho M, Almeida JM, Sampaio JP** (2003) Assessment of yeast diversity in a marine environment in the south of Portugal by microsatellite-primed PCR. Antonie Van Leeuwenhoek 84:217-227.

76. **Gadanho M, Libkind D, Sampaio JP** (2006) Yeast diversity in the extreme acidic environments of the Iberian Pyrite Belt. Microb Ecol 52:552-563.

77. **Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR et al.** (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature 438:1105-1115.

78. **Galgoczy DJ, Cassidy-Stone A, Llinas M, O'Rourke SM, Herskowitz I, DeRisi JL, Johnson AD** (2004) Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 101:18069-18074.

79. **Gillissen B, Bergemann J, Sandmann C, Schroeer B, Bölker M, Kahmann R** (1992) A two-component regulatory system for self/non-self recognition in *Ustilago maydis*. Cell 68:647-657.

80. **Giraud T, Yockteng R, López-Villavicencio M, Refrégier G, Hood ME** (2008) Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. Eukaryot Cell 7:765-775.

81. **Glass NL, Vollmer SJ, Staben C, Grotelueschen J, Metzenberg RL, Yanofsky C** (1988) DNAs of the two mating-type alleles of *Neurospora crassa* are highly dissimilar. Science 241:570-573.

82. **Goddard MR, Godfray HC, Burt A** (2005) Sex increases the efficacy of natural selection in experimental yeast populations. Nature 434:636-640.

83. **Grandel A, Romeis T, Kämper J** (2000) Regulation of pathogenic development in the corn smut fungus *Ustilago maydis*. Mol Plant Pathol 1:61-66.

84. **Gryganskyi AP, Lee SC, Litvintseva AP, Smith ME, Bonito G, Porter TM, Anishchenko IM, Heitman J, Vilgalys R** (2010) Structure, function, and phylogeny of the mating locus in the *Rhizopus oryzae* complex. PLoS One 5:e15273.

85. **Guindon S, Gascuel O** (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696-704.

86. **Gustin MC, Albertyn J, Alexander M, Davenport K** (1998) MAP kinase pathways in the yeast *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev 62:1264-1300.

87. **Hadany L, Comeron JM** (2008) Why are sex and recombination so common? Ann N Y Acad Sci 1133:26-43.

88. **Hagen DC, McCaffrey G, Sprague GF, Jr.** (1986) Evidence the yeast *STE3* gene encodes a receptor for the peptide pheromone a factor: gene sequence and implications for the structure of the presumed receptor. Proc Natl Acad Sci U S A 83:1418-1422.

89. **Hajlaoui MR, Bélanger RR** (1993) Antagonism of the yeast-like phylloplane fungus *Sporothrix flocculosa* against *Erysiphe graminis* var. *tritici*. Biocontrol Sci Technol 3:427-434.

90. **Hall TA** (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41 95-98.

91. **Hamamoto M, Boekhout T, Nakase T** (2011) *Sporobolomyces* Kluyver & van Niel. In: Kurtzman CP, Fell JW, Boekhout T (eds) The Yeasts, a taxonomic study, Volume 3, Part Vc, 5th edn. Elsevier, Amsterdam, The Netherlands, pp 1929-1990.

92. **Hammami W, Castro CQ, Rémus-Borel W, Labbé C, Bélanger RR** (2011) Ecological basis of the interaction between *Pseudozyma flocculosa* and powdery mildew fungi. Appl Environ Microbiol 77:926-933.

93. **Hartmann HA, Kahmann R, Bölker M** (1996) The pheromone response factor coordinates filamentous growth and pathogenicity in *Ustilago maydis*. EMBO J 15:1632-1641.

94. **Hartmann HA, Krüger J, Lottspeich F, Kahmann R** (1999) Environmental signals controlling sexual development of the corn Smut fungus *Ustilago maydis* through the transcriptional regulator Prf1. Plant Cell 11:1293-1306.

95. **Haskins RH, Thorn JA, Boothroyd B** (1955) Biochemistry of the Ustilaginales. XI. Metabolic products of *Ustilago zeae* in submerged culture. Can J Microbiol 1:749-756.

96. **Hawksworth DL, Kirk PM, Sutton BC, Pelger DN** (1995) Ainsworth and Bisby's Dictionary of the Fungi, 8th edn. CABI, Oxon, UK.

97. **Hawksworth DL** (2011) Naming *Aspergillus* species: progress towards one name for each species. Med Mycol 49 Suppl 1:S70-76.

98. **Hegner J, Siebert-Bartholmei C, Kothe E** (1999) Ligand recognition in multiallelic pheromone receptors from the basidiomycete *Schizophyllum commune* studied in yeast. Fungal Genet Biol 26:190-197.

99. **Heitman J** (2006) Sexual reproduction and the evolution of microbial pathogens. Curr Biol 16:R711-725.

100. **Heitman J** (2010) Evolution of eukaryotic microbial pathogens via covert sexual reproduction. Cell Host Microbe 8:86-99.

101. **Henninger W, Windisch S** (1975) A new yeast of *Sterigmatomyces*, *S. aphidis* sp. n. Arch Microbiol 105:49-50.

102. **Herskowitz I, Rine J, Strathern J** (1992) Mating-type determination and mating-type interconversion in *Saccharomyces cerevisiae*. In: Broach JR, Pringle J, Jones EW (eds) The molecular and cellular biology of the yeast *Saccharomyces*, vol. 2, Gene expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA, pp 583-656.

103. **Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF et al.** (2007) A higher-level phylogenetic classification of the Fungi. Mycol Res 111:509-547.

104. **Hicks JB, Herskowitz I** (1976) Interconversion of yeast mating types I. Direct observations of the action of the homothallism (*HO*) gene. Genetics 83:245-258.

105. **Hinata K, Watanabe M, Yamakawa S, Satta Y, Isogai A** (1995) Evolutionary aspects of the *S*-related genes of the *Brassica* self-incompatibility system: synonymous and nonsynonymous base substitutions. Genetics 140:1099-1104.

106. **Hofmann C, Shepelev M, Chernoff J** (2004) The genetics of Pak. J Cell Sci 117:4343-4354.

107. **Hood ME** (2002) Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. Genetics 160:457-461.

108. **Hood ME, Antonovics J, Koskella B** (2004) Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. Genetics 168:141-146.

109. **Hsueh YP, Shen WC** (2005) A homolog of Ste6, the *a*-factor transporter in *Saccharomyces cerevisiae*, is required for mating but not for monokaryotic fruiting in *Cryptococcus neoformans*. Eukaryot Cell 4:147-155.

110. **Hsueh YP, Idnurm A, Heitman J** (2006) Recombination hotspots flank the *Cryptococcus* mating-type locus: implications for the evolution of a fungal sex chromosome. PLoS Genet 2:e184.

111. **Hsueh YP, Fraser JA, Heitman J** (2008) Transitions in sexuality: recapitulation of an ancestral tri- and tetrapolar mating system in *Cryptococcus neoformans*. Eukaryot Cell 7:1847-1855.

112. **Hull CM, Davidson RC, Heitman J** (2002) Cell identity and sexual development in *Cryptococcus neoformans* are controlled by the mating-type-specific homeodomain protein Sxi1α. Genes Dev 16:3046-3060.

113. **Hull CM, Boily MJ, Heitman J** (2005) Sex-specific homeodomain proteins Sxi1α and Sxi2a coordinately regulate sexual development in *Cryptococcus neoformans*. Eukaryot Cell 4:526-535.

114. **Hwang-Shum JJ, Hagen DC, Jarvis EE, Westby CA, Sprague GF, Jr.** (1991) Relative contributions of *MCM1* and *STE12* to transcriptional activation of a- and α-specific genes from *Saccharomyces cerevisiae*. Mol Gen Genet 227:197-204.

115. **Ianiri G, Wright SA, Castoria R, Idnurm A** (2011) Development of resources for the analysis of gene function in Pucciniomycotina red yeasts. Fungal Genet Biol doi:10.1016/j.fgb.2011.03.003.

116. **Idnurm A, Walton FJ, Floyd A, Heitman J** (2008) Identification of the sex genes in an early diverged fungus. Nature 451:193-196.

117. **Isoda H, Kitamoto D, Shinmoto H, Matsumura M, Nakahara T** (1997) Microbial extracellular glycolipid induction of differentiation and inhibition of the protein kinase C activity of human promyelocytic leukemia cell line HL60. Biosci Biotechnol Biochem 61:609-614.

118. **James TY, Srivilai P, Kues U, Vilgalys R** (2006) Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of mating-type-specific pheromone receptor function. Genetics 172:1877-1891.

119. **James TY** (2007) Analysis of mating-type locus organization and synteny in mushroom fungi: beyond model species. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 317-330.

120. **James TY, Lee M, van Diepen LT** (2011) A single mating-type locus composed of homeodomain genes promotes nuclear migration and heterokaryosis in the white-rot fungus *Phanerochaete chrysosporium*. Eukaryot Cell 10:249-261.

121. **Jones DT, Taylor WR, Thornton JM** (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275-282.

122. **Kaffarnik F, Müller P, Leibundgut M, Kahmann R, Feldbrügge M** (2003) PKA and MAPK phosphorylation of Prf1 allows promoter discrimination in *Ustilago maydis*. EMBO J 22:5817-5826.

123. **Kahmann R, Schirawski J** (2007) Mating in the smut fungi: from *a* to *b* to the downstream cascades. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 377-387.

124. **Kamada T** (2002) Molecular genetics of sexual development in the mushroom *Coprinus cinereus*. Bioessays 24:449-459.

125. **Kamiya Y, Sakurai A, Tamura S, Takahashi N** (1978a) Structure of rhodotorucine *A*, a novel lipopeptide, inducing mating tube formation in *Rhodosporidium toruloides*. Biochem Biophys Res Commun 83:1077-1083.

126. **Kamiya Y, Sakurai A, Tamura S, Takahashi N, Abe K, Tsuchiya E, Fukui S** (1978b) Isolation of rhodotorucine *A*, a peptidyl factor inducing the mating tube formation in *Rhodosporidium toruloides*. Agric Biol Chem 42:1239-1243.

127. **Kämper J, Reichmann M, Romeis T, Bölker M, Kahmann R** (1995) Multiallelic recognition: nonself-dependent dimerization of the bE and bW homeodomain proteins in *Ustilago maydis*. Cell 81:73-83.

128. **Kämper J** (2004) A PCR-based system for highly efficient generation of gene replacement mutants in *Ustilago maydis*. Mol Genet Genomics 271:103-110.

129. **Kämper J, Kahmann R, Bölker M, Ma LJ, Brefort T et al.** (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature 444:97-101.

130. **Karlson P, Luscher M** (1959) Pheromones': a new term for a class of biologically active substances. Nature 183:55-56.

131. **Kelly M, Burke J, Smith M, Klar A, Beach D** (1988) Four mating-type genes control sexual differentiation in the fission yeast. EMBO J 7:1537-1547.

132. **Kirk MP, Cannon PF, Minter DW, Stalpers JA** (2008) Dictionary of the Fungi, 10th edn. CABI, Oxon, UK.

133. **Kitamoto D, Akiba S, Hioki C, Tabuchi T** (1990) Extracellular accumulation of mannosylerythritol lipids by a strain of *Candida antarctica*. Agric Biol Chem 54 31-36.

134. **Klix V, Nowrousian M, Ringelberg C, Loros JJ, Dunlap JC, Pöggeler S** (2010) Functional characterization of *MAT1-1*-specific mating-type genes in the homothallic ascomycete *Sordaria macrospora* provides new insights into essential and nonessential sexual regulators. Eukaryot Cell 9:894-905.

135. **Kondrashov AS** (1993) Classification of hypotheses on the advantage of amphimixis. J Hered 84:372-387.

136. **Kües U, Asante-Owusu RN, Mutasa ES, Tymon AM, Pardo EH, O'Shea SF, Göttgens B, Casselton LA** (1994a) Two classes of homeodomain proteins specify the multiple a mating types of the mushroom *Coprinus cinereus*. Plant Cell 6:1467-1475.

137. **Kües U, Göttgens B, Stratmann R, Richardson WV, O'Shea SF, Casselton LA** (1994b) A chimeric homeodomain protein causes self-compatibility and constitutive sexual development in the mushroom *Coprinus cinereus*. EMBO J 13:4054-4059.

138. **Kües U** (2000) Life history and developmental processes in the basidiomycete *Coprinus cinereus*. Microbiol Mol Biol Rev 64:316-353.

139. **Kües U, Walser PJ, Klaus MJ, Aebi M** (2002) Influence of activated *A* and *B* mating-type pathways on developmental processes in the basidiomycete *Coprinus cinereus*. Mol Genet Genomics 268:262-271.

140. **Kurtzman CP, Robnett CJ** (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. Antonie Van Leeuwenhoek 73:331-371.

141. **Kwon-Chung KJ** (1976) Morphogenesis of *Filobasidiella neoformans*, the sexual state of *Cryptococcus neoformans*. Mycologia 68:821-833.

142. **Kwon-Chung KJ, Bennett JE** (1978) Distribution of a and α mating types of *Cryptococcus neoformans* among natural and clinical isolates. Am J Epidemiol 108:337-340.

143. **Lahn BT, Page DC** (1999) Four evolutionary strata on the human X chromosome. Science 286:964-967.

144. **Lee N, Bakkeren G, Wong K, Sherwood JE, Kronstad JW** (1999) The mating-type and pathogenicity locus of the fungus *Ustilago hordei* spans a 500-kb region. Proc Natl Acad Sci U S A 96:15026-15031.

145. **Lee SC, Corradi N, Doan S, Dietrich FS, Keeling PJ, Heitman J** (2010a) Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. PLoS One 5:e10539.

146. **Lee SC, Ni M, Li W, Shertz C, Heitman J** (2010b) The evolution of sex: a perspective from the fungal kingdom. Microbiol Mol Biol Rev 74:298-340.

147. **Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS, Heitman J** (2002) Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. Eukaryot Cell 1:704-718.

148. **Leubner-Metzger G, Horwitz BA, Yoder OC, Turgeon BG** (1997) Transcripts at the mating type locus of *Cochliobolus heterostrophus*. Mol Gen Genet 256:661-673.

149. **Leveleki L, Mahlert M, Sandrock B, Bölker M** (2004) The PAK family kinase Cla4 is required for budding and morphogenesis in *Ustilago maydis*. Mol Microbiol 54:396-406.

150. **Li W, Metin B, White TC, Heitman J** (2010) Organization and evolutionary trajectory of the mating type (*MAT*) locus in dermatophyte and dimorphic fungal pathogens. Eukaryot Cell 9:46-58.

151. **Librado P, Rozas J** (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

152. **Lin SS, Pranikoff T, Smith SF, Brandt ME, Gilbert K, Palavecino EL, Shetty AK** (2008) Central venous catheter infection associated with *Pseudozyma aphidis* in a child with short gut syndrome. J Med Microbiol 57:516-518.

153. **Lin X, Hull CM, Heitman J** (2005) Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature 434:1017-1021.

154. **Lin X, Heitman J** (2007) Mechanisms of homothallism in Fungi and transitions between heterothallism and homothallism. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 35-57.

155. **Lin X, Litvintseva AP, Nielsen K, Patel S, Floyd A, Mitchell TG, Heitman J** (2007) αADα hybrids of *Cryptococcus neoformans*: evidence of same-sex mating in nature and hybrid fitness. PLoS Genet 3:1975-1990.

156. **Lin X, Patel S, Litvintseva AP, Floyd A, Mitchell TG, Heitman J** (2009) Diploids in the *Cryptococcus neoformans* serotype A population homozygous for the α mating type originate via unisexual mating. PLoS Pathog 5:e1000283.

157. **LoBuglio KF, Pitt JL, Taylor JW** (1993) Phylogenetic analysis of two ribosomal DNA regions indicates multiple independent losses of a sexual *Talaromyces* state among asexual *Penicillium* species in subgenus *Biverticillium* Mycologia 85 592-604.

158. **López-Villavicencio M, Aguileta G, Giraud T, de Vienne DM, Lacoste S, Couloux A, Dupont J** (2010) Sex in *Penicillium*: combined phylogenetic and experimental approaches. Fungal Genet Biol 47:693-706.

159. **Lynch M, Conery JS** (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155.

160. **Mahlert M, Vogler C, Stelter K, Hause G, Basse CW** (2009) The *a2* mating-type-locus gene *lga2* of *Ustilago maydis* interferes with mitochondrial dynamics and fusion, partially in dependence on a Dnm1-like fission component. J Cell Sci 122:2402-2412.

161. **Martin T, Lu SW, van Tilbeurgh H, Ripoll DR, Dixelius C, Turgeon BG, Debuchy R** (2010) Tracing the origin of the fungal α1 domain places its ancestor in the HMG-box superfamily: implication for fungal mating-type evolution. PLoS One 5:e15199.

162. **Martinez-Espinoza AD, García-Pedrajas MD, Gold SE** (2002) The Ustilaginales as plant pests and model systems. Fungal Genet Biol 35:1-20.

163. **Martinez D, Larrondo LF, Putnam N, Gelpke MD, Huang K et al.** (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. Nat Biotechnol 22:695-700.

164. **May G, Matzke E** (1995) Recombination and variation at the *A* mating-type of *Coprinus cinereus*. Mol Biol Evol:794-802.

165. **May G, Shaw F, Badrane H, Vekemans X** (1999) The signature of balancing selection: fungal mating compatibility gene evolution. Proc Natl Acad Sci U S A 96:9172-9177.

166. **Maynard-Smith J** (1978) The Evolution of Sex, 1st edn. Cambridge University Press, Cambridge, UK.

167. **McClelland CM, Fu J, Woodlee GL, Seymour TS, Wickes BL** (2002) Isolation and characterization of the *Cryptococcus neoformans MATa* pheromone gene. Genetics 160:935-947.

168. **McGinnis W, Krumlauf R** (1992) Homeobox genes and axial patterning. Cell 68:283-302.

169. **Menkis A, Jacobson DJ, Gustafsson T, Johannesson H** (2008) The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. PLoS Genet 4:e1000030.

170. **Menkis A, Whittle CA, Johannesson H** (2010) Gene genealogies indicates abundant gene conversions and independent evolutionary histories of the mating-type chromosomes in the evolutionary history of *Neurospora tetrasperma*. BMC Evol Biol 10:234.

171. **Merino ST, Nelson MA, Jacobson DJ, Natvig DO** (1996) Pseudohomothallism and evolution of the mating-type chromosome in *Neurospora tetrasperma*. Genetics 143:789-799.

172. **Metin B, Findley K, 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 Genet 6:e1000961.

173. **Metzenberg RL, Glass NL** (1990) Mating type and mating strategies in *Neurospora*. Bioessays 12:53-59.

174. **Metzker ML** (2010) Sequencing technologies - the next generation. Nat Rev Genet 11:31-46.

175. **Michaelis S, Chen P, Berkower C, Sapperstein S, Kistler A** (1992) Biogenesis of yeast *a*-factor involves prenylation, methylation and a novel export mechanism. Antonie Van Leeuwenhoek 61:115-117.

176. **Middelhoven WJ** (1997) Identity and biodegradative abilities of yeasts isolated from plants growing in an arid climate. Antonie Van Leeuwenhoek 72:81-89.

177. **Moore TD, Edman JC** (1993) The α-mating type locus of *Cryptococcus neoformans* contains a peptide pheromone gene. Mol Cell Biol 13:1962-1970.

178. **Morrow CA, Fraser JA** (2009) Sexual reproduction and dimorphism in the pathogenic basidiomycetes. FEMS Yeast Res 9:161-177.

179. **Muirhead CA, Glass NL, Slatkin M** (2002) Multilocus self-recognition systems in fungi as a cause of trans-species polymorphism. Genetics 161:633-641.

180. **Muller HJ** (1964) The relation of recombination to mutational advance. Mutat Res 106:2-9.

181. **Müller P, Aichinger C, Feldbrügge M, Kahmann R** (1999) The MAP kinase kpp2 regulates mating and pathogenic development in *Ustilago maydis*. Mol Microbiol 34:1007-1017.

182. **Munkacsi AB, Stoxen S, May G** (2007) Domestication of maize, sorghum, and sugarcane did not drive the divergence of their smut pathogens. Evolution 61:388-403.

183. **Nadal M, García-Pedrajas MD, Gold SE** (2008) Dimorphism in fungal plant pathogens. FEMS Microbiol Lett 284:127-134.

184. **Narisawa K, Yamaoka Y, Katsuya K** (1994) Mating type of isolates derived from the spermogonial state of *Puccinia coronata* var. *coronata*. Mycoscience 35:131-135.

185. **Nichols CB, Fraser JA, Heitman J** (2004) PAK kinases Ste20 and Pak1 govern cell polarity at different stages of mating in *Cryptococcus neoformans*. Mol Biol Cell 15:4476-4489.

186. **Niculita-Hirzel H, Labbé J, Kohler A, le Tacon F, Martin F, Sanders IR, Kües U** (2008) Gene organization of the mating type regions in the ectomycorrhizal fungus *Laccaria bicolor* reveals distinct evolution between the two mating type loci. New Phytol 180:329-342.

187. **Nielsen K, Heitman J** (2007) Sex and virulence of human pathogenic fungi. Adv Genet 57:143-173.

188. **O'Donnell K, Ward TJ, Geiser DM, Corby Kistler H, Aoki T** (2004) Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. Fungal Genet Biol 41:600-623.

189. **Ohm RA, de Jong JF, Lugones LG, Aerts A, Kothe E et al.** (2010) Genome sequence of the model mushroom *Schizophyllum commune*. Nat Biotechnol 28:957-963.

190. **Olesnicky NS, Brown AJ, Dowell SJ, Casselton LA** (1999) A constitutively active G-protein-coupled receptor causes mating self-compatibility in the mushroom *Coprinus*. EMBO J 18:2756-2763.

191. **Olesnicky NS, Brown AJ, Honda Y, Dyos SL, Dowell SJ, Casselton LA** (2000) Self-compatible *B* mutants in *Coprinus* with altered pheromone-receptor specificities. Genetics 156:1025-1033.

192. **Paoletti M, Rydholm C, Schwier EU, Anderson MJ, Szakacs G et al.** (2005) Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. Curr Biol 15:1242-1248.

193. **Poirot O, O'Toole E, Notredame C** (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res 31:3503-3506.

194. **Posada D, Crandall KA** (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

195. **Raju NB** (1992) Functional heterothallism resulting from homokaryotic conidia and ascospores in *Neurospora tetrasperma*. Mycol Res 96:103-116.

196. **Raju NB, Perkins DD** (1994) Diverse programs of ascus development in pseudohomothallic species of *Neurospora*, *Gelasinospora*, and *Podospora*. Dev Genet 15:104-118.

197. **Ramesh MA, Malik SB, Logsdon JM, Jr.** (2005) A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. Curr Biol 15:185-191.

198. **Raper JR** (1966) Genetics of sexuality in Higher Fungi. The Roland Press, New York.

199. **Raudaskoski M, Kothe E** (2010) Basidiomycete mating type genes and pheromone signaling. Eukaryot Cell 9:847-859.

200. **Reeves JH** (1992) Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. J Mol Evol 35:17-31.

201. **Richman A** (2000) Evolution of balanced genetic polymorphism. Mol Ecol 9:1953-1963.

202. **Riquelme M, Challen MP, Casselton LA, Brown AJ** (2005) The origin of multiple *B* mating specificities in *Coprinus cinereus*. Genetics 170:1105-1119.

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

204. **Rouquier S, Blancher A, Giorgi D** (2000) The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. Proc Natl Acad Sci U S A 97:2870-2874.

205. **Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC** (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873-876.

206. **Saitou N, Nei M** (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

207. **Sampaio JP, Gadanho M, Santos S, Duarte FL, Pais C, Fonseca Á, Fell JW** (2001) Polyphasic taxonomy of the basidiomycetous yeast genus *Rhodosporidium*: *Rhodosporidium kratochvilovae* and related anamorphic species. Int J Syst Evol Microbiol 51:687-697.

208. **Sampaio JP, Gadanho M, Bauer R, Weiss M** (2003) Taxonomic studies in the Microbotryomycetidae: *Leucosporidium golubevii* sp. nov., *Leucosporidiella* gen. nov. and the new orders Leucosporidiales and Sporidiobolales. Mycol Prog 2:53-68.

209. **Sampaio JP** (2011a) *Rhodosporidium* Banno. In: Kurtzman CP, Fell JW, Boekhout T (eds) The Yeasts, a taxonomic study, Volume 3, Part Vb, 5th edn. Elsevier, Amsterdam, The Netherlands, pp 1523-1540.

210. **Sampaio JP** (2011b) *Rhodotorula* Harrison. In: Kurtzman CP, Fell JW, Boekhout T (eds) The Yeasts, a taxonomic study, Volume 3, Part Vc, 5th edn. Elsevier, Amsterdam, The Netherlands, pp 1873-1928.

211. **Sampaio JP** (2011c) *Sporidiobolus* Nyland. In: Kurtzman CP, Fell JW, Boekhout T (eds) The Yeasts, a taxonomic study, Volume 3, Part Vb, 5th edn. Elsevier, Amsterdam, The Netherlands, pp 1549-1562.

212. **Schäfer AM, Kemler M, Bauer R, Begerow D** (2010) The illustrated life cycle of *Microbotryum* on the host plant *Silene latifolia*. Botany 88:875-885.

213. **Schirawski J, Heinze B, Wagenknecht M, Kahmann R** (2005) Mating type loci of *Sporisorium reilianum*: novel pattern with three *a* and multiple *b* specificities. Eukaryot Cell 4:1317-1327.

214. **Schirawski J, Mannhaupt G, Münch K, Brefort T, Schipper K et al.** (2010) Pathogenicity determinants in smut fungi revealed by genome comparison. Science 330:1546-1548.

215. **Schulz B, Banuett F, Dahl M, Schlesinger R, Schäfer W, Martin T, Herskowitz I, Kahmann R** (1990) The *b* alleles of *U. maydis*, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif. Cell 60:295-306.

216. **Schurko AM, Logsdon JM, Jr.** (2008) Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. Bioessays 30:579-589.

217. **Seifert KA, Gams W** (2001) The taxonomy of anamorphic fungi. In: McLaughlin DJ, McLaughlin EJ, Lemke PA (eds) The Mycota VII, Part A, Systematics and Evolution. Springer-Verlag, Berlin, pp 307-347.

218. **Sexton AC, Howlett BJ** (2006) Parallels in fungal pathogenesis on plant and animal hosts. Eukaryot Cell 5:1941-1949.

219. **Smith DG, Garcia-Pedrajas MD, Hong W, Yu Z, Gold SE, Perlin MH** (2004) An ste20 homologue in *Ustilago maydis* plays a role in mating and pathogenicity. Eukaryot Cell 3:180-189.

220. **Smulian AG, Sesterhenn T, Tanaka R, Cushion MT** (2001) The ste3 pheromone receptor gene of *Pneumocystis carinii* is surrounded by a cluster of signal transduction genes. Genetics 157:991-1002.

221. **Sprague GF, Jr., Thorner JW** (1992) Pheromone response and signal transduction during the mating process of *Saccharomyces cerevisiae*. In: Broach JR, Pringle J, Jones EW (eds) The molecular and cellular biology of the yeast *Saccharomyces*, vol. 2, Gene expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA, pp 657-744.

222. **Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, Taylor JW** (2009) The fungi. Curr Biol 19:R840-845.

223. **Stajich JE, Wilke SK, Ahrén D, Au CH, Birren BW et al.** (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). Proc Natl Acad Sci U S A 107:11889-11894.

224. **Stanke M, Diekhans M, Baertsch R, Haussler D** (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637-644.

225. **Stanton BC, Hull CM** (2007) Mating-type locus control of cell identity. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 59-73.

226. **Stanton BC, Giles SS, Staudt MW, Kruzel EK, Hull CM** (2010) Allelic exchange of pheromones and their receptors reprograms sexual identity in *Cryptococcus neoformans*. PLoS Genet 6:e1000860.

227. **Statzell-Tallman A, Fell JW** (1998) *Sporidiobolus* Nyland. In: Kurtzman CP, Fell JW (eds) The Yeasts, a taxonomic study, 4th edn. Elsevier, Amsterdam, The Netherlands, pp 693-699.

228. **Steinberg G, Perez-Martin J** (2008) *Ustilago maydis*, a new fungal model system for cell biology. Trends Cell Biol 18:61-67.

229. **Stoll M, Begerow D, Oberwinkler F** (2005) Molecular phylogeny of *Ustilago*, *Sporisorium*, and related taxa based on combined analyses of rDNA sequences. Mycol Res 109:342-356.

230. **Sugita T, Takashima M, Poonwan N, Mekha N, Malaithao K, Thungmuthasawat B, Prasarn S, Luangsook P, Kudo T** (2003) The first isolation of ustilaginomycetous anamorphic yeasts, *Pseudozyma* species, from patients' blood and a description of two new species: *P. parantarctica* and *P. thailandica*. Microbiol Immunol 47:183-190.

231. **Swiezynski KM, Day PR** (1960) Heterokaryon formation in *Coprinus lagopus*. Genet Res 1:114-128.

232. **Tamura K, Nei M** (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512-526.

233. **Tamura K, Dudley J, Nei M, Kumar S** (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599.

234. **Taylor J, Jacobson D, Fisher M** (1999) The evolution of asexual fungi: reproduction, speciation and classification. Annu Rev Phytopathol 37:197-246.

235. **Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC** (2000) Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol 31:21-32.

236. **Taylor SS, Buechler JA, Yonemoto W** (1990) cAMP-dependent protein kinase: framework for a diverse family of regulatory enzymes. Annu Rev Biochem 59:971-1005.

237. **Thompson JD, Higgins DG, Gibson TJ** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.

238. **Tsong AE, Tuch BB, Li H, Johnson AD** (2006) Evolution of alternative transcriptional circuits with identical logic. Nature 443:415-420.

239. **Tsong AE, Tuch BB, Johnson AD** (2007) Rewiring transcriptional circuitry: mating-type regulation in *Saccharomyces cerevisiae* and *Candida albicans* as a model for evolution. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 75-89.

240. **Turgeon BG, Bohlmann H, Ciuffetti LM, Christiansen SK, Yang G, Schäfer W, Yoder OC** (1993) Cloning and analysis of the mating type genes from *Cochliobolus heterostrophus*. Mol Gen Genet 238:270-284.

241. **Turgeon BG, Yoder OC** (2000) Proposed nomenclature for mating type genes of filamentous ascomycetes. Fungal Genet Biol 31:1-5.

242. **Turgeon BG, Debuchy R** (2007) *Cochliobolus* and *Podospora*: mechanisms of sex determination and the evolution of reproductive lifestyle. In: Heitman J, Kronstad JW, Taylor JW, Casselton LA (eds) Sex in fungi: Molecular determination and evolutionary implications. ASM Press, Washington D.C., pp 93-121.

243. **Tymon AM, Kües U, Richardson WV, Casselton LA** (1992) A fungal mating type protein that regulates sexual and asexual development contains a POU-related domain. EMBO J 11:1805-1813.

244. **Tzung KW, Williams RM, Scherer S, Federspiel N, Jones T et al.** (2001) Genomic evidence for a complete sexual cycle in *Candida albicans*. Proc Natl Acad Sci U S A 98:3249-3253.

245. **Urban M, Kahmann R, Bölker M** (1996) Identification of the pheromone response element in *Ustilago maydis*. Mol Gen Genet 251:31-37.

246. **Valério E, Gadanho M, Sampaio JP** (2008) *Sporidiobolus johnsonii* and *Sporidiobolus salmonicolor* revisited. Mycol Prog 7:125-131.

247. **Velagapudi R, Hsueh YP, Geunes-Boyer S, Wright JR, Heitman J** (2009) Spores as infectious propagules of *Cryptococcus neoformans*. Infect Immun 77:4345-4355.

248. **Votintseva AA, Filatov DA** (2009) Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. Genetics 182:1391-1396.

249. **Votintseva AA, Filatov DA** (2010) DNA polymorphism in recombining and non-recombing mating-type-specific loci of the smut fungus *Microbotryum*. Heredity doi:10.1038/hdy.2010.140.

250. **Wahl R, Zahiri A, Kämper J** (2010) The *Ustilago maydis b* mating type locus controls hyphal proliferation and expression of secreted virulence factors in planta. Mol Microbiol 75:208-220.

251. **Wang L, Lin X** (2011) Mechanisms of unisexual mating in *Cryptococcus neoformans*. Fungal Genet Biol doi:10.1016/j.fgb.2011.02.001.

252. **Wang P, Nichols CB, Lengeler KB, Cardenas ME, Cox GM, Perfect JR, Heitman J** (2002) Mating-type-specific and nonspecific PAK kinases play shared and divergent roles in *Cryptococcus neoformans*. Eukaryot Cell 1:257-272.

253. **Weismann A** (1904) The evolution theory. Edward Arnold, London, UK.

254. **Whelan S, Goldman N** (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691-699.

255. **Whittle CA, Nygren K, Johannesson H** (2011a) Consequences of reproductive mode on genome evolution in fungi. Fungal Genet Biol doi:10.1016/j.fgb.2011.02.005.

256. **Whittle CA, Sun Y, Johannesson H** (2011b) Degeneration in codon usage within the region of suppressed recombination in the mating type chromosomes of *Neurospora tetrasperma*. Eukaryot Cell 10:594-603.

257. **Wright S** (1960) On the number of self-incompatibility alleles maintained in equilibrium by a given mutation rate in a population of a given size: a re-examination. Biometrics 16:61-85.

258. **Xin G, Glawe D, Doty SL** (2009) Characterization of three endophytic, indole-3-acetic acid-producing yeasts occurring in *Populus* trees. Mycol Res 113:973-980.

259. **Xu J, Saunders CW, Hu P, Grant RA, Boekhout T et al.** (2007) Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. Proc Natl Acad Sci U S A 104:18730-18735.

260. **Xue C, Hsueh YP, Heitman J** (2008) Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in fungi. FEMS Microbiol Rev 32:1010-1032.

261. **Yamazaki S, Katsuya K** (1988) Mating type of pine gall rust fungus, *Cronartium quercuum*. Proc Japan Acad 64 Ser. B:197-200.

262. **Yang Z** (1993) Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol 10:1396-1401.

263. **Yang Z** (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39:306-314.

264. **Yee AR, Kronstad JW** (1993) Construction of chimeric alleles with altered specificity at the *b* incompatibility locus of *Ustilago maydis*. Proc Natl Acad Sci U S A 90:664-668.

265. **Yee AR, Kronstad JW** (1998) Dual sets of chimeric alleles identify specificity sequences for the *bE* and *bW* mating and pathogenicity genes of *Ustilago maydis*. Mol Cell Biol 18:221-232.

266. **Yi R, Tachikawa T, Ishikawa M, Mukaiyama H, Bao D, Aimi T** (2009) Genomic structure of the *A* mating-type locus in a bipolar basidiomycete, *Pholiota nameko*. Mycol Res 113:240-248.

267. **Yi R, Mukaiyama H, Tachikawa T, Shimomura N, Aimi T** (2010) *A*-mating-type gene expression can drive clamp formation in the bipolar mushroom *Pholiota microspora* (*Pholiota nameko*). Eukaryot Cell 9:1109-1119.

268. **Yockteng R, Marthey S, Chiapello H, Gendrault A, Hood ME et al.** (2007) Expressed sequences tags of the anther smut fungus, *Microbotryum violaceum*, identify mating and pathogenicity genes. BMC Genomics 8:272.

269. **Yue C, Osier M, Novotny CP, Ullrich RC** (1997) The specificity determinant of the Y mating-type proteins of *Schizophyllum commune* is also essential for Y-Z protein binding. Genetics 145:253-260.

270. **Yun SH, Berbee ML, Yoder OC, Turgeon BG** (1999) Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. Proc Natl Acad Sci U S A 96:5592-5597.

271. **Zarnack K, Feldbrügge M** (2007) mRNA trafficking in fungi. Mol Genet Genomics 278:347-359.

272. **Zarnack K, Eichhorn H, Kahmann R, Feldbrügge M** (2008) Pheromone-regulated target genes respond differentially to MAPK phosphorylation of transcription factor Prf1. Mol Microbiol 69:1041-1053.

273. **Zarnack K, Feldbrügge M** (2010) Microtubule-dependent mRNA transport in fungi. Eukaryot Cell 9:982-990.

274. **Zerbino DR, Birney E** (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821-829.

275. **Zhao X, Murata T, Ohno S, Day N, Song J, Nomura N, Nakahara T, Yokoyama KK** (2001) Protein kinase Cα plays a critical role in mannosylerythritol lipid-induced differentiation of melanoma B16 cells. J Biol Chem 276:39903-39910.

APPENDIX I

Additional information pertaining to Chapters 2 and 3

I.1. GenBank accession numbers of sequences used in Figure 2.7

Table I.1. Sequence accession numbers of LSU rDNA (D1/D2 region) and ITS sequences of species used to construct the phylogenetic tree in Figure 2.7. Species are listed in alphabetical order. (Based on the data of Stoll et al., 2005).

Species name	Strain	D1/D2	ITS
Leucocintractia scleriae	MP 2074	AJ236154	AY740025
Macalpinomyces eriachnes	M 56574	AY740091	AY740038
Moesziomyces bullatus	Ust. Exs. 833	AY740153	
Pseudozyma abaconensis	CBS 8380	FJ008047	FJ008053
Pseudozyma alboarmeniaca	CBS 9961	AB117961	
Pseudozyma antarctica	JCM 10317	AB089361	AB089360
Pseudozyma aphidis	JCM 10318	AB089363	AB089362
Pseudozyma crassa	JCM 12455	AB117962	
Pseudozyma flocculosa	JCM 10321	AB089365	AB089364
Pseudozyma fusiformata	JCM 3931	AB089367	AB089364
Pseudozyma graminicola	CBS 10092	AB180728	
Pseudozyma hubeiensis	AS 2.2493	DQ008953	DQ008954
Pseudozyma jejuensis	CBS 10454	DQ866822	EF079966
Pseudozyma parantarctica	JCM 11752	AB089357	AB089356
Pseudozyma prolifica	CBS 319.87	AJ235298	AF294700
Pseudozyma pruni	CBS 10937	EU379943	EU379942
Pseudozyma rugulosa	JCM 10323	AB089371	AB089370
Pseudozyma shanxiensis	AS2.2523	DQ008955	DQ008956
Pseudozyma siamensis	CBS 9960	AB117963	
Pseudozyma thailandica	JCM 11753	AB089355	AB089354
Pseudozyma tsukubaensis	JCM 10324	AB089373	AB089372
Sporisorium aegypticum	Ust. Exs. 756	AY740129	AY344970
Sporisorium andropogonis	MP 2666	AY740096	AY740043
Sporisorium chrysopogonis	Ust. Exs. 407	AY740131	AY344973
Sporisorium consanguineum	H.U.V. 19145	AY740101	AY740048
Sporisorium elionuri	MP 2601	AY740157	
Sporisorium fastigiatum	MP 1976	AY740133	AY344978
Sporisorium holwayi	MP 1271	AF453941	AY344980

Table I.1. Continued.

Species name	Strain	D1/D2	ITS
Sporisorium mishrae	Ust. Exs. 967	AY740136	AY344983
Sporisorium modestum	M 56617	AY740054	AY740107
Sporisorium nealii	M 56621	AY740108	AY740055
Sporisorium neglectum	RB 2056	AY740109	AY740056
Sporisorium occidentale	Ust. Exs. 758	AY740137	AY344985
Sporisorium penniseti	MP 2367	AY740130	AY344971
Sporisorium polliniae	Ust Exs. 690	AY740138	AY344987
Sporisorium pseudechinolaenae	Ust. Exs. 853	AY740139	AY344989
Sporisorium pulverulentum	M56627	AY740162	
Sporisorium reilianum	Ust. Exs. 527	AY740163	
Sporisorium scitamineum	MP 541	AJ236138	AY740070
Sporisorium sorghi	MP 2036a	AF009872	AY740021
Sporisorium trachypogonis-plumosi	M 56635	AY740113	AY740060
Sporisorium tumefaciens	Ust. Exs. 231	AY740128	AY344969
Sporisorium veracruzianum	MP 735	AY740142	AY747075
Ustanciosporium taubertianum	MP 2276	AJ236156	AY740024
Ustilago austro-africana	M 56516	AY740115	AY740061
Ustilago avenae	RB 3092	AY740116	AY740062
Ustilago bouriquetii	M 56517	AY740167	
Ustilago bullata	MP 2363	AF453935	AY344998
Ustilago calamagrostidis	M 56518	AY740119	AY740065
Ustilago crameri	Ust. Exs. 995	AY740143	AY344999
Ustilago cynodontis	MS 1	AY740168	
Ustilago hordei	Ust. Exs. 784	AF453943	AY345003
Ustilago kolleri	F 947/ GD 1300	AY740122	AY740068
Ustilago maydis	RB 3093	AF453938	AY345004
Ustilago nuda	H.U.V. 17782	AJ236139	AY740069
Ustilago spermophora	F 565/ H.U.V. 13634	AY740171	
Ustilago striiformis	H.U.V. 18286	AY740172	
Ustilago trichophora	MP 2473	AY740148	AY345009
Ustilago vetiveriae	H.U.V. 17954	AY740149	AY345011
Ustilago xerochloae	Ust. Exs. 1000	AY740150	AY345012

I.2. RNA isolation and purification from *Pseudozyma* and *Ustilago* strains

RNA isolation protocol

1. Defrost tubes containing cells previously frozen in liquid nitrogen.
2. Centrifuge at 8000 rpm for 10 min, at 4ºC.
3. Transfer cells with a minimum volume of supernatant for 2 ml tubes containing sterile glass beads (∅ 0,45-0,55 µm).
4. Add 1 ml of TRIzol® (Invitrogen, Cat. No. 15596-026).
5. Disrupt cells in a vortex: 6 cycles of 30 s vortex + 30 s on ice.
6. Incubate at 30ºC for 5 min.
7. Add 0.2 ml of chloroform and shake vigorously by hand for 15 min (confirm that the tubes are firmly closed).
8. Incubate at 30ºC for 3 min.
9. Centrifuge at 12 000 × *g* for 15 min, at 4ºC.
10. Transfer the aqueous phase to a new sterile 1.5 ml tube (previously chilled on ice).
11. Add 0.5 ml of isopropanol.
12. Incubate at 30ºC for 10 min.
13. Centrifuge at 12 000 × *g* for 10 min, at 4ºC.
14. Remove the supernatant and wash pellet with 1 ml of 75% ethanol (-20ºC).
15. Centrifuge at 7500 × *g* for 15 min, at 4ºC.
16. Remove the supernatant and dry pellet at room temperature (do not allow pellet to completely dry out as it will affect its solubilization).
17. Dissolve the pellet with 40 µl of sterile and RNase-free water.
18. Quantify RNA on a UV spectrophotometer or equivalent.
19. Store RNA samples at -80ºC.

RNA purification protocol using the RNeasy® Mini kit (Quiagen, Cat. No. 74106)

1. Add 350 µl of RLT buffer to the RNA sample (maximum of 100 µg), add 1 volume of 70% ethanol and mix well by pipetting (add 10 µl of β-mercaptoethanol to 1 ml of RLT buffer immediately before using).
2. Transfer the sample to an RNeasy spin column placed in a 2 ml collection tube. Close the lid gently, and centrifuge at ≥8000 × *g* (≥10 000 rpm) for 15 s. Discard the flow-through.

On-Column DNase treatment (RNase-Free DNase Set, Quiagen, Cat. no. 79254)

3. Add 350 µl Buffer RW1 to the RNeasy spin column. Close the lid gently, and centrifuge at ≥8000 × *g* (≥10 000 rpm) for 15 s to wash the spin column membrane. Discard the flow-through.

4. Add 10 µl DNase I stock solution (prepared according to the manufacture instructions) to 70 µl of buffer RDD. Mix by gently inverting the tube, and centrifuge briefly to collect residual liquid from the sides of the tube.
5. Add the DNase I incubation mix (80 µl) directly to the RNeasy spin column membrane, and incubate at room temperature (20–30°C) for 15 min.
6. Add 350 µl of Buffer RW1 to the RNeasy spin column. Close the lid gently, and centrifuge at ≥8000 × *g* (≥10 000 rpm) for 15 s. Discard the flow-through and transfer column to a new 2 ml tube.
7. Add 500 µl of buffer RPE (with previously added ethanol) to the RNeasy spin column. Close the lid gently, and centrifuge at ≥8000 × *g* (≥10 000 rpm) for 15 s to wash the spin column membrane. Discard the flow-through.
8. Repeat step 7, but centrifuge for 2 min to dry completely the spin column membrane.
9. Transfer the column to a new 2 ml tube and repeat centrifugation for 1 min to remove any residuals of buffer RPE. Transfer the column to a new 1.5 ml tube.
10. Add 30 µl of RNase-free water directly to the spin column membrane. Close the lid gently, and centrifuge at ≥8000 × *g* (≥10 000 rpm) for 1 min to elute the RNA.
11. If the expected RNA yield is >30 µg, repeat step 10 using another 30 µl of RNase-free water.

I.3. NCBI Trace Archive sequences numbers used to obtain selected genomic regions of *Sporidiobolus salmonicolor* (IAM 12258 = CBS 483)

SsSTE20	***SsRHA1***	***SsRHA2***	***SsRHA3***
>gnl|ti| 1243964000	>gnl|ti| 1244034102	>gnl|ti| 1244010230	>gnl|ti| 1244013519
>gnl|ti| 1243931275	>gnl|ti| 1244025649	>gnl|ti| 1243911835	>gnl|ti| 1244014847
>gnl|ti| 1244008391	>gnl|ti| 1244026866	>gnl|ti| 1243919179	>gnl|ti| 1243996949
>gnl|ti| 1244012976	>gnl|ti| 1243949840	>gnl|ti| 1244012751	>gnl|ti| 1243918143
>gnl|ti| 1244008568		>gnl|ti| 1243952391	>gnl|ti| 1244015954
>gnl|ti| 1243932804		>gnl|ti| 1243997749	
>gnl|ti| 1243983780			
>gnl|ti| 1243949766			
>gnl|ti| 1244033153			
>gnl|ti| 1243947537			

I.4. GenBank accession numbers of the novel *Rhodosporidium toruloides* sequences described in Chapter 3 (see Figure 3.2)

RtSTE20.A1	PYCC 4416	Fragments 1 and 2 (Figure 3.2)	GenBank: **EU386160**
RtSTE20.A2 – RHA2.A2	PYCC 4661	Fragments 3 to 6 (Figure 3.2)	GenBank: **EU386161**
RtSTE20.A1	PYCC 5082	Fragments 3 and 4 (Figure 3.2)	GenBank: **EU401861**
RtSTE20.A2	PYCC 4417	Fragments 3 and 4 (Figure 3.2)	GenBank: **EU401862**
RtSTE3.A1	PYCC 4416	Partial *STE3.A1* sequence	GenBank: **EU401863**

APPENDIX II

Additional information pertaining to Chapter 4

II.1. NCBI Trace Archive sequences used to assemble the *STE3.A2* gene of *Rhodosporidium babjevae* (WP1) and the *HD1/HD2* region of both *Sporidiobolus salmonicolor* (IAM 12258 = CBS 483) and *R. babjevae*

R. babjevae

STE3.A2

>gnl|ti| 2154151284
>gnl|ti| 2195334063
>gnl|ti| 2195327270
>gnl|ti| 2154446490
>gnl|ti| 2154537287
>gnl|ti| 2154452092
>gnl|ti| 2154291912
>gnl|ti| 2154443469

S. salmonicolor

HD1/HD2

>gnl|ti| 1244034137
>gnl|ti| 1243919455
>gnl|ti| 1244020867
>gnl|ti| 1243931890
>gnl|ti| 1243965174
>gnl|ti| 1244003711
>gnl|ti| 1243998420
>gnl|ti| 1244025540
>gnl|ti| 1244020332
>gnl|ti| 1243943891
>gnl|ti| 1243937469
>gnl|ti| 1243958334
>gnl|ti| 1244003699
>gnl|ti| 1243967269
>gnl|ti| 1243996862
>gnl|ti| 1244031458
>gnl|ti| 1243974274
>gnl|ti| 1243951950
>gnl|ti| 1243993653

R. babjevae

HD1/HD2

>gnl|ti| 2154331804
>gnl|ti| 2154456123
>gnl|ti| 2154540940
>gnl|ti| 2154229020
>gnl|ti| 2195319826
>gnl|ti| 2154235755
>gnl|ti| 2195334783
>gnl|ti| 2154503699
>gnl|ti| 2195316925
>gnl|ti| 2154229021
>gnl|ti| 2154430327
>gnl|ti| 2195318461
>gnl|ti| 2154533013
>gnl|ti| 2154534672
>gnl|ti| 2154333225
>gnl|ti| 2195322807
>gnl|ti| 2195332082
>gnl|ti| 2154197503
>gnl|ti| 2154499347
>gnl|ti| 2195335759
>gnl|ti| 2154197502
>gnl|ti| 2154447297
>gnl|ti| 2154500783
>gnl|ti| 2154233232
>gnl|ti| 2154419698
>gnl|ti| 2154326269
>gnl|ti| 2154367586
>gnl|ti| 2154430851
>gnl|ti| 2154233233
>gnl|ti| 2154343709
>gnl|ti| 2154368702
>gnl|ti| 2154321820
>gnl|ti| 2154418949

II.2. NCBI Trace Archive sequences used to assemble eight genes from the pheromone receptor region and four genes from the *HD1/HD2* region in *S. salmonicolor* (IAM 12258 = CBS 483)

Genes at the pheromone receptor region

PAN6

(pantoate-beta-alanine ligase)

>gnl|ti| 1243999491
>gnl|ti| 1243944265
>gnl|ti| 1243939898
>gnl|ti| 1244024936
>gnl|ti| 1243921720

STE20

(p21-activated protein kinase)

>gnl|ti| 1243947537
>gnl|ti| 1243983780
>gnl|ti| 1244033153
>gnl|ti| 1243931275
>gnl|ti| 1243964000
>gnl|ti| 1244010134
>gnl|ti| 1244008391
>gnl|ti| 1244012976
>gnl|ti| 1243932804
>gnl|ti| 1244008568
>gnl|ti| 1243986237
>gnl|ti| 1243931275

KAP95

(Karyopherin β 1 - importin)

>gnl|ti| 1243914633
>gnl|ti| 1243945047
>gnl|ti| 1244013243
>gnl|ti| 1244003594
>gnl|ti| 1243947825
>gnl|ti| 1244019354
>gnl|ti| 1244010118
>gnl|ti| 1244023708
>gnl|ti| 1243943827
>gnl|ti| 1244005440

LSm7

(small nuclear ribonucleoprotein)

>gnl|ti| 1243966515
>gnl|ti| 1244006137
>gnl|ti| 1244019455
>gnl|ti| 1244006358
>gnl|ti| 1243968129
>gnl|ti| 1244035821

RibL18ae

(Ribosomal L18ae protein)

>gnl|ti| 1244029876
>gnl|ti| 1244027010
>gnl|ti| 1243938823
>gnl|ti| 1244024596

RNAPOL

(DNA directed RNA pol. III, 30/40 KDa subunit)

>gnl|ti| 1243928029
>gnl|ti| 1244017874
>gnl|ti| 1243938919
>gnl|ti| 1244027106
>gnl|ti| 1244002511

GPD

(glyceraldehyde-3-phosphate dehydrogenase)

>gnl|ti| 1244012762
>gnl|ti| 1243987204
>gnl|ti| 1243961274

MIP

(mitochondrial intermediate peptidase)

>gnl|ti| 1243979024
>gnl|ti| 1243993537
>gnl|ti| 1244034662
>gnl|ti| 1244008528
>gnl|ti| 1243938225
>gnl|ti| 1243918485
>gnl|ti| 1243959176
>gnl|ti| 1243964071

Genes at the *HD1/HD2* region

IsocL **(isocitrate lyase)**	***NGP1*** (*NOG2*) **(Nucleolar GTP-binding protein 2)**	***AKOR2*** **(aldo-keto reductase)**	***RPB2*** **(RNA pol. II second largest subunit)**
>gnl|ti| 1244032592	>gnl|ti| 1243932700	>gnl|ti| 1244016242	>gnl|ti| 1243971757
>gnl|ti| 1243913568	>gnl|ti| 1244009264	>gnl|ti| 1243951077	>gnl|ti| 1244020882
>gnl|ti| 1243931944	>gnl|ti| 1244000123	>gnl|ti| 1243949884	>gnl|ti| 1243983637
>gnl|ti| 1243952262	>gnl|ti| 1243941162	>gnl|ti| 1244024160	>gnl|ti| 1243920596
>gnl|ti| 1243963834	>gnl|ti| 1244019186	>gnl|ti| 1244021707	>gnl|ti| 1243971149
>gnl|ti| 1243965732	>gnl|ti| 1243984108	>gnl|ti| 1244024160	>gnl|ti| 1243965032
>gnl|ti| 1243981027	>gnl|ti| 1243948626	>gnl|ti| 1243966759	>gnl|ti| 1243961708
>gnl|ti| 1243971150	>gnl|ti| 1244029720	>gnl|ti| 1243913061	>gnl|ti| 1244035646
	>gnl|ti| 1243918656	>gnl|ti| 1243949884	>gnl|ti| 1243917934
	>gnl|ti| 1244014951		>gnl|ti| 1243920588
	>gnl|ti| 1243966796		>gnl|ti| 1243922988
	>gnl|ti| 1243966495		>gnl|ti| 1243938766
	>gnl|ti| 1243933163		>gnl|ti| 1243961708
	>gnl|ti| 1244034658		
	>gnl|ti| 1243921840		

II.3. NCBI Trace Archive sequences used to assemble nine *S. salmonicolor* genes that are located in four different scaffolds of the *Sporobolomyces roseus* genome

sdhA **(succinate dehydrogenase flavoprotein subunit)**	***URA3*** **(orotidine 5'-phosphate decarboxylase)**	***HXT1*** **(hexose transporter)**
>gnl|ti| 1243932965	>gnl|ti| 1244022690	>gnl|ti| 1243923129
>gnl|ti| 1243949976	>gnl|ti| 1244004916	>gnl|ti| 1243923641
>gnl|ti| 1244011373	>gnl|ti| 1243996898	>gnl|ti| 1244007487
>gnl|ti| 1243982038	>gnl|ti| 1243942104	>gnl|ti| 1243978746
>gnl|ti| 1244024325	>gnl|ti| 1243939980	>gnl|ti| 1244008309
	>gnl|ti| 1243942104	>gnl|ti| 1243932413
	>gnl|ti| 1244011207	>gnl|ti| 1244025679
		>gnl|ti| 1244028271
		>gnl|ti| 1243928371
		>gnl|ti| 1243949421
		>gnl|ti| 1243981984
		>gnl|ti| 1244014361
		>gnl|ti| 1243923622
		>gnl|ti| 1243953511
		>gnl|ti| 1243935956
		>gnl|ti| 1243943748
		>gnl|ti| 1244021128
		>gnl|ti| 1243914662
		>gnl|ti| 1243923263
		>gnl|ti| 1244019696
		>gnl|ti| 1243916660
		>gnl|ti| 1243927197

RAN1

(protein serine/ threonine kinase)

>gnl|ti| 1243988647
>gnl|ti| 1244016377
>gnl|ti| 1244006386
>gnl|ti| 1244005972
>gnl|ti| 1243915389
>gnl|ti| 1243988647

aldA

(NAD-aldehyde dehydrogenase)

>gnl|ti| 1244020809
>gnl|ti| 1244016587
>gnl|ti| 1243935797
>gnl|ti| 1243978923
>gnl|ti| 1243958883
>gnl|ti| 1243988822
>gnl|ti| 1243930459
>gnl|ti| 1244032672
>gnl|ti| 1244033401
>gnl|ti| 1244014959
>gnl|ti| 1243978923
>gnl|ti| 1243986112
>gnl|ti| 1244010416
>gnl|ti| 1243964402
>gnl|ti| 1244032672
>gnl|ti| 1244033401
>gnl|ti| 1244014959
>gnl|ti| 1243978923
>gnl|ti| 1243986112
>gnl|ti| 1244010416
>gnl|ti| 1243964402

PAL

(hexose transporter)

>gnl|ti| 1244023897
>gnl|ti| 1243917414
>gnl|ti| 1244028263
>gnl|ti| 1243999288
>gnl|ti| 1243988576
>gnl|ti| 1243941753
>gnl|ti| 1243997975
>gnl|ti| 1243979468

DMC1

(meiosis-specific recombinase)

>gnl|ti| 1244026568
>gnl|ti| 1243954225
>gnl|ti| 1244015333
>gnl|ti| 1244030098

GEF1

(phenylalanine ammonia-lyase)

>gnl|ti| 1243992869
>gnl|ti| 1243919507
>gnl|ti| 1243920956
>gnl|ti| 1243975744
>gnl|ti| 1243981595
>gnl|ti| 1243982772
>gnl|ti| 1244026095
>gnl|ti| 1243982553

LACC

(laccase protein)

>gnl|ti| 1244008544
>gnl|ti| 1243941934
>gnl|ti| 1243992777
>gnl|ti| 1243956635
>gnl|ti| 1244018080
>gnl|ti| 1243928557
>gnl|ti| 1244013399
>gnl|ti| 1243954711

APPENDIX III

Additional information pertaining to Chapter 5

III.1. List of species/ strains used in Chapter 5.

Table III.1. List of species/strains used in Chapter 5 and relevant information pertaining to them. GenBank accessions numbers are given for each genomic region analysed. Sequences retrieved from the CBS culture collection database ('CBS seq.') or from genome projects ('JGI genome') are indicated. Mating behaviour as originally determined by classical mating tests ('old') and as now reassigned based on the molecular data ('new') is indicated when needed: A1, A2, As and Sf stands for mating type A1, A2, asexual and self-fertile strains, respectively. The "molecular mating type" as identified by PCR detection of the pheromone receptor alleles *STE3.A1* and *STE3.A2* is depicted by yellow and blue circles, respectively. White circles indicate strains for which PCR detection produced negative results and 'n.d.' stands for 'not determined'. Strains highlighted in boldface were used in Figure 5.1 and those indicated by thick-lined blue or yellow circles were used in Figure 5.2. Sequence accession numbers of the *HD1/HD2* alleles used in Figure 5.7 are highlighted in boldface. Abbreviations: *R.*, *Rhodosporidium*; *Rh.*, *Rhodotorula*; *S.*, *Sporidiobolus*; *Sp.*, *Sporobolomyces*; T, Type strain; LT, Lectotype, AUT, Authentic strain. Undescribed species included in this study are shadowed in light yellow and those for which no molecular mating type was accessed are shadowed in gray.

Species	**Strains**	**D1/D2 LSU** (GenBank)	**ITS region** (GenBank)	**Mating behaviour**		***STE3 A1***	***STE3 A2***	***STE3.A1*** (GenBank)	***STE3.A2*** (GenBank)	***HD1/HD2*** (GenBank)	**Isolation source and location**
				Old	New						
R. babjevae	A122			As		●				**HM143851**	Seawater, South of Portugal
R. babjevae	A125			As		●					Seawater, South of Portugal
R. babjevae	A267			As		●					Seawater, South of Portugal
R. babjevae	A270			As		●					Seawater, South of Portugal
R. babjevae	CBS 322			As		●					Water supply of brewery, Germany
R. babjevae	CBS 2366			As			●			**JN246619**	Air, Japan
R. babjevae	CBS 6020			As		●				**JN246620**	Soil from drill core, Japan
R. babjevae	**CBS 7808^{T}**	AF070420	AF444542	A1		●		JN246636		**HM143857**	Herbaceous plant, Moscow, Russia
R. babjevae	**CBS 7809**	AF389828	AF387773	A2			●		JN246573	**HM143856**	Herbaceous plant, Moscow, Russia
R. babjevae	CBS 9071			A2			●			**JN246621**	Barley kernels, North Dakota, USA
R. babjevae	CBS 9072			As			●		JN246575	HM143852	Barley kernels, North Dakota, USA
R. babjevae	CRUB 1025			A1		●				**JN246614**	Nahuel Huapi, Patagonia, Argentina
R. babjevae	CRUB 1033			Sf			●			JN246615	Verde Lagoon, Patagonia, Argentina
R. babjevae	CRUB 1034			As		●				JN246616	Verde Lagoon, Patagonia, Argentina
R. babjevae	CRUB 1037			As		●				**HM143854**	Fonk Lake, Patagonia, Argentina
R. babjevae	CRUB 1113			As			●				Verde Lagoon, Patagonia, Argentina
R. babjevae	PTZ 001			A1		●		JN246641			Herbaceous plant, Prioksko-terrasny reserve, Moscow, Russia

Species	Strains	D1/D2 LSU (GenBank)	ITS region (GenBank)	Mating behaviour		*STE3 A1*	*STE3 A2*	*STE3.A1* (GenBank)	*STE3.A2* (GenBank)	*HD1/HD2* (GenBank)	Isolation source and location
				Old	New						
R. babjevae	PTZ 111			A1		●					Herbaceous plant, Prioksko-terrasny reserve, Moscow, Russia
R. babjevae	PYCC 4651			As		●					Leaf, Estoril, Portugal
R. babjevae	PYCC 4777			As		●		JN246638		**HM143855**	Flower, Oeiras, Portugal
R. babjevae	**PYCC 4915**	AF389825	AF387770	As		●		JN246639		**HM143853**	Pond water, Silwood, UK
R. babjevae	VKM Y-1310			A1		●		JN246640		**JN246617**	Silage, Kiev region, Ukraine
R. babjevae	VKM Y-1629			A2			●			**JN246618**	Soil, Krasnodar region, Russia
R. babjevae	VKM Y-1631			A2			●				Soil, Krasnodar region, Russia
R. babjevae	VKM Y-2911			Sf		●	●	JN246637	JN246574		Soil, Komi Republic, Russia
Rh. graminis	AJ 130			As			●		JN246578	**JN246634**	Soil in *Nothofagus* forest, Tongario National Park, New Zealand
Rh. graminis	AJ 132			As			●			**JN246635**	Soil in *Podocarpus* forest, Pureora Forest Park, New Zealand
Rh. graminis	**CBS 2826^{T}**	AF070431	AF444505	As			●		JN246576	**JN246632**	Grass, New Zealand
Rh. graminis	**CBS 3043**	AF387130	AF387780	As			●		JN246577	**JN246633**	Leaf of *Citrus* sp., Indonesia
Rh. graminis	WP1			n.d.			●		JGI Genome	**JGI Genome**	Isolated as an endophyte within *Populus trichocarpa* (black cottonwood), Western Washington state, USA
Rh. glutinis	**C13A1**	JN246532	JN246543	As		●		JN246643		**JN246631**	Fruiting body of *Cyttaria hariotii,* Patagonia, Argentina
Rh. glutinis	**CBS 20^{T}**	AF070430	AF387775	As		●		JN246644		JN246626	Air, unknown location
Rh. glutinis	CBS 2367			As		●		JN246647		**JN246627**	Fruit of *Phyllodendron* sp., Botanic Garden, Florence, Italy
Rh. glutinis	**JCM 3907**	DQ531948	AB073237	As		●		JN246642		**JN246628**	Unknown
Rh. glutinis	PYCC 4107			As		●		JN246646		JN246629	Seawater, Florida, USA
Rh. glutinis	**PYCC 5599**	AF387132	AF387777	As		●		JN246645		JN246630	Core samples from stratigraphic drillings, Japan
R. diobovatum	CBS 324			A1		●		JN246649		**JN246623**	Unknown substrate, Portugal
R. diobovatum	**CBS 6084**	AF189914	AB073236	A2			●		JN246579	**JN246624**	Seawater, South-eastern barrier reef, Florida, USA
R. diobovatum	**CBS 6085^{T}**	AF070421	AF387782	A1		●		JN246648		**JN246625**	Seawater, South-eastern barrier reef, Florida, USA
R. diobovatum	PYCC 2920			A2			●		JN246580	**JN246622**	Seawater, Scripps pier, USA
"Rh. pinicorticis"	**PYCC 4826^{T}**	AF387134	JN246544	As		○	○				Bark of *Pinus pinaster*, Sesimbra, Portugal
R. kratochvilovae	A113			As			●				Under thermocline, Algarve, Portugal
R. kratochvilovae	CBS 7293			Sf		●	●				Unknown
R. kratochvilovae	**CBS 7436^{T}**	AF071436	AF444520	Sf		●	●	JN246651	JN246584		Unknown

Species	Strains	D1/D2 LSU (GenBank)	ITS region (GenBank)	Mating behaviour		*STE3 A1*	*STE3 A2*	*STE3.A1* (GenBank)	*STE3.A2* (GenBank)	*HD1/HD2* (GenBank)	Isolation source and location
				Old	New						
R. kratochvilovae	CRUB 0014			As		●		JN246652			*Rubus idaeus*, Patagonia, Argentina
R. kratochvilovae	CRUB 0121			A2			●				Lake Toncek, Argentina
R. kratochvilovae	CRUB 1035			A2			●		JN246583		Lake Escondido, Argentina
R. kratochvilovae	PYCC 4776			A2			●				Plant leaf, Oeiras, Portugal
R. kratochvilovae	PYCC 4785			A2			●		JN246585		Humus, Arrábida, Portugal
R. kratochvilovae	**PYCC 4818**	AF1899116	AF444586	A1		●		JN246653			Wood chips, Setúbal, Portugal
R. kratochvilovae	**PYCC 4819**	AF189917	AF444587	A2			●		JN246581		Moss, Oeiras, Portugal
R. kratochvilovae	PYCC 5068			A2			●				Soil, Azores, Portugal
R. kratochvilovae	PYCC 5085			A2			●		JN246582		Soil, Azores, Portugal
R. kratochvilovae	PYCC 5244			A1		●		JN246650			Culture contaminant, Oeiras, Portugal
Rh. araucariae	**CBS 6031^{T}**	AF070427	AF444510	As			●		JN246586		Rotting bark of *Araucaria araucana* (monkey-puzzle tree) Llaima, Chile
Rh. araucariae	CBS 7797			As			●		JN246587		Unknown
Rh. araucariae	**PYCC 4820**	AF387135	JN246545	As		●	●	JN246654	JN246588		Plant litter, Oeiras, Portugal
Rh. araucariae	PYCC 4821			As			●				Humus, Arrábida Natural Park, Portugal
Rh. araucariae	ZP 280			As			●				Plant litter, Sintra, Portugal
"Rh. hamamotoiana"	**PYCC 4724**	AF387137	JN246547	As			●		JN246589		Soil, Alentejo, Portugal
"Rh. hamamotoiana"	**PYCC 4824^{T}**	AF387138	JN246546	As		●		JN246656			Rotten wood, Gerês, Portugal
"Rh. sesimbrana"	**PYCC 4691^{T}**	AF387144	JN246548	As		●		JN246657			Bark of *Pinus pinaster*, Sesimbra, Portugal
R. paludigenum	CBS 4477			A2[(a)]		●		JN246658			Sea-water, Biscayne Bay, Florida, USA
R. paludigenum	CBS 6565			A1[(a)]			●		JN246591		*Rhizophora mangle* (mangrove), swamp, Southern Florida, USA
R. paludigenum	**CBS 6566^{T}**	AF363640	AF363640	A2			●		JN246590		*Juncus roemerianus* (black rush), marsh, Florida, USA
R. paludigenum	**CBS 6567**	AF070424	AF444493	A1		●		JN246659			*Rhizophora mangle* (mangrove), swamp, Southern Florida, USA
[(a)] The molecular mating type of strains CBS 4477 and CBS 6565 is not in agreement with the original mating type designation. However, when crossing CBS 6566^{T} and CBS 4474 a positive result was obtained, indicating that the analysed strains are in fact of opposite mating types. A confirmation of this result is needed for the original strains deposited at the CBS culture collection.											
R. toruloides	A405			A2			●		JN246596		Rio Tinto, Spain
R. toruloides	A421			A1		●					Rio Tinto, Spain
R. toruloides	**CBS 14**	AF207884	AB049028	A1		●		JN246660			Wood pulp from Coniferae, Sweden

Species	Strains	D1/D2 LSU (GenBank)	ITS region (GenBank)	Mating behaviour		*STE3 A1*	*STE3 A2*	*STE3.A1* (GenBank)	*STE3.A2* (GenBank)	*HD1/HD2* (GenBank)	Isolation source and location
				Old	New						
R. toruloides	CBS 315			A1		●		JN246663			Tokyo, Japan (type strain of *Rhodotorula rubescens*)
R. toruloides	CBS 350			A1		●					Wood pulp from conifer, Sweden
R. toruloides	CBS 5745			A2			●		JN246594		Soil, Nishiyama oil field, Japan
R. toruloides	**CBS 6016^{T}**	CBS seq.	CBS seq.	Sf		●	●	JN246662	JN246595		Unknown, [Hybrid]: cross of IFO 0559 x IFO 0880 (CBS 14 X CBS 349)
R. toruloides	PYCC 4350			A2			●				Unknown
R. toruloides	PYCC 4661			A2			●				Woodland soil, Alentejo, Portugal
R. toruloides	PYCC 4786			A2			●				Dry leaf, Jardim Botânico, Rio de Janeiro, Brasil
R. toruloides	PYCC 4943			A2			●				Pond of solar saltern, Alcochete, Portugal
R. toruloides	**PYCC 5081**	JN246533	JN246549	A2			●		JN246593		Soil, Azores, Portugal
R. toruloides	PYCC 5082			A1		●		JN246661			Soil, Azores, Portugal
R. toruloides	PYCC 5109			A2			●		JN246592		Salt farm, Alcochete, Portugal
R. toruloides	SDY 231			A1		●					Acidic mine water, S. Domingos, Portugal
Rh. dairenensis	**CBS 2202**	JN246535	JN246550	As		○					Soil, USA
Rh. dairenensis	**CBS 4406^{LT}**	AF070429	AF444501	As			●		JN246597		Air, unknown location
Rh. dairenensis	CBS 7294			As		○					Unknown
Rh. dairenensis	**PYCC 4784**	AY033551	AF444683	As			●		JN246598		Plant litter, Arrábida, Portugal
Rh. dairenensis	PYCC 4897			As			●				Flower of *Canna indica*, Azores, Portugal
Rh. dairenensis	PYCC 4944			As			●				Salt farm, Alcochete, Portugal
Rh. dairenensis	PYCC 5601			As		○					Core samples from stratigraphic drillings, Japan
Rh. mucilaginosa	**CBS 17**	AF189960	AF444503	As		○	○				Unknown
Rh. mucilaginosa	**CBS 316^{T}**	AF070432	AF444541	As		○	○				Unknown
Rh. mucilaginosa	CBS 325			As		○	○				Air, Tokyo, Japan
Rh. mucilaginosa	CBS 326			As		○	○				Air, unknown location
Rh. mucilaginosa	CBS 992			As		○	○				Nail of 12-year-old girl, Austria
Rh. mucilaginosa	CBS 1011			As		○	○				Ulcer of man, Madagascar
Rh. mucilaginosa	CBS 2377			As		○	○				Skin, scrapings from axillary region of man,
Rh. mucilaginosa	CBS 2404			As		○	○				Fermenting Kentucky tobacco, Salerno, Italy

Species	Strains	D1/D2 LSU (GenBank)	ITS region (GenBank)	Mating behaviour		*STE3 A1*	*STE3 A2*	*STE3.A1* (GenBank)	*STE3.A2* (GenBank)	*HD1/HD2* (GenBank)	Isolation source and location
				Old	New						
Rh. mucilaginosa	CBS 5804			As		○	○				Larvae of *Drosophila pilimanae* (fruit fly), Hawaii, USA
Rh. mucilaginosa	CBS 5951			As		○	○				Gut of *Guara rubra* (red ibis bird), Paris Zoological Park, France
Rh. pacifica	**SY-246**	AB193175	AB193175	A1		○	○				Sediments, deep-sea floor (900 m), Iheya Ridge, Pacific Ocean
Rh. pacifica	**SY-96**^{T}	AB026006	AB026006	A2		○	○				Sediments, deep-sea floor (3700 m), Yap Trench, Pacific Ocean
R. sphaerocarpum	**CBS 5939**^{T}	AF070425	AF444499	A2	A1	○	○				Seawater, Marguerite Bay, Antarctica
R. sphaerocarpum	CBS 5940			A1	A2	○	○				Seawater, Gerlache Straits near Anvers Island, Antarctica
R. sphaerocarpum	**CBS 5941**	AF189919	AB073259	Sf		○	○				Sea water, Gerlache Straits near Brabant Island, Antarctica
R. sphaerocarpum	CBS 6985			A2	A1	○	○				Salt farm, Seto Inland Sea, Yoshima, Japan
R. sphaerocarpum	**PYCC 2988**	JN246534	JN246551	A2[(b)]	A1	◐		JN246664			Unknown
R. sphaerocarpum	PYCC 5105			A2	A1	○	○				Salt farm, Alcochete, Portugal
R. sphaerocarpum	PYCC 5106			A2	A1	○	○				Salt farm, Alcochete, Portugal
R. sphaerocarpum	PYCC 5107			A1	A2	○	○				Salt farm, Alcochete, Portugal
R. sphaerocarpum	PYCC 5108			A1	A2	○	○				Salt farm, Alcochete, Portugal
R. sphaerocarpum	ZP 417			A2	A1	○	○				Mud from salt marsh, Tróia, Portugal
[(b)] The molecular mating type of strain PYCC 2988 indicates that, despite originally designated as MAT A2, in fact is a MAT A1 strain. However, no positive results were obtained for the other strains of the same mating type.											
Rh. taiwanensis	BCRC 23118^{T}	GU646863	GUG46862			n.d.	n.d.				
"Rh. pyritica"	PYCC 4782			As		○	○				Wood of *Quercus suber*, Arrábida, Portugal
"Rh. pyritica"	**PYCC 5597**	AF387139	JN246552	As		○	○				Unknown (IAM 12959)
"Rh. pyritica"	RT 4.5.3			As		○	○				Acidic water, Rio Tinto, Spain
"Rh. pyritica"	**SDY 119**^{T}	AY731799	JN246553	As		○	○				Acid mine water, S. Domingos, Portugal
"R. feracious"	**PYCC 5380**	AF387143	JN246554	As		○	○				Soil, Azores, Portugal
"R. feracious"	**SDY 020**^{T}	AY731800	JN246555	Sf		○	○				Acidic mine water, S. Domingos, Portugal
"R. feracious"	SDY 080			As		○	○				Acidic mine water, S. Domingos, Portugal
"R. feracious"	SDY 176			Sf		○	○				Acidic mine water, S. Domingos, Portugal
R. azoricum	**CBS 4648**	**DQ531944**	JN246556	A1	A2		●		JN246600		Cacao, unknown location
R. azoricum	**CBS 8948**	AF321978	CBS Seq.	A2	A1	◐		JN246655			Soil near thermal hot spring, São Miguel island, Azores, Portugal
R. azoricum	**CBS 8949**^{T}	AF321977	CBS Seq.	A1	A2		●		JN246599		Soil, São Miguel island, Azores, Portugal
R. fluviale	**CBS 6568**^{T}	AF189915	AY015432	Sf			●		JN246601		Brackish water, near mouth of Miami river, Florida, USA

Species	Strains	D1/D2 LSU (GenBank)	ITS region (GenBank)	Mating behaviour		*STE3 A1*	*STE3 A2*	*STE3.A1* (GenBank)	*STE3.A2* (GenBank)	*HD1/HD2* (GenBank)	Isolation source and location
				Old	New						
S. microsporus	**CBS 7041^{T}**			Sf			●		JN246602		Herbaceous culm, Jamaica
S. ruineniae	**CBS 5004**	AF387127	JN246558	A1	A2	○	○				Foliage of cacao tree (*Theobroma cacao*), Bogor, Indonesia
S. ruineniae	CBS 5078			As		○	○				Foliage of cacao tree (*Theobroma cacao*), Bogor, Indonesia
S. ruineniae	**CBS 5811**	AF070434	AB030339	As		○	○				Dung of goat, Pakistan
S. ruineniae	**CBS 9111**	AF387128	JN246557	A2	A1	●		JN246665			Garden soil, Taiwan
Sp. odoratus	**CBS 9115**	AF387125	CBS seq.	As		○	○				Basidiocarp of *Myxarium nucleatum*, Sesimbra, Portugal
Sp. odoratus	**ZP 470**	JN246540	JN246559	As		○	○				Culture contaminant, Caparica, Portugal
Sp. poonsookiae	**CBS 9095^{T}**	AF387124	AB030327	As		n.d.	n.d.				Dead leaf of *Mangifera indica*, Bangkok, Thailand
Sp. nylandii	**CBS 9093^{T}**	AF387123	AB030323	As		n.d.	n.d.				Dead leaf of *Oryza sativa*, Bangkok, Thailand
"*R. oreadum*"	**ZP 292^{T}**	JN246537	JN246560	A1		●		JN246666			Soil under snow, Serra da Estrela, Portugal
"*R. oreadum*"	**ZP 295**	JN246538	JN246561	A2		○	○				Soil under snow, Serra da Estrela, Portugal
Rh. colostri	A326			As			●		JN246605		Olo river, Portugal
Rh. colostri	C18B1			As			●				Fruiting body of *Cyttaria hariotii,* Patagonia, Argentina
Rh. colostri	**C21C1**	JN246536	JN246562	As		●		JN246667			Fruiting body of *Cyttaria hariotii,* Patagonia, Argentina
Rh. colostri	**CBS 348^{T}**	AY372177	JN246563	As			●		JN246603		Colostrum from woman
Rh. colostri	CBS 2201			As			●		JN246604		Unknown
Rh. colostri	CBS 9002			As		●		JN246668			Flowering plant of *Helleborus foetidus*, Marburg, Germany
Rh. colostri	ZP 970			As			●				Leaves of *Nothofagus moorei*, Lamington National Park, Australia
Rh. colostri	ZP 984			As			●				*Cyttaria gunnii* on *Nothofagus cunninghamii*, Mt. Field Park, Tasmania
Rh. colostri	ZP 985			As			●				*Cyttaria gunnii* on *Nothofagus menziesii*, Lake Daniels, New Zealand
Rh. colostri	ZP 986			As			●				*Cyttaria gunnii* on *Nothofagus menziesii* (floor), Hasst Pass, New Zealand
R. lusitaniae	**CBS 7604^{T}**	AF070423	AY015430	Sf			●		JN246606		Soil of woodland, Lisbon, Portugal
R. lusitaniae	PYCC 4599			Sf			●		JN246607		Dried leaf material, Arrábida Natural Park, Portugal
R. lusitaniae	**PYCC 4642**	JN246539	AB087497	Sf			●		JN246608		Dried leaf material, Arrábida Natural Park, Portugal
Sp. ruberrimus	**CBS 7500^{AUT}**	AF070442	AY070004	As			●		JN246609		Air, Fukuoka, Japan
Sp. ruberrimus	**CRUB 1041**	JN246541	JN246564	As		●		JN246669			Aquatic environment, Patagonia, Argentina
Sp. phaffii	**CBS 9129^{T}**	AY070011	AY069995	As		n.d.	n.d.				Wilting leaf of *Nerium indicum*, Yunnan, China

Species	Strains	D1/D2 LSU (GenBank)	ITS region (GenBank)	Mating behaviour		*STE3 A1*	*STE3 A2*	*STE3.A1* (GenBank)	*STE3.A2* (GenBank)	*HD1/HD2* (GenBank)	Isolation source and location
				Old	New						
Sp. phaffii	**CH 2.049**	AY070010	AY069994	As		n.d.	n.d.				Wilting leaf of *Ehretia corylifolia*, Yunnan, China
Sporobolomyces sp.	**IAM 13481**	JGI Genome	JGI Genome	n.d.		●		JGI Genome			Leaf of willow, Japan (identified as *Sporobolomyces roseus*)
S. pararoseus	**CBS 484**	AF070437	AF417115	A1	A2		●		JN246610		Air, USA
S. pararoseus	**CBS 491^{T}**	AF189977	AY015429	A2	A1	●		JN246670			Soil, Japan
S. pararoseus	CBS 499			A1	A2		●		JN246611		Air in dairy, USA
S. pararoseus	**CBS 7716**	EU003452	JN246565	A2	A1	●		JN246671			Soil, USSR
Sp. patagonicus	**CBS 9657^{T}**	AY158655	AY552328	As		n.d.	n.d.				Subsurface water, lake Fonck, Patagonia, Argentina
Sp. patagonicus	**CBS 9658**	AY158656	AY552329	As		n.d.	n.d.				Subsurface water, lake Hess, Patagonia, Argentina
Sp. salmoneus	**CBS 488^{T}**	AY070017	AY070005	As		n.d.	n.d.				Etiolated grass, under a wooden plank, Delft, The Netherlands
Sp. salmoneus	**CBS 4217^{T}**	AF070440	AY069993	As		n.d.	n.d.				Air, unknown location
Sp. carnicolor	**CBS 4215^{AUT}**	AY158641	AY069991	As		n.d.	n.d.				Unknown
Sp. japonicus	**CBS 5744^{T}**	AY070009	AY069992	As		n.d.	n.d.				Oil brine, Yabase oil field, Akita Prefecture, Japan
Sp. bannaensis	**CBS 9204^{T}**	AY274823	AY274824	As		n.d.	n.d.				Leaf of *Theobroma cacao*, Xishuang Banna, Yunnan, China
Sp. beijingensis	**CBS 9730^{T}**	AY364837		As		n.d.	n.d.				Wilting leaf of *Sorbus pohuashanensis*, Baihua Mountain, Beijing, China
S. longiusculus	**CBS 9654^{T}**	AY158657	AY552327	A1	A2		●		JN246612		Subsurface water, Fonck lake, Patagonia, Argentina
S. longiusculus	**CBS 9655**	AY552326	JN246566	A2	A1	○					Subsurface water, Ilon lake, Patagonia, Argentina
S. longiusculus	CBS 9656			A2	A1	○					Subsurface water, Ilon lake, Patagonia, Argentina
Sp. blumeae	**CBS 9094^{T}**	AY070007	AB030331	As		●		JN246672			Dead leaf of *Blumea* sp., Bangkok, Thailand
S. metaroseus	**CBS 5541**	EU003458	EU003480	Sf		●		JN246673			Flower of *Fumaria* sp., France
S. metaroseus	**CBS 7683^{T}**	EU003461	EU003482	Sf		○	○				Leaves, Portugal
Sp. jilinensis	**CBS 9728^{T}**	AY364838		As		●		JN246674			Wilting leaf of *Pinus koraiensis*, Changbai Mountain, Jilin province, China
S. johnsonii	CBS 1522			As	As	◐				GU474690	Fodder yeast, Germany
S. johnsonii	CBS 2630			A2	A1	◐					Air, The Netherlands
S. johnsonii	**CBS 2634**	JN246542	JN246567	A1	A2		◐		GU474647	GU474685	*Fragaria sp.*, Japan
S. johnsonii	CBS 2643			A2	A1	◐					Unknown substrate, Germany
S. johnsonii	CBS 4209			A2	A1	◐					Fruit body of *Exidia* sp., Japan
S. johnsonii	**CBS 5470^{T}**	AF070435	AY015431	Sf		●	◐	GU474645	GU474648	GU474692	Leaf of *Rubus idaeus* with dead pustule of *Phragmidium rubi-idaei*, USA

Species	Strains	D1/D2 LSU (GenBank)	ITS region (GenBank)	Mating behaviour		*STE3 A1*	*STE3 A2*	*STE3.A1* (GenBank)	*STE3.A2* (GenBank)	*HD1/HD2* (GenBank)	Isolation source and location
				Old	New						
S. johnsonii	**CBS 7795**	EF592139	EF592112	As		●		HM133777			Culture contaminant, Russia
S. johnsonii	CBS 8241			Sf		●	●				Unknown
S. johnsonii	PYCC 4351			A2	A1	●		GU474644		GU474689	Leaf of a tree, Portugal
S. salmonicolor	**CBS 483**	EF592144	EF592118	A2	A1	●		GU474642		GU474673	Rusted leaf of *Citrus* sp., France
S. salmonicolor	CBS 487			A1	A2		●				Air, Japan
S. salmonicolor	**CBS 490^{T}**	AF070439	AY015434	A1	A2		●		GU474641	**GU474656**	Culture contaminant
S. salmonicolor	CBS 495			A1	A2		●				Culture contaminant, Japan
S. salmonicolor	CBS 496			A1	A2		●			GU474658	Scorched skin of orange, Delft, Netherlands
S. salmonicolor	CBS 497			A2	A1	●					Unknown
S. salmonicolor	CBS 1012			A1	A2		●		HM133782	GU474649	Leaf of *Aristolochia* sp., Botanic Laboratory, Delft, The Netherlands
S. salmonicolor	CBS 1013			A1	A2		●			GU474650	Air in dairy, USA
S. salmonicolor	CBS 1039			A1	A2		●			GU474651	Air in dairy, USA
S. salmonicolor	CBS 2635			A1	A2		●				Infected skin of a man, Germany
S. salmonicolor	CBS 2641			As			●				Air, Netherlands
S. salmonicolor	CBS 2647			A2	A1	●				GU474670	Gut of *Drosophila* sp., locality unknown
S. salmonicolor	CBS 2648			A2	A1	●					Exudate of *Kalopanax vicinifolium* var. *typicum*, Japan
S. salmonicolor	CBS 2873			A1	A2		●				Extract of oak bark, France
S. salmonicolor	CBS 4029			As			●				Soil, New Zealand
S. salmonicolor	CBS 4030			A1	A2		●				Soil, New Zealand
S. salmonicolor	CBS 4474			A2	A1	●		HM133770		GU474672	Blistered skin of a man, Bonn, Germany
S. salmonicolor	CBS 5937			A1	A2		●				Carious dentine, South Africa
S. salmonicolor	CBS 6322			A1	A2		●		HM133783	GU474660	Rotting wood shavings, Chile
S. salmonicolor	CBS 6470			A1	A2		●				Air, Germany
S. salmonicolor	CBS 6529			A2	A1	●					Culture contaminant, R.J. Bandoni
S. salmonicolor	CBS 6530			A1	A2		●				Bog, locality unknown
S. salmonicolor	CBS 6781			A1	A2		●				Bituminous soil, oil field, Cumanà, Venezuela
S. salmonicolor	CBS 6832			A1	A2		●		GU474646	**GU474664**	Cerebro-spinal fluid, India
S. salmonicolor	CBS 7260			A1	A2		●			GU474665	Slime flux of *Quercus pendunculata*, Russia

Species	Strains	D1/D2 LSU (GenBank)	ITS region (GenBank)	Mating behaviour		*STE3 A1*	*STE3 A2*	*STE3.A1* (GenBank)	*STE3.A2* (GenBank)	*HD1/HD2* (GenBank)	Isolation source and location
				Old	New						
S. salmonicolor	IHMT 2446/96			A2	A1	●				GU474676	Skin lesion, Portugal
S. salmonicolor	ML 2241			A2	A1	●		GU474643		**GU474678**	Isolated by J. W. Fell
S. salmonicolor	NRRL Y-17498			A2	A1	●					Ice core, Greenland
S. salmonicolor	PYCC 4558			A2	A1	●		HM133774		**GU474679**	Polluted river water, Oeiras, Portugal
S. salmonicolor	PYCC 4623			A2	A1	●				GU474680	Soil, USSR
S. salmonicolor	PYCC 5245			A1	A2		●				Unknown substrate, Portugal
S. salmonicolor	RJB 948			A2	A1	●					Isolated by R. J. Bandoni
S. salmonicolor	RJB 950			A1	A2		●				Isolated by R. J. Bandoni
S. salmonicolor	RJB 8219			A1	A2		●				Isolated by R. J. Bandoni
S. salmonicolor	ZP 392			A1	A2		●				Fruit body of *Dacrymyces* sp., Germany
S. salmonicolor	ZP 648			A2	A1	●				**GU474682**	Leaf of *Ligustrum* sp., Monte de Caparica, Portugal

III.2. Primers and specific PCR conditions used in Chapter 5 to amplify several genomic regions in red yeasts.

Two PCR protocols (A and B) described below were used to amplify the genomic regions under study in Chapter 5, which are given in Table III.2. For each genomic region is indicated which of two protocols was used. Primer annealing sites are shown in Figure III.1.

PCR Protocols

A. For PCR screening, reactions were performed in a final volume of 10 μl with the following components: 1X DreamTaq Buffer (Fermentas, Canada), 0.20 mM of each of the four dNTPs (GE Healthcare), 1% DMSO, 1.0 μM of each primer, 100 ng of genomic DNA, and 0.5 U DreamTaq DNA polymerase (Fermentas, Canada). Thermal cycling consisted of a 5-minute denaturation step at 95°C, followed by 35 cycles of denaturation at 95°C for 30 s, 30 s at the annealing temperature (variable), and extension at 72°C (variable time). The annealing temperatures, extension times and primer sequences are given for each case. A final extension of 7 min at 72°C was performed at the end of each reaction. PCRs reactions were adjusted to 50 μl final volume when sequencing was required.

B. For the long range PCR (LR-PCR), reactions were performed in a final volume of 50 μl and contained the following components: 1X Long PCR buffer, 2 mM of $MgCl_2$, 0.2 mM of each of the four dNTPs (GE healthcare), 1.0% DMSO, 1.0 μM of primer, 150 ng of genomic DNA and 2.5 U Long PCR enzyme mix (Fermentas, Canada). Thermal cycling consisted of a 3-minute denaturation step at 94ºC, followed by an initial round of 10 cycles of denaturation at 95ºC for 20 s, annealing for 30 s (variable temperature), extension at 68ºC (variable time), and a second round of 25 cycles increasing the extension time in each cycle (variable). A final extension of 10 minutes at 68ºC was performed. The annealing temperatures, extension times and primer sequences are given for each case.

Table III.2. L**ist of primers and specific PCR conditions used to amplify the indicated regions under study in Chapter 5.** Degenerate nucleotides follow the IUPAC nomenclature. All primers were obtained from MWG Biotech AG.

Species	Primer Name	Sequence (5' → 3')	Annealing/ Extension
Amplification of *STE3.A1* allele – Protocol A			
R. babjevae, *Rh. glutinis*, *R. diobovatum*, *Rh. araucariae*, *R. kratochvilovae*, "*Rh. hamamotoiana"*, *R. sphaerocarpum*, *R. azoricum*, *"R. oreadum"*, *Rh. colostri*,	MC159 MC164	ACGYTGCTSTTCATYTTYTGGTG GAAGATSGCRCASGAGATKGCGTA	60.0ºC / 60 s
"Rh. sesimbrana", *R. paludigenum*, *S. ruineniae*	MC190 MC191	ACCYTGCTYCTCATYTTYTGG ATGGCGCTYTTKGACGAYTC	59.0ºC / 60 s
R. toruloides	MC057 MC058	CATCTTCTGGTGTKTGGTCAMCGT GGGASGSTGTAGGCGATTTG	60.0ºC / 60 s
S. salmonicolor, *S. johnsonii*	MC053 MC054	CATCTTCTTCCTCCTCCTGCC GGGTAGGGAGGAAGCGAGAT	59.5ºC / 90 s
Amplification of the intervening region between *RibL18ae* and *RNAPOL* genes (encompasses the *STE3.A1*) – Protocol B			
S. metaroseus , *S. pararoseus*	MC135 MC138	GAGAAGGAGCCKCARCCCAA CCTGTCCKCCKCGCATCTT	60ºC / 7 min – 10 cy. 60ºC / 7 min + 5 s/ cy. – 25 cy.
	MC188 MC189	ACCGWCCMTCGACYTTCTACT ACTCACRCTCTCVACACGCTC	60ºC / 4 min – 10 cy. 60ºC / 4 min + 2 s/ cy. – 25 cy.
	MC160 MC192	ACTYTCMTCTACATYTTYTGGTG CACCCTCACYTATATTTTCTGG	Primer-walking sequencing (*S. metaroseus*) Primer-walking sequencing (*S. pararoseus*)
Sp. ruberrimus, *Sp. blumeae*	MC188 MC189	ACCGWCCMTCGACYTTCTACT ACTCACRCTCTCVACACGCTC	60ºC / 4 min – 10 cy. 60ºC / 4 min + 2 s/ cy. – 25 cy.
	MC160	ACTYTCMTCTACATYTTYTGGTG	Primer-walking sequencing
Sp. jilinensis	MC188 MC189	ACCGWCCMTCGACYTTCTACT ACTCACRCTCTCVACACGCTC	60ºC / 4 min – 10 cy. 60ºC / 4 min + 2 s/ cy. – 25 cy.
	MC193	CACCTTGATCTACGTTTTCTGG	Primer-walking sequencing
No amplification of the *STE3.A1* allele was obtained in the following species using any combination of the above mentioned primers: *Rh. graminis*, *"Rh. pinicortis"*, *R. dairenensis, Rh. mucilaginosa, Rh. pacifica, "Rh. pyritica", "R. feracious", R. fluviale*, *S. microsporus*, *Sp. odoratus*, *R. lusitaniae*, *S. longiusculus*			

Table III.2. Continued.

Species	Primer Name	Sequence (5' → 3')	Annealing/ Extension
Amplification of *STE3.A2* allele – Protocol A			
R. babjevae, *Rh. graminis*, *R. kratochvilovae*,	MC124 MC125	AACAKCGSGCCWCTGTACTGGCAG GCSACSACCCGAYSAAAGTTGCG	63.0ºC / 60 s
S. salmonicolor, *S. johnsonii*	MC126 MC127	AATRCCGCTCCGCTYTACTGGCA TGTTCGGGTCTACAGGTTGGAGAG	62.0ºC / 90 s
R. diobovatum, *Rh. araucariae*, *"Rh. hamamotoiana"*, *R. paludigenum*, *R. toruloides*, *Rh. dairenensis*, *R. azoricum*, *R. fluviale*, *S. microsporus*, *R. lusitaniae*	MC166 MC169	AKCGTKCCKCTSTACTGGCA CCGAAMGCGGCRAARAARAT	58.0ºC / 90 s
Rh. colostri	MC166 MC168	AKCGTKCCKCTSTACTGGCA CYGAAMGCSGCGAAGAARAT	59.0ºC / 60 s
Amplification of the intervening region between *LSm7* and *RNAPOL* genes (encompasses the *STE3.A2*) – Protocol B			
Sp. ruberrimus	MC123 MC189	AAAGGSTACGACCAGYTSCTCAAYCT ACTCACRCTCTCVACACGCTC	62ºC / 4 min – 10 cy. 62ºC / 4 min + 2 s/ cy. – 25 cy.
	MC195 MC198	TACAGGGTAARGGTATCTT CTGACAGGACGCATAGATACAT	54ºC / 2 min – 10 cy. 54ºC / 2 min + 1 s/ cy. – 25 cy.
Sp. pararoseus	MC123 MC189	AAAGGSTACGACCAGYTSCTCAAYCT ACTCACRCTCTCVACACGCTC	62ºC / 4 min – 10 cy. 62ºC / 4 min + 2 s/ cy. – 25 cy.
	MC196 MC199	GCGAAGTHGCWGCTCGAATC CCGATAGCCGTCCAACAGG	60ºC / 2 min – 10 cy. 60ºC / 2 min + 1 s/ cy. – 25 cy.
S. longiusculus	MC123 MC189	AAAGGSTACGACCAGYTSCTCAAYCT ACTCACRCTCTCVACACGCTC	62ºC / 4 min – 10 cy. 62ºC / 4 min + 2 s/ cy. – 25 cy.
	MC194 MC197	CCGTGAAGGACGCCTCGAT CTCTCGCTCAATGACTACCTT	59ºC / 2 min – 10 cy. 59ºC / 2 min + 1 s/ cy. – 25 cy.
	MC167	AATRCCGCKCCKCTYTACTGGCA	Primer-walking sequencing
No amplification of the *STE3.A2* allele was obtained in the following species using any combination of the abovementioned primers			
Rh. glutinis, *"Rh. sesimbrana"*, *"Rh. pinicortis"*, *Rh. mucilaginosa*, *Rh. pacifica*, *R. sphaerocarpum*, *"Rh. pyritica"*, *"R. feracious"*, *Sp. odoratus*, *"R. oreadum"*, *S. longiusculus*, *Sp. blumeae*, *Sp. jilinensis*			

Table III.2. Continued.

Species	Primer Name	Sequence (5' → 3')	Annealing/ Extension	Amplified region	GenBank Accession Number
Confirm synteny in the vicinity of the *STE3.A1* allele in selected red yeast species (see Figures 5.3 and 5.4) **– Protocols A** and **B** Amplicons were obtained for the following red species but only fully sequenced in those for which a GenBank Accession number is given					
R. babjevae, R. diobovatum	MC109 MC122	CGTCCGCCGCCCGAACATCAAG TGGCGMGTYCCCGAACCYTACAAC	65ºC / 4 min – 10 cy. 65ºC / 4 min + 2 s/ cy. – 25 cy.	*Rib L18ae – RibL6* (includes *STE3.A1*)	
R. kratochvilovae					JN246653
Rh. araucariae					JN246654
R. azoricum					JN246655
R. toruloides	MC109 MC161	CGTCCGCCGCCCGAACATCAAG AAGAYSGCCGGCAKGCTCCARTA	65ºC / 4 min – 10 cy. 65ºC / 4 min + 2 s/ cy. – 25 cy.	*Rib L18ae – RtSTE3.A1*	
	MC057 MC022	CATCTTCTGGTGTKTGGTCAMCGT GGCGGATGGATTGGTTGAGC	60ºC / 6 min – 10 cy. 60ºC / 6 min + 2 s/ cy. – 25 cy.	*RtSTE3.A1 – RHA1*	(Coelho et al. 2008)
R. sphaerocarpum	MC109 MC122	CGTCCGCCGCCCGAACATCAAG TGGCGMGTYCCCGAACCYTACAAC	65ºC / 4 min – 10 cy. 65ºC / 4 min + 2 s/ cy. – 25 cy.	*Rib L18ae – RibL6* (includes *STE3.A1*)	JN246664
	MC159 MC164	ACGYTGCTSTTCATYTTYTGGTG GAAGATSGCRCASGAGATKGCGTA	60.0ºC / 60 s		
Sp. ruberrimus, *Sp. blumeae*	MC188 MC189	ACCGWCCMTCGACYTTCTACT ACTCACRCTCTCVACACGCTC	60ºC / 4 min – 10 cy. 60ºC / 4 min + 2 s/ cy. – 25 cy.	*Rib L18ae – RNAPOL* (includes *STE3.A1* and *RibL6* genes)	
	MC160	ACTYTCMTCTACATYTTYTGGTG	Primer-walking sequencing		
S. pararoseus	MC135 MC138	GAGAAGGAGCCKCARCCCAA CCTGTCCKCCKCGCATCTT	60ºC / 4 min – 10 cy. 60ºC / 4 min + 2 s/ cy. – 25 cy.	*Rib L18ae – RNAPOL* (includes *STE3.A1* and *RibL6* genes)	JN246670
	MC188 MC189	ACCGWCCMTCGACYTTCTACT ACTCACRCTCTCVACACGCTC	Primer-walking sequencing		
	MC192	CACCCTCACYTATATTTTCTGG			
S. metaroseus	MC135 MC138	GAGAAGGAGCCKCARCCCAA CCTGTCCKCCKCGCATCTT	60ºC / 4 min – 10 cy. 60ºC / 4 min + 2 s/ cy. – 25 cy.	*Rib L18ae – RNAPOL* (includes *STE3.A1* and *RibL6* genes)	
	MC188 MC189	ACCGWCCMTCGACYTTCTACT ACTCACRCTCTCVACACGCTC	Primer-walking sequencing		
	MC160	ACTYTCMTCTACATYTTYTGGTG			
Sp. jilinensis	MC188 MC189	ACCGWCCMTCGACYTTCTACT ACTCACRCTCTCVACACGCTC	60ºC / 4 min – 10 cy. 60ºC / 4 min + 2 s/ cy. – 25 cy.	*Rib L18ae – RNAPOL* (includes *STE3.A1* and *RibL6* genes)	
	MC193	CACCTTGATCTACGTTTTCTGG	Primer-walking sequencing		
Sp. ruineniae	MC109 MC122	CGTCCGCCGCCCGAACATCAAG TGGCGMGTYCCCGAACCYTACAAC	65ºC / 4 min – 10 cy. 65ºC / 4 min + 2 s/ cy. – 25 cy..	*Rib L18ae – RNAPOL* (includes SruRHA1.A2)	

Table III.2. Continued.

Species	Primer Name	Sequence (5' → 3')	Annealing/ Extension	Amplified region	GenBank Accession Number
Confirm synteny in the vicinity of the *STE3.A2* allele in selected red yeast species (see Figures 5.3 and 5.4) **– Protocols A** and **B** Amplicons were obtained for the following red species but only fully sequenced in those for which a GenBank Accession number is given					
R. babjevae, Rh. graminis, *R. lusitaniae*	MC122 MC123	TGGCGMGTYCCCGAACCYTACAAC AAAGGSTACGACCAGYTSCTCAAYCT	(LR-PCR) 63ºC / 4 min – 10 cy. 63ºC / 4 min + 2 s/ cy. – 25 cy.	*LSm7 – RibL6* (includes *STE3.A2*)	
R. diobovatum					JN246579
R. kratochvilovae					JN246585
Rh. araucariae					JN246588
Rh. colostri					JN246603
S. johnsonii					JN246613
R. paludigenum, R. toruloides	MC123 MC124	AAAGGSTACGACCAGYTSCTCAAYCT AACAKCGSGCCWCTGTACTGGCAG	62.0ºC / 2 min	*STE3.A2 – LSm7*	
Sp. ruberrimus	MC123 MC189	AAAGGSTACGACCAGYTSCTCAAYCT ACTCACRCTCTCVACACGCTC	62ºC / 4 min – 10 cy. 62ºC / 4 min + 2 s/ cy. – 25 cy.	*LSm7 – RNAPOL* (includes *STE3.A2 and RibL6* genes)	
	MC195 MC198	TACAGGGTAARGGTATCTT CTGACAGGACGCATAGATACAT	54ºC / 2 min – 10 cy. 54ºC / 2 min + 1 s/ cy. – 25 cy.		
Sp. pararoseus	MC123 MC189	AAAGGSTACGACCAGYTSCTCAAYCT ACTCACRCTCTCVACACGCTC	62ºC / 4 min – 10 cy. 62ºC / 4 min + 2 s/ cy. – 25 cy.	*LSm7 – RNAPOL* (includes *STE3.A2 and RibL6* genes)	JN246610
	MC196 MC199	GCGAAGTHGCWGCTCGAATC CCGATAGCCGTCCAACAGG	60ºC / 2 min – 10 cy. 60ºC / 2 min + 1 s/ cy. – 25 cy.		JN246611
S. longiusculus	MC123 MC189	AAAGGSTACGACCAGYTSCTCAAYCT ACTCACRCTCTCVACACGCTC	62ºC / 4 min – 10 cy. 62ºC / 4 min + 2 s/ cy. – 25 cy.	*LSm7 – RNAPOL* (includes *STE3.A2 and RibL6* genes)	JN246612
	MC194 MC197	CCGTGAAGGACGCCTCGAT CTCTCGCTCAATGACTACCTT	59ºC / 2 min – 10 cy. 59ºC / 2 min + 1 s/ cy. – 25 cy.		
	MC167	AATRCCGCKCCKCTYTACTGGCA	Primer-walking sequencing		
S. salmonicolor	MC123 MC189	AAAGGSTACGACCAGYTSCTCAAYCT ACTCACRCTCTCVACACGCTC	62ºC / 4 min – 10 cy. 62ºC / 4 min + 2 s/ cy. – 25 cy.	*LSm7 – RNAPOL* (includes *STE3.A2 and RibL6* genes)	
	MC126 MC122	AATRCCGCTCCGCTYTACTGGCA TGGCGMGTYCCCGAACCYTACAAC	62ºC / 2 min – 10 cy. 62ºC / 2 min + 1 s/ cy. – 25 cy.		

Table III.2. Continued.

Species	Primer Name	Sequence (5' → 3')	Annealing/ Extension
Amplification of *HD1/HD2* alleles – Protocol B			
R. babjevae, *Rh. glutinis*, *Rh. graminis*, *R. diobovatum*,	MC118 MC120	TGTTGRYGAACCAGGTRTCGAKCTG AGGATGKCGAGGACSTCGKGSGTGAA	65ºC / 1 min 30 s – 10 cy. 65ºC / 1 min 30 s + 1 s/ cy. – 25 cy.
S. salmonicolor, *S. johnsonii*	(Coelho et al. 2010)		

Species	Primer Name	Sequence (5' → 3')	Annealing/ Extension	Amplified region	GenBank Accession Number
Amplification intervening region between *STE20* and *KAP95* genes (encompasses the *RHA2*) – Protocol A					
S. salmonicolor (ML 2241, NRRL Y-17498) *S. johnsonii* (CBS 1522)	MC040 MC073	GCCCGAAGTCGTCAAGCAGAAGGA TTCCCSAAYGGCCARCTCAAGGAGCC	63.0ºC / 3 min	*STE20 – KAP95* (includes *RHA2* gene)	JN246569 JN246568 JN246571
	MC113 MC115	AGCCRGCGCTTSGGATCAGTTAC AAGGAAGGATGCTYACTTGCGG	Primer-walking sequencing		
S. salmonicolor (PYCC 4558)	MC040 MC073	GCCCGAAGTCGTCAAGCAGAAGGA TTCCCSAAYGGCCARCTCAAGGAGCC	63.0ºC / 3 min	*STE20 – KAP95* (includes *RHA2* gene)	JN246570
	MC113 MC115 MC130	AGCCRGCGCTTSGGATCAGTTAC AAGGAAGGATGCTYACTTGCGG TCCGAACAATGGATCAACAGGC	Primer-walking sequencing		
S. jonhsonii (strains: PYCC 4351)	MC040 MC073	GCCCGAAGTCGTCAAGCAGAAGGA TTCCCSAAYGGCCARCTCAAGGAGCC	63.0ºC / 3 min	*STE20 – KAP95* (includes *RHA2* gene)	JN246572
	MC114 MC115	GGAATCGGGACCCCAAGAAAC AAGGAAGGATGCTYACTTGCGG	Primer-walking sequencing		

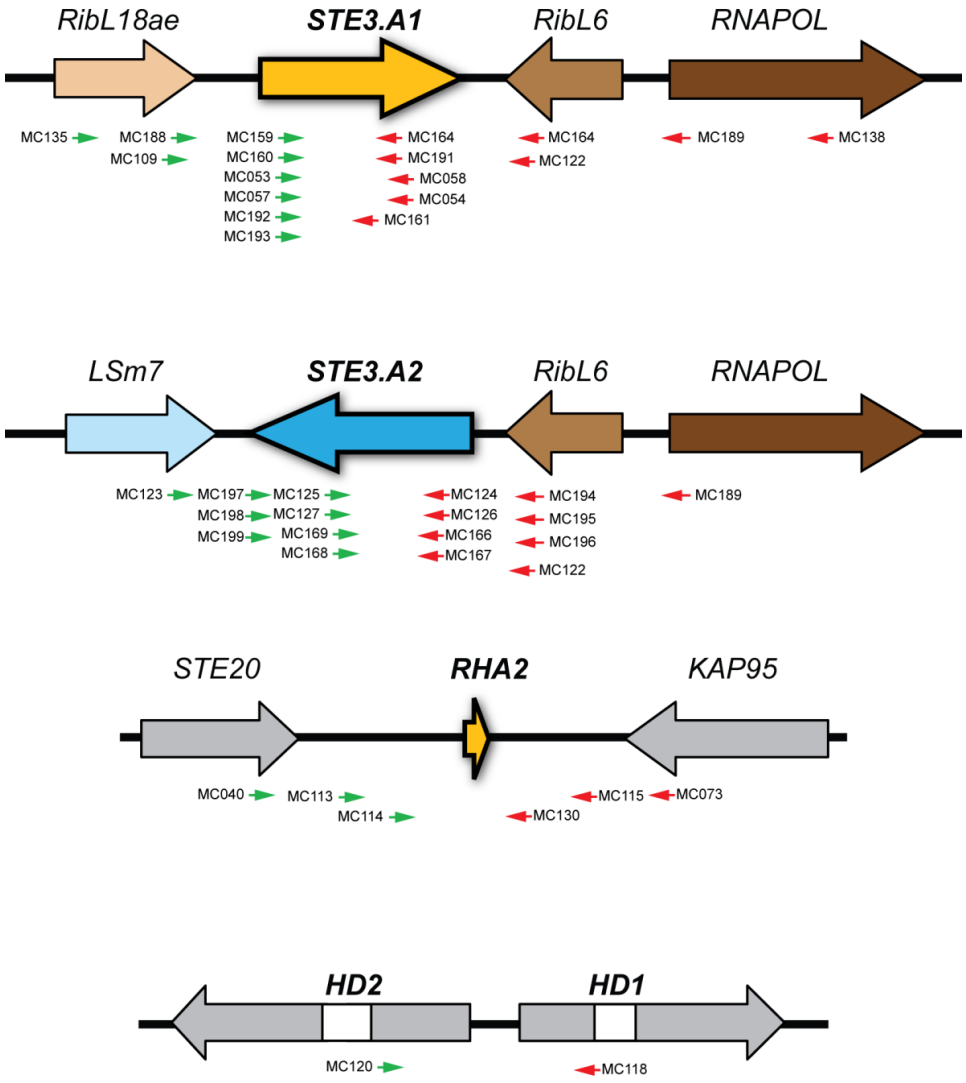

Figure III.1. Annealing sites of the primers used to amplify the genomic regions under study in Chapter 5.