UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de Biologie Moléculaire et Cellulaire - UPR 9002 CNRS - Strasbourg

Architecture et Réactivité de l'ARN

THÈSE présentée par :

Yacine DAHMAN

soutenue le : **19 Septembre 2018**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

Caractérisation de la RNase P nucléaire de *Candida glabrata* et amélioration des outils d'édition de son génome.

THÈSE dirigée par :

Mr. JOSSINET Fabrice	Maître de conférence, Université de Strasbourg

RAPPORTEURS :

Mme. FAIRHEAD Cécile	Professeur des universités, Université Paris Sud
Mme. BUSCAINO Alessia	Maître de conférence, Université du Kent, Canterbury

---

AUTRES MEMBRES DU JURY :

Mr. WESTHOF Eric	Professeur des universités, Université de Strasbourg

À khalti Rabia, Sidi & Ama

Acknowledgment / Remerciements

I am sincerely thankful to the thesis committee members, Dr. Alessia BUSCAINO, Pr. Cecile FAIRHEAD and Pr. Eric WESTHOF who accepted to read and evaluate this work.

I wish to thank Dr. Pascale ROMBY director of the UPR 9002 for giving me the opportunity to work in her department , and for her kindness throughout my thesis.

I would like to thank Fabrice JOSSINET for welcoming me in his team and for supporting me during these three years of thesis.

Je souhaiterais remercier les membres actuels de l'équipe, Delphine RICHER et Laurence DESPONS. J'aimerais aussi remercier particulièrement Antinea BABARIT, que j'ai eu la chance d'encadrer pendant ma thèse.

Merci aussi à tous les membres de l'UPR 9002 avec qui j'ai partagé de bons moments ces trois dernières années. Merci à Tracy et Leandra pour leur efficacité hors pair.

J'aimerais remercier tout les membres de l'IBMC et spécialement les drosophilistes de l'UPR 9022.

Un grand merci à Dr. Pascaline ULLMAN, pour m'avoir fait confiance et m'avoir donner la chance de pouvoir enseigner durant ces trois années de thèse. Merci tout particulièrement à Dr. Salah BOUZOUBAA qui m'a beaucoup aidé et épaulé dans cette expérience d'enseignement.

J'aimerais aussi remercier mes anciens professeurs et notamment Eric WESTHOF, Philippe CARBON et Evelyne MISLINSKI-CARBON pour m'avoir tant appris et pour m'avoir transmis votre passion pour la biologie moléculaire.

Je souhaiterais remercier Dr. Dominique GAGLIARDI, qui a toujours été bienveillant à mon égard et qui a résolument déclenché en moi l'envie de faire une thèse.

Je ne remercierai jamais assez mes chers parents, Houria et Abdelkader DAHMAN, mon frère Azzedine et ma soeur Sabrina pour leur soutien indéfectible. Merci à jedda Ama Zohra qui m'a toujours encouragé même à 1500 km de distance.

Merci évidement à Catherine pour son soutien, sa patience et son aide précieuse durant la préparation de cette thèse.

Table of contents

General introduction

1. The *Fungi* kingdom

The *Fungi* kingdom is the largest and oldest group of living organisms. While approximately 100,000 fungal species have been identified so far, recent studies estimate the existence of 5.1 million different species (Blackwell 2011). Fungi are mostly multicellular eukaryotes with a characteristic wall composed of chitin or chitosan. Their reproductive mode can be sexual or asexual. They are heterotrophic for carbon, absotrophic and their mode of nutrition is mostly saprophyte. Fungi interact with species from all known groups of organisms, through mutualist or parasitic relationships. The classification of fungi was originally based on their observable morphology.

The first publication describing the fungi systematics was published in 1729 by the Italian botanist Pier Antonio Micheli in the *Nova Planetarium Genera*. Technological advances allowing the study of new features of fungi enabled to progressively refine their classification. Mycologists were then able to classify fungi according to their reproductive mode, growth temperature, nutrient use, secondary metabolite production and cell composition. Analysis of these new data led to the publication of the first phylogeny of fungi in 1962 (Alexopoulos 1962). According to this study, fungi formed the *Mycota* division within the reign of the *Plantae*. *Mycota* were separated into two subdivisions : *Myxomycotina* and *Eumycotina* (introducing the term "True Fungi"), containing 10 different classes. In the late 1990s, sequencing techniques and their application to phylogeny led to significant changes in the understanding of fungi. DNA sequence analysis of the small ribosomal subunit gene initially showed that fungi were phylogenetically closer to animals than plants (Wainright et al. 1993), thereby transforming fungi into a kingdom *per se*. The sequence analysis was then extended to a larger number of genes.

First, housekeeping genes found in single copies such as actin (*ACT*) (Helgason et al. 2003) or elongation factor 1 translation (*TEF1*) (Rehner and Buckley 2005) were compared. The subsequent analysis suggested that the kingdom of *Fungi* was a monophyletic group. The emergence of next-generation sequencing (NGS) promoted the publication of good quality fungal genomes, reaching today more than 1000 sequenced genomes (Aylward et al. 2017). The accessibility of these sequencing data allowed the emergence of phylogenomics, a new field based on the analysis and comparison of these entire genomes (Delsuc et al. 2005).

The fungal consensus system in 2018 is based on this methodology and presents the *Fungi* kingdom as a non-monophyletic group subdivided into 9 distinct phyla: *Ascomycota*, *Basidiomycota*, *Mucormycota*, *Zoopagomycota*, *Blastocladiomycota*, *Chytridiomycota*, *Cryptomycota*, *Neocallimastigomycota*, and *Microsporidia*. *Basidomycota* and *Ascomycota* form together a « sister taxa » called *Dikaryomycota* which comprises more than 98% of the identified species (Shaffer 1975).

The species that will be covered in this thesis belong to the *Saccharomycotina* subphylum of *Ascomycota*. *Ascomycota* is the largest clade of the *Fungi* kingdom with more than 65 000 estimated species (Kirk et al. 2008). *Ascomycot*a are divided into 3 distinct groups: *Pezizomycotina* (filamentous species like *Neurospora crassa*), *Taphrinomycotina* (*Schizosaccharomyces pombe)* and the subphylum of *Saccharomycotina* comprising *Saccharomyces cerevisiae*, the yeast model (Schoch et al. 2009). *Saccharomycotina* includes the *Saccharomycetales* group associated with the species *Yarrowia lipolytica*. *Saccharomycetales* are subdivided into two clades: the CTG clade, encompassing the species translating the CTG codons into serine rather than leucine; and the *Saccharomycetacea* clade (Figure 1). Finally the *Saccharomycetaceae* are subdivided into two clusters: the species whose common ancestor underwent a complete duplication of the genome (post-WGD species) and the species which diverged before this event (Gabaldón et al. 2013).

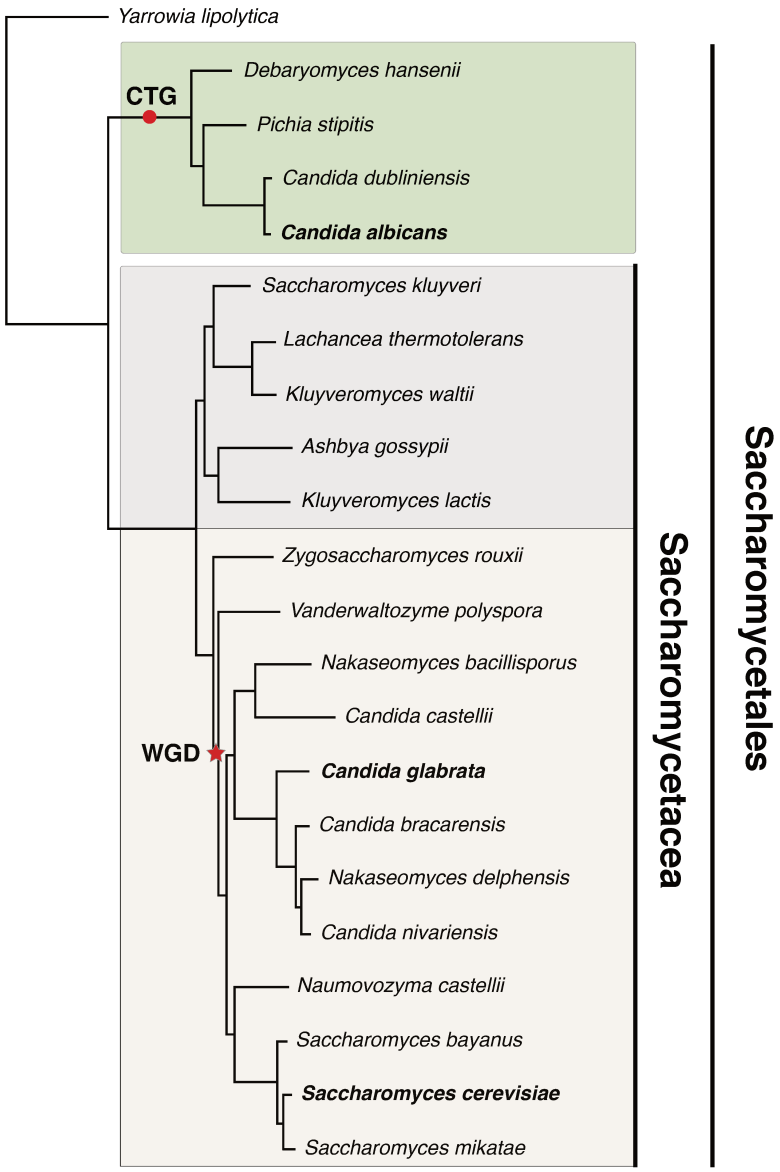

Figure 1. Phylogeny of *Saccharomycotina* species. (adapted from Gabaldon et al. 2013). The tree was built by aligning 603 orthologous gene sequences. Major evolutionary events, such as the complete duplication of the genome (WGD) and the transition of the genetic code to the CTG clade are respectively indicated by a star and a full circle.

In many languages the term "yeast" is derived from the use of these species in the fermentation of sugar into carbon dioxide for baking and into alcohol for brewing (Kurtzman et al. 2011). Across literature, the term "yeast" is often used as a synonymous for *Saccharomyces cerevisiae* and yeasts have long been referred to mostly unicellular species in the *Saccharomycotina* subphylum. This definition has changed since the discovery of yeast belonging to the *Basidiomycota* phylum. The consensus definition of yeast is then: "yeasts, whether ascomycetes or basidiomycetes, are generally characterized by budding or fission as the primary means of asexual reproduction, and have sexual states that are not enclosed in fruiting bodies" (Kurtzman et al. 2011). Yeasts are found in all ecosystems, and some species have been beneficial to humans by direct interaction as part of the normal microbiota (Sam et al. 2017). Humans have also isolated and domesticated numerous strains. For example, *Saccharomyces cerevisiae* and yeasts like *Hanseniaspora* spp., *Pychia* spp., *Debaryomyces* spp., and *Candida* spp. are still used extensively for the production of fermented food and beverages (Hittinger et al. 2018). The domestication of yeasts has made possible to set up the necessary conditions for their culture in the laboratory. Yeasts such as *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* have thus become eukaryotic models of choice in molecular and cellular biology (Botstein et al. 1997, Yanagida 2002).

2. Human fungal infection.

The number of deaths caused by fungal infections is as high as for malaria (Brown, Denning et al. 2012). Recent studies estimate that approximately 1.7 billion people worldwide are affected by fungal infections (Brown, Denning et al. 2012).

2.1. Fungal infection of healthy humans.

Healthy people are mostly affected by superficial fungal infections. These infections represent the fourth most common cause of human disease (Hay 2017). They preferentially affect skin and keratinized structures and are generally caused by two large groups of fungi: Dermatophytes, with species of the *Trichophyton*, *Microsporum* and *Epidermophyton* genus, and *Malassezia* species (Weitzman and Summerbell 1995, Gupta et al. 2004). Superficial infections also occur in the oral and vaginal mucosa and are mainly caused by the yeast *Candida albicans*. These infections are usually non life-threatening and well restrained. Nevertheless, healthy humans are also plagued by invasive infections, i.e. infections that contaminate deep tissues, much more threatening for the patients health. These species causing invasive fungal infections are contracted by ingestion, inhalation or contact with cutaneous lesion. They are represented by three distinct groups: i) *Entomophthoromycota*, fungi first described as infectious agents for insects, often restricted to tropical regions. These fungi develop in the facial or intestinal mucosa depending on the mode of contamination. They induce respiratory, digestive and disfigurement issues (Isa-Isa et al. 2012); ii) *Onygenales*, fungi belonging to the *Ascomycota* phylum. They are found worldwide but each pathogenic species in this group is geographically restricted. Depending on the species involved, these fungi develop in the mucous membranes of the respiratory system, inducing severe pneumonia, but also in the bones, subcutaneous tissues and in the digestive tract (Bradsher et al. 2003, Marques 2012); iii) *Cryptococci*, fungi belonging to the *Basidiomycota* phylum, represented by *Cryptococcus neoformans* and *Cryptococcus gattii*. They contaminate humans by

inhalation or contact with a cutaneous lesion and can be spread from person to person. *Cryptococcus* species usually develops in the lungs inducing pneumonia but can also reach the central nervous system and cause meningitis (Byrnes Iii et al. 2010). Although the consequences of invasive fungal infections are often dramatic, their occurrence remains rare in healthy individuals. The human immune system provides an effective response to invasive infections, which are generally controlled before developing into systemic infections.

2.2.Immune response to fungal infection.

The surface epithelium, including the skin and mucous membranes of the oral cavity, gastrointestinal, pulmonary and genito-urinary tracts, are the first barrier faced by pathogenic fungi (Moens and Veldhoen 2012). These epitheliums consist of a combination of: i) a mechanical barrier due to mucus secretion and tight-junctions between cells (France and Turner 2017); ii) a chemical barrier resulting from the secretion of antimicrobial peptide (AMP) by epithelial cells (Abiko et al. 2002); and iii) a microbial barrier set up by microbiota species, inducing strong competition for access to nutrients and adhesion sites (Dollive et al. 2013). If a pathogenic fungus crosses the epithelial barrier, innate immune cells (dendritic cells, natural killer cells, neutrophils, monocyte and macrophages) activate inflammatory responses and phagocytosis (Drummond et al. 2014). These cells recognize specific molecules presented at the surface of the pathogen fungi cell-wall, called the pathogen associated molecular patterns (PAMPs) (Sorrell and Chen 2009). PAMPs are detected by the pattern recognition receptors (PRR), involving the Toll-like receptors (TLRs) and C-type lectin-like receptors (CLRs) in the case of fungi (Netea et al. 2007, Netea et al. 2008). Adaptive immunity also plays a role in the response to fungal infections. Dendritic cells form the major link between innate immunity and adaptive immunity cells (Ramirez-Ortiz and Means 2012). Dendritic cells present fungal antigens to naive T cells which will then differentiate into memory and cytotoxic cells to specifically respond to the pathogen.

Understanding these defense mechanisms through the characterization of immune cells, the PAMPs and their recognition by PRR, paves the way to use a vaccination strategy to prevent fungal infection. Phase 2 clinical trials are already carried out for *Candida albicans* (Sui et al. 2017). However, the vaccination strategy faces two major problems: i) Vaccine development requires many efforts and substantial financial resources, which are unfortunately poorly invested in mycology (Spellberg 2011, Brown et al. 2012); ii) Vaccination is ineffective for immunocompromised patients, the most risky population for life-threating fungal infections and who show a dramatically high mortality rate after infection: up to 95% for *Aspergillus fumigatus*, 70% for *Cryptococcus* and 75% for *Candida albicans* (Brown et al. 2012).

2.3.Fungal infection of immunocompromised humans.

Since three decades, the increase in the frequency of invasive fungal infections is in correlation with the number of immunocompromised individuals. Paradoxically, this increase is linked to major medical advances, namely: i) Better care for people infected with HIV, allowing a longer lifespan and thus increasing the probability of invasive fungal infection (Samji et al. 2013); ii) Treatment of cancer patients by radiotherapy and chemotherapy, which considerably reduces their immunity (Galluzzi et al. 2015); iii) Organ transplants which effectiveness is increased by anti-rejection treatments (Fishman 2017); iv) Transplantation of stem cells for the regeneration of the immune system which is followed by compromised immune defenses (Kontoyiannis et al. 2010); v) Treatment of chronic diseases, such as asthma, and autoimmune diseases like inflammatory bowel disease by immunosuppressors (Kennedy et al. 2000); and vi) General medical advances leading to a significant increase in life expectancy and a subsequent increase in the number of elderly persons (Kauffman and Yoshikawa 2001). In parallel to these situations leading to immunodeficiency, the extensive use of catheters and invasive surgical procedures correlates significantly with the increased incidence of invasive fungal infections (Zilberberg and Shorr 2009). Individuals with weakened immunity are then susceptible to infections by species usually controlled in healthy individuals and are frequently associated to the human microbiota. These species are called opportunistic pathogens and the vast majority of invasive opportunistic fungal infections are linked to: *Cryptococcus* sp., *Aspergillus* sp. and *Candida* sp. (Brown et al. 2012).

2.4. The *Candida* genus.

The name « *Candida* » does not refer to a monophyletic group but rather to the definition of an asexual yeast, related to ascomycetes or basidiomycetes, that divide by multilateral budding (Kurtzman et al. 2011). Infections due to these species are called "candidiasis" and the term "candidemia" refers to systemic candidiasis. About 200 species form this group, of which 32 species are human pathogens (Pfaller et al. 2010). *Candida* species are most frequently isolated fungal species in the case of nosocomial infections (Pelroth 2007). Their incidence seems more related to the use of invasive medical procedures (venous or urinary catheter and invasive surgery) than to the immunodeficiency level of patients (McKinnon et al. 2001, Charles et al. 2003). In France, 35 876 cases of invasive fungal infections were reported between 2001 and 2010, among which 43.4% were due to *Candida* species (Bitar et al. 2014). *Candida albicans* is the most frequently isolated species worldwide for candidemia (more than 50%) (Pfaller et al. 2010). Nevertheless the frequency of candidemia due to « non-*Candida albicans*" species (NAC species) shifted from 10-40% from 1970 to 1990, to 35-65% in the 2000s (Krcmery and Barnes 2002). Today the most globally isolated NAC species causing candidemia are : *Candida glabrata* (13%), *Candida tropicalis* (7%), *Candida parapsilosis* (6%) and *Candida krusei* (2%) (Pfaller et al. 2010). The extensive use of azole compounds for the treatment of candidemia appears to have contributed to the emergence of naturally resistant species such as *Candida tropicalis* and *Candida glabrata*. Nevertheless, this shift is multifactorial and significant variations in the distribution of NAC species have been identified (Pfaller et al. 2010). For example *Candida tropicalis* is more frequently isolated in the southern hemisphere, and *Candida glabrata* and *Candida parapsilosis* in North America and Europe. Moreover, *Candida glabrata* is more frequently isolated in elderly people while *Candida parapsilosi*s is more common in young children (Yapar 2014).

3. *Candida glabrata*, an opportunistic pathogen

3.1.Taxonomy and morphological features

The yeast *C. glabrata* was first described in 1917 as part of the normal human intestinal flora (Anderson 1917). Spread on Sabouraud Dextrose Agar medium (SDA), the colonies appear round, smooth and creamy in colour. Based on this morphology, the name *Cryptococcus glabratus* has been assigned. This yeast was described as not forming hyphae and was therefore classified as *Torulopsis*, to become *Torulopsis glabrata* (Lodder and De Vries 1938). The present name was given when the definition of the « *Candida* » genus has been changed to « strains of species or varieties with pseudohyphae absent, rudimentary or well developed » (Yarrow and Meyer 1978). The absence of sexual reproduction and its inability to change mate reinforces *Candida* genus affiliation. *C. glabrata* is an haploid yeast of 1-4 µM, not capable of metabolizing galactose but only trehalose and glucose, and is naturally auxotrophic for nicotinic acid, pyridoxine and thiamine (Rodrigues et al. 2014). Analysis of ribosomal DNA sequences placed *C. glabrata* in the phylum of ascomycetes, and were the first evidence that *C. glabrata* is phylogenitically closer to *S. cerevisiae* than other species of the *Candida* genus (Kurtzman and Robnett 1998). *C. glabrata* is now included in the *Nakaseomyces* clade (Kurtzman 2003). This clade comprises 6 species, including 3 environmental species: *Candida castelli*, *Nakaseomyces delphensis* and *Nakaseomyces bacillisporus*; and 3 species pathogenic to humans: *Candida bracariensis*, *Candida nivariensis* and *Candida glabrata* (Alcoba-Florez et al. 2005, Correia et al. 2006). The reservoir of *C. glabrata* is almost exclusively human. Its circulation occurs directly from individual to individual, or indirectly after contamination of an abiotic surface. However, recent studies suggest that domestic or closely related birds possess strains of *C. glabrat*a of human origin in their microbiota gut and participate in their circulation (Al-Yasiri et al. 2016).

Table 1. Characteristics of Candida glabrata (adapted from Rodrigues et al. 2014)

Feature	***Candida glabrata* characteristics**
Ploidy	Haploid
Hyphae/Pseudohyphae	Absent
Colonies on SDA	Small and cream-color
Cell size	1-4 μM
Biochemical reactions	Ferments and assimilates glucose and trehalose
Pathogenicity	Opportunistic pathogen
Mating genes	Present
Sexual cycle	Unknown
Auxotrophy	Niacin, thiamine, pyridoxine

SDA = Sabouraud Dextrose Agar

3.2. Clinical relevance

C. glabrata was considered to be a saprophyte fungi commensal for humans until the late 1980s, when it was first identified as directly involved in candidemia of immunocompromised patients (Just et al. 1989). *C. glabrata* was then rapidly considered as an emerging pathogen (Hazen 1995). Today *C. glabrata* is the second cause of candidemia in Europe and North America (Guinea 2014). The mortality rate linked to candidemia by *C. glabrata* is one of the highest among *Candida* species and appears in 40% of the cases (Tortorano et al. 2006). The major risk factors to contract an infection by *C. glabrata* are associated with hospitalization in intensive care units, which are exacerbated by : i) Previous treatment with fluconazole (Colombo et al. 2013); ii) Installation of mechanical ventilation and central venous catheter, allowing *C. glabrata,* already present in patients or from exogenous origin (tools or hands of contaminated practitioners), to access to deep tissues (Gupta et al. 2015); ii) Advanced age of the hospitalized patient (Sampaio and Pais 2014). The combination of these factors accounts for *C. glabrata* being the most frequently isolated species in hospitalized elderly people.

4. Pathogenicity of *Candida glabrata*

Pathogenicity can be difficult to define and its meaning varies between authors. In this thesis, the choice was made to define pathogenicity as the capacity of an organism to cause disease in a host. According to this definition, pathogenicity is based on two factors: i) infectivity, which can be defined as the ability to infect and colonize host tissues; and ii) virulence, which is the capacity to cause damages to the host.

A commensal yeast has a high probability of leading to invasive infection if it exhibits the following characteristics: i) a high dependence to the human body for growth; ii) an ability to survive in the host despite stress caused by lack of nutrients and the immune system ; iii) an ability to reach the blood system and to adhere to deep tissues.

Concerning the opportunistic pathogen *C. glabrata*, pathogenicity is largely due to its infectivity and results more from an accident than from an active trait as for specialist pathogens. The pathogenic potential of *C. glabrata* is only expressed when the cell reaches the wrong tissue in the wrong host (Gabaldon and Carrete 2016). The pathogenic trait of *C. glabrata* has been traditionally interpreted by comparison with *S. cerevisiae*, whose genome appeared very similar (Dujon et al. 2004). Differences between *S. cerevisiae* and *C. glabrata* were interpreted as an adaptation to humans (Roetzer et al. 2010) but the sequencing of new non-pathogenic species close to *C. glabrata* (Gabaldón et al. 2013) have changed this paradigm.

4.1. Dependence on the human body for the growth of *C. glabrata*

Unlike *S. cerevisiae*, *C. glabrata* is intrinsically auxotrophic for pyridoxine, thiamine and nicotinic acid. These auxotrophies, and notably the inability to synthesize nicotinic acid, an essential precursor of NAD+ for cell viability, have been described as formal evidence for high host dependence (Domerge 2015). These hypotheses were refuted after the sequencing and phylogenetic analysis of the 5 other species of the *Nakaseomyces* clade (Gabaldón et al. 2013). Indeed, auxotrophy for nicotinic acid was found to be present in all species of the clade, especially in strictly environmental species (*C. castelli*, *N. delphensis* and *N. bacillisporus*) non-comensal for humans.

4.2. Stress resistance : starvation and host immune system

When growing on mucosal surfaces, *C. glabrata* has to undergo the intrinsic rarefaction of nutrients (Hood and Skaar 2012) that can also result from the competition with other micro-organisms (Wargo and Hogan 2006). Unlike *C. albicans*, *C. glabrata* does not destroy tissues to sustain its nutritional needs : it does not form hyphae and does not secrete proteases. However, *C. glabrata* is able to use alternative and non-fermentable carbon sources such as lactate (Ueno et al. 2011) and pexophagy, mechanism allowing the recycling of peroxysomes (Roetzer et al. 2010). Mucosal surfaces and other human tissues are generally poor in free iron, which is sequestered intracellularly in associations with proteins. *C. glabrata* promotes mitophagy to survive in iron-poor environment (Nagi et al. 2016). Concerning host iron acquisition, the strategy is not yet well described. *C. albicans* destroys erythrocytes and obtains iron by secreting haemolysins (Almeida et al. 2009). Recent studies have shown that *C. glabrata* expressed a hemolysin, but the conditions of its expression are elusive (Luo et al. 2004). Nevertheless *C. glabrata* appears to differ drastically from all pathogenic fungal organisms in its exclusive use of non-protein bound iron, producing an extracellularly non-protein ferric reductant (Gerwien et al. 2017). Note that these analyses have not yet been performed on other species of the *Nakaseomyces* clade.

These features also confer resistance to phagocytosis. *C. glabrata* was early described as capable of surviving these anti-microbial mechanisms and is able to replicate within phagocytic cells (Otto and Howard 1976). Once engulfed by a phagocytic cell, *C. glabrata* is restricted in a compartment called the phagosome, where nutrients are extremely limited (Flannagan et al. 2009). Within phagocytic cells this compartment is matured by fusing to the endosomal vesicles thereby forming a phagolysosome, characterized by a high acidity and large number of reactive oxygen species (ROS), reactive nitrogen species (RNS) and hydrolitic enzymes (Kinchen and Ravichandran 2008). *C. glabrata* is resistant to ROS by detoxifying the phagosome via the Ca1p catalase (Cuéllar-Cruz et al. 2008). However, this characteristic seems to be shared by all pathogenic fungals and is even found in *S. cerevisiae*, which is nevertheless not able to multiply in phagocytes (Seider et al. 2011). *C. glabrata* inhibits phagolysosome maturation by preventing acidification and phagosome fusion with the lysosome (Seider et al. 2011). These mechanisms seem to be an active strategy involving chromatin remodeling (Rai et al. 2012). *C. glabrata* persists in phagocytes and does not actively escape from them, meaning that it does not induce apoptosis or breaches in their membranes. Instead, yeast cells are released when the phagocyte ruptures and the event correlates with a high concentration of intracellular *C. glabrata* (Seider et al. 2011). Again, the majority of genes possibly involved in starvation resistance and phagocytic stress are shared by all *Nakaseomyces* (Gabaldon and Carrete 2016).

4.3.Penetration and adhesion to deep tissues

The ability to form hyphae is a major advantage for the protrusion of a fungi into deep tissues. This is particularly the case for *C. albicans*, which hyphae formation is necessary for its escape from phagocytes and for the passage through epithelia (Phan et al. 2000). The inability to form hyphae of *C. glabrata* suggests a passive invasion. This mechanism does not damage tissues and therefore induces a very weak inflammatory response (Jayatilake et al. 2006, Li and Dongari-Bagtzoglou 2007). However *C. glabrata* is still able to cross the epitheliums in-vitro (Perez-Torrado et al.

2012) and in-vivo (Jacobsen et al. 2011). Invasion of deep tissues could be due to epithelial endocytosis (Li and Dagari 2007) as well as co-infection with other species as *C. albicans* (Soll 2002; Coco et al 2008) (Alves et al. 2014). Most of the time, access to deep tissue is possible via medical-induced breaches in epithelia (Pelroth 2007).

C. glabrata has the ability to adhere to epithelia but also to an abiotic surface (Izumida et al. 2014). It is therefore frequently found on medical devices such as catheters or dental prostheses for example (Ramage et al. 2005, Tay et al. 2014). *C. glabrata* contaminates these different surfaces by forming a biofilm, which is defined as « a microbially derived sessile community characterized by cells that are irreversibly attached to a substratum or interface or to each other, are embedded in a matrix of extracellular polymeric substances that they have produced, and exhibit an altered phenotype with respect to growth rate and gene transcription » (Donlan and Costerton 2002). The composition of the matrix of biofilms formed by *C. glabrata* is currently elusive, and has only been described as rich in protein and carbohydrates such as ß-(1,3)-D-glucan (Silva et al. 2009). The formation of a biofilm considerably increases the resistance to antifungal cells within (Taff et al. 2013) and is considered to be a major risk of candidemia (Filler and Kullberg 2002). Therefore, catheter patients with fungal infections must absolutely have their implants removed for effective treatment (Ramage et al. 2005).

The adhesion of *C. glabrata* to a surface is made possible by the presentation of proteins on the cell-wall surface, called adhesins (Cormack et al. 1999). The genome of *C. glabrata* contains 67 genes coding for predicted adhesin-like proteins (de Groot et al. 2008). Adhesins are divided into 7 different subgroups among which the EPA (for Epithelial Adhesins) subgroup is most represented with 18 to 23 genes depending on the isolate (de Groot et al. 2008). The expression of EPA gene is very low in in-vitro culture under standard conditions (Timmermans et al. 2018). Most of these genes are found in the subtelomeric regions and their expression is regulated by telomeric silencing (Castaño et al. 2005). This chromosomal localization predisposes EPA genes

to undergo chromosomal rearrangements, which can be the source of their expansion and the variability in their number depending on the strain studied.

The number of copies of *EPA* genes is suggested to correlate with the pathogenicity of a strain. This has been demonstrated by the sequencing of the genome of a clinical isolate presenting a large number of *EPA* gene duplicates (Vale-Silva et al. 2017). The variation in the number of *EPA* genes is also in correlation with the degree of infectivity of the species within the *Nakaseomyces*. Indeed *C. glabrata,* which is considered to be the most pathogenic species of the clade contains 17 to 23 *EPA* genes and is followed by *C. bracariensis*, with 12 *EPA* genes, *C. nivariensi*s with 9 *EPA* genes and only a single copy in the genome of the environmental species *N. delphensis (Gabaldón et al. 2013)*. The adherence of *C. glabrata*, via *EPA* gene expression, is likely to represent the major reason for of its infectivity.

As mentioned earlier, *C. glabrata* exhibits features that are more referring to an opportunistic colonizer than to a highly virulent pathogen. Unlike *C. albicans*, also considered an opportunistic pathogen, *C. glabrata* induces very little damage to the colonized tissues. Instead, its infection strategy resides more in stealth and passive evasion. *C. glabrata* does not appear to show clear features resulting from adaptation to humans and the comparative genomics of the *Nakaseomyces* clade supports this idea (Gabaldon and Carrete 2016). Moreover, the recent analysis of 33 different isolates of *C. glabrata* would even suggest that human are only a secondary niche (Carreté et al. 2018).

Although the virulence level of *C. glabrata* can be considered as low, the high mortality rate associated with its infections is a real concern. As we will discuss bellow, the threat represented by *C. glabrata* is based on the combination of strong adhesion capacity and an increased resistance to antifungal agents.

5. Antifungals and resistance

Microorganism-related infections are considered as a major public health issue when the medical arsenal available is not sufficient to cope with its pathogenicity. Fungal infections caused by *Candida* species have historically been treated by two major classes of antifungals : polyenes and azole compounds (Sheehan et al. 1999). These two families of molecules act on ergosterol, a sterol specific to fungi, which constitutes the plasma membrane. Additionally, echinochandins, a new class of antifungals molecules targeting the cell wall, is commonly used since their approval in 2003.

All these molecules have proven their efficacy in the treatment of fungal infections, but seems no longer efficient regarding the increased frequency of resistance events in *C. glabrata.*

5.1. Polyenes

Polyenes molecules, represented by amphotericin B and nystatin, bind directly to the membrane ergosterols. The binding induces the depolarization of the sterol and thus the formation of pores (Lemke et al. 2005). As a result, the cells die by emptying themselves of their contents via the formed pores. Polyenes present the inconvenient of being non-specific and strongly inducing nephrotoxicity (Pound et al. 2011). Given the consequences, these antifungals are used only as a last resort for the treatment of invasive fungal infection. *C. glabrata* appears sensitive to the action of polyenes, however several cases of amphotericin B resistance have been studied (Vandeputte et al. 2007). This resistance has been described as concomitant with low ergosterol levels in the plasma membrane linked to a mutation in the *ERG6* gene involved in ergosterol biosynthesis (Vandeputte et al. 2008).

5.2. Azole compounds

The mechanism of action of azole compounds, including fluconazole (the most commonly used), ketoconazole (the first to have been marketed) and itraconazole, is based on inhibition of ergosterol synthesis (Ghannoum and Rice 1999). Azoles block the P-450 lanosterol demethylase enzyme, encoded by the *ERG11* gene in *C. glabrata*, thereby preventing the conversion of lanosterol to ergosterol. As previously mentioned, *C. glabrata* exhibits an intrinsic low susceptibility to azoles. This resistance is due to modifications of the P-450 lanosterol demethylase enzyme leading to a loss of affinity for azoles (Silva et al. 2016). In addition, *C. glabrata* actively resists via an energy-dependent efflux pump mechanism that exports the molecule out of the cell. These efflux pumps are composed of membrane transport proteins belonging to the ATP binding cassette transporters family (Sanglard et al. 1999). In *C. glabrata* these proteins are encoded by the genes *CDR1* and *CDR2* which expression is controlled by the transcription factor Pdr1p (Vermitsky and Edlind 2004). Overexpression of *CDR* genes increases tolerance of *C. glabrat*a to azoles (Ferrari et al. 2011). The resistance of *C. glabrata* clinical isolates to azoles is amplified by the acquisition of several mutations in the *PDR1* gene leading to its hyperactivity (Ferrari et al. 2009). The acquisition of this resistance is very quick and has been described as appearing after only 4 days of azole treatment of naïve strains (Borst et al. 2005). Azoles have been the most widely used molecules in the treatment of candidiasis and candemia. As previously mentioned, this extensive use seems to have prompted the emergence NAC species by : i) the selection of intrinsically azole tolerant species such as *C. glabrata*; and ii) the selection of strains that rapidly acquired mutations involving higher tolerance. As a result of the increased resistance events frequency, azoles are no longer considered as first-line agents for the treatment of *C. glabrata* infections (Pappas et al. 2015).

5.3. Echinochandins

Echinochandins, including caspofungin and micofungin, constitute a new class of antifungals approved since 2003 for the treatment of *Candida* spp. related infections (Johnson and Perfect 2003). The efficiency of echinochandins coupled with the low associated side effects have made this class the first-line agent for the treatment of candidemia, according to European and American guidelines (Cornely et al. 2012, Pappas et al. 2015). These molecules block ß-(1,3)-D-glucan synthase, preventing the synthesis of essential glucan for the fungi cell wall (Douglas et al. 1997). The ß-(1,3)-D-glucan enzyme complex in *Candida glabrata* consists of a protein complex which subunits are encoded by the *FSK1*, *FSK2* and *FSK3* genes (Katiyar et al. 2012). Clinical isolates with acquired resistance to echinochandins have rapidly emerged and are linked to the acquisition of mutations in *FKS* genes (Shields et al. 2012, Alexander et al. 2013). In addition, several isolates already described as azole resistant have shown combined resistance to echinochandins (Pfaller et al. 2012).

Combination of strong adhesion capacity and rapid acquisition of resistance has contributed to make *C. glabrata* a major health concern, especially alarming for hospital care (Schelenz 2008, Savastano et al. 2016). Regarding the mechanisms involved in this combination, we can find a common denominator : the dynamics of the genome of *C. glabrata*.

Features including its plasticity (duplication of *EPA* genes) and its mutable potential (*ERG*, *FKS* genes) indicate that the role of the genomic structure of *C. glabrata* is prominent in its biology.

6. The genome of *Candida glabrata*

Genome sequencing of *C. glabrata* was first performed in 2004 with the CBS138 strain (Dujon et al. 2004). Phylogenetic analysis confirmed previous studies indicating that *C. glabrata* is closer to *S. cerevisiae* than to *C. albicans* (Kurtzman and Robnett 1998). The sequenced genome size is 12.3 Mb, formed by 13 chromosomes. The genome was annotated by homology with *S. cerevisiae*, and completed by transcriptomics data (Linde et al. 2015). *C. glabrata* contains 5293 protein coding genes; 6 ribosomal RNA genes, 68 non-coding RNA genes and 230 tRNA genes. A large number of genes of *S. cerevisiae*, 4372, present an ortholog in *C. glabrata*. Protein alignment of orthologous protein-coding genes between *S. cerevisiae* and *C. glabrata* indicates an amino acid identity of 54%. Although *C. glabrata* is considered phylogenetically close to *S. cerevisiae*, it should be noted that this similarity is equivalent to the one between humans and zebrafish (Gabaldón et al. 2013). As all *Nakaseomycetes*, the genome of *C. glabrata* is relatively small compared to *S. cerevisiae* with a size difference of 11.7%, it contains almost twice as less introns (129 instead of 287 in *S. cerevisiae*), and no transposons. Compared to *S. cerevisiae*, *C. glabrata* appears to have lost a significant number of genes after the WGD event. The majority of these genes are also absent in *Nakaseomyces.* and are as follows: *BNA* genes, involved in *de novo* biosynthesis of nicotinic acid; *PHO* family of acid phosphatase genes; *PAU* genes family coding for mannoprotein cell-wall, involved in anaerobiosis; *DAL* genes involved in alantoin degradation; the *CRF1* gene, repressor of the ribosomal proteins expression regulation; the *GAL* genes, involved in the pathway of galactose, the *JEN1* gene, coding for a lactate permease, copies of the *COX* (A and B) gene pairs, coding for cytochrome oxidase and the *CYC1*/*CYC7* pair coding for cytochrome C1 and its isoform.

The loss of these genes does not seem to be a characteristic of *C. glabrata*. Nevertheless, *C. glabrata* is the species among *Nakaseomyces* that presents the largest tandem arrays. These structures are genomic rearrangements resulting from the tandem

duplication of a locus, leading to the formation of a gene family. The genome of *C. glabrata* contains five arrays: an array of eight *MNT3* genes, coding for alpha-1,3-mannosyltransferase; an array of eight *YPS* genes, coding for a protease required for cell wall remodeling and for survival in macrophages (Bairwa et al. 2014); an array of five unknown genes probably involved in carbohydrate metabolism; an array of three *PMU* genes coding for phosphomutase, involved in phosphate activity induced during phosphate-starvation (Orkwis et al. 2010); and an array of three genes coding for an aryl-alcohol deshydrogenase. The majority of these genes code for proteins that seem to confer an advantage to starvation and survival in macrophages, and hence could most likely play a role in the pathogenicity of *C. glabrata.*

C. glabrata is strictly haploid and asexual, even though it possesses all the genes necessary for meiosis, mating and mating-type switching (Gabaldón et al. 2013). Like all *Nakaseomyces* species, *C. glabrata* possesses the *HO* gene and the *HMR* and *HML* cassettes allowing mating-type switching. The *HO* gene is weakly expressed under in laboratory condition, but its overexpression in the type strain CBS138 allowed the switch from mating-type alpha to mating-type a (Edskes and Wickner 2013). Moreover the sequencing of 33 different isolates of *C. glabrata* provided evidence that genetic material exchange has recently occurred between different isolate, thereby validating the existence of an active sexual mechanism (Carreté et al. 2018).

The different isolates of *C. glabrata* show a high genomic diversity. This variability is reflected first of all by the karyotype which differs even within isolates from the same patient (Shin et al. 2007). Chromosomal changes are generally associated with polymorphism in their size (Muller et al. 2009) following reciprocal or non-reciprocal translocations of chromosomal arms, or inversions. Karyotype diversity is associated with aneuploidy. These modification of the chromosome number implies the generation of new small chromosomes containing a centromere and telomeres (Poláková et al. 2009), or the complete duplication of a chromosome (Carreté et al. 2018).

The karyotype variability was until then often associated with a response to antifungal agents, but spontaneous events of complete chromosome duplication have been observed in rich medium culture without any selection pressure (Carreté et al. 2018). *C. glabrata* isolates also exhibit high diversity in DNA content. Comparison of 33 isolates of *C. glabrata* with the reference strain CBS138 showed that 580 genes are unique to the type strain. The greatest genetic variations affect the genes encoding the proteins involved in adhesion, and phenotypically corroborate with the adhesion capacity of the isolates studied. Most of these genes are located in subtelomeric regions which plasticity seems to be linked to a protein involved in telomeric silencing, ESC1, and its sequence mutations (Carreté et al. 2018).

C. glabrata, as well as the other *Nakaseomyces*, have however a genomic characteristic that is invariable within isolates. Indeed, the non-coding RNAs composing the ribonucleoprotein complex of the spliceosome (U1 RNA); the maturation complex of the 18S-rRNA (U3 RNA); the telomerase (TLC1 RNA) and the RNase P (RPR1 RNA) contains large insertions increasing remarkably the size of their transcripts.

Material and Methods

A. Material

I. Primers and DNA fragments

Primers or DNA fragments (gene blocks) were purchased lyophilized, from Integrated DNA Technologies (IDT®), then were resuspended in H_2O mQ to a final concentration of 100 µM. Primers and DNA fragments that have been used are listed in the following table.

Primers (1/2).

Purpose	Fragment	Orientation	Sequence (5'-3')	Tm (°C)
Cagl0K07678-*ShBle*	Upstream homology	Forward	TGACTTGGGTGTCTCTTTATTTGTATTGC	68,4
		Reverse	AGCATCTCTGCCCATATCATAGGA	68,3
	Sh ble	Forward	ATCCCTTCCTATGATATGGGCAGAGATGCTC CCACACACCATAGCTTCAAAATG	66,9
		Reverse	AGTTCTCAATTCTTCCAGCGTAGACCTCCAGC TTGCAAATTAAAGCCTTCGAGC	69,5
	Downstream homology	Forward	TGGAGGTCTACGCTGGAAGAA	66,3
		Reverse	CGCAACGCCAATACGATATCATTAACG	72,3
Cagl0K07678-HA	Upstream homology - Synthetic Terminator	Forward	ACTCACCAGTCTCTTTTGAAACTGC	66
		Reverse	TTGAAGCTATGGTGTGTGGGTTTGAAAGATGATA CTCTTTATTTCTAGACAGTTATATATTA AGCGTAATCTGGAACGTCATATGG	64
	Sh Ble	Forward	CCCACACACCATAGCTTCAAAATG	68,3
		Reverse	CTGAGATGCTTGAACCAATGGGTGCTTGCAA ATTAAAGCCTTCGAGC	69,5
	Downstream homology	Forward	ACCCATTGGTTCAAGCATCTCAG	68,5
		Reverse	TTTCCGTAATGAAGCTGTCGATGC	70
GFP	Upstream homology- Synthetic promoter	Forward	CACACACAAGCATACTGACCC	63,6
		Reverse	CACCACCCCGGTGAACAGCTCCTCGCCCTTGCTC ACCATCCGGCGCCTCCACTCACGCCAACAGTGCT CTTTTATAAGC	65,8
	GFP	Forward	ATGGTGAGCAAGGGCGAG	66,9
		Reverse	TTAGGATCTCTTGTACAGCTCG	60,6
	Synthetic Terminator - *Sh Ble*	Forward	ACGAGCTGTACAAGAGATCCTAATATATAACTGTC TAGAAATAAAGAGTATCATCTTTCAAACCCACAC ACCATAGCTTCAAAT	66,1
		Reverse	GCTTGCAAATTAAAGCCTTCGA	66,4
	Downstream homology	Forward	GGGACGCTCGAAGGCTTTAATTTGCAAGCGAAAG CTCGATAATTCGAGGCAGTTCCAG	66,4
		Reverse	GCACCAAAGATATTATGTGTAGGTGT	64,6
	GFP-RT-qPCR	Forward	AGGACGACGGCAACTACAAG	64,6
		Reverse	AAGTCGATGCCCTTCAGCTC	66,1

Primers (2/2).

Purpose	Fragment	Orientation	Sequence (5'-3')	Tm (°C)
CgRPR1::*ShBle*	Upstream homology	Forward	GCCATCTTAAAAAGTCCCTACGATGC	58
		Reverse	AAGAGTCAGAACTCTCAAGAGTACCCA	58
	Sh ble	Forward	ACTCTTGAGAGTTCTTCCCACACACCATAGC TTCAAAAT	58
		Reverse	ACCACATGTAACCAACCGCGTCTTGCAAATT AAAGCCTTCGAG	58
	Downstream homology	Forward	ACGCGGTTGGTTACATGTGGT	58
		Reverse	CATGACAGTCACATCCCTAGACAGG	58
Chromosome B1	Insertion site	Forward	CTGGACAGCAACCGGGATC	58
		Reverse	CATGACAGTCACATCCCTAGACAGG	58
	Sh Ble	Forward	CCCACACACCATAGCTTCAAAATG	58

DNA fragments.

Purpose	Fragment	Sequence (5'-3')
Switch ScRPR1	Ty Terminator	TTTTTTTCTTATTTTTTGT
	CgRPR1 promoter	CATTAGATAGATATTGGGGCCATCTTAAAAAGTCCCTA CGATGCTCTGTTAGGACCCTCAGTGGTTCAGTGGTTTT TGAAGCTTCCCCTTCCATAGGGGCGACTTCCAGGGCC ATTGTACGAATGGAAGTGTTAACTTTTTGAAGCCACTA TTTAAAGAGGGAACGATCCAGAGTTCGAAAC
RPR2-HA	Neutral linker	GGTTCTGGTGGTGGTTCTGGTGGTGGTTCT
	3X-HA	TACCCATACGATGTTCCTGACTATGCGGGCTATCCCTAT GACGTCCCGGACTATGCAGGATCCTATCCATATGACGT TCCAGATTACGCT
Synthetic promoters	Sh ble promoter	CCCACACACCATAGCTTCAAAATGTTTCTACTCCTTTT TTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCG TACCACTTCAAAACACCCAAGCACAGCATACTAAATT TTCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCGT ACTAAAGGTTTGGAAAAGAAAAAAGAGACCGCCTCG TTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTT ATCACGTTTCTTTTTCTTGAAATTTTTTTTTTTAGTTTT TTTCTCTTTCAGTGACCTCCATTGATATTTAAGTTAATA AACGGTCTTCAATTTCTCAAGTTTCAGTTTCATTTTTC TTGTTCTATTACAACTTTTTTTACTTCTTGTTCATTAGA AAGAAAGCATAGCAATCTAATCTAAGGGGCGGTGTTG ACAATTAATCATCGGCATAGTATATCGGCATAGTATAAT ACGACAAGGTGAGGAACTAAACC
	CYC1 terminator	CACGTCCGACGGCGGCCCACGGGTCCCAGGCCTCGG AGATCCGTCCCCCTTTTCCTTTGTCGATATCATGTAATT AGTTATGTCACGCTTACATTCACGCCCTCCCCCCACAT CCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTG AAGTCTAGGTCCCTATTTATTTTTTTATAGTTATGTTAGT ATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTT CTGTACAGACGCGTGTACGCATGTAACATTATACTGAA AACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCT TTAATTTGCAAGGTTGTAATCGAGCTCGAATT
	Synthetic terminator	TATATAACTGTCTAGAAATAAAGAGTATCATCTTTCAA
	UAS C	TAGCATGTGA
	UAS E	TAGCATGTGA
	UAS F	GGCGCGCCCCTCCTTGAA
	Spacer 1	TTAATTAACTTGTAATATTCTAATCAAGCT
	Spacer 2	TAAGTT
	TATA-box	TATAAAAG
	Core promoter	GCACTGTTGGGCGTGAGTGGAGGCGCCGG

CRISPR-Cas9 single guide RNA.

Coordinates	Fragment	Sequence (5'-3')	PAM
Chromosome L : 880147-880167	sgRPR2	TAGAGTGCCAGTGCCACCAG	TGG
Chromosome L: 879849-879823	sgRPR1	TGAGAGTTCTGACTCTTACG	CGG
Chromosome B: 89861-89889	sgChrB1	CGAATTATCGAGCTTTCGTA	CGG

II. Microscope

All pictures of GFP-transformed yeasts were taken with the LSM700 confocal microscope from Zeiss. Images were analyzed with the ImageJ software.

III. Strains

E.coli

For cloning experiments, DH5α *E. coli strains* (F- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK-, mK+) *pho*A *sup*E44 λ- *thi*-1 *gyr*A96 *rel*A1) were successively washed in a cold calcium chloride (CaCl2) solution to induce chemo-competence.

S. cerevisiae

The *S.cerevisiae* S154 and S2 strains of *S. cerevisiae* were used for complementation experiments. These strains are diploid and isogenic to the reference strain S288C (Mortimer RK and Johnston JR, 1986). The S154 strain (BY 4719 x BY 4738) with the genotype *ura3Δ0/ura3Δ0, trp1Δ63/trp1Δ63* is auxotrophic for uracil (U) and tryptophan (T). The S2 strain was generated from the S154 strain, in which one the two allele of the *RPR1* gene was replaced by an *URA3* wild type copy. This strain, with the *ura3Δ0*/*ura3Δ0*, *trp1Δ63*/*trp1Δ63*, *RPR1*/*rpr1*Δ::*URA3* genotype is thus prototrophic for uracil and auxotrophic for tryptophan. The deletion in the *RPR1* locus includes the promoter and terminator sequences of the gene (positions -141 to +1550)

C. glabrata

C. glabrata strains used for experiments are the following : the CBS138 wild type strain, the auxotrophic strains ΔTL and ΔHTL (his3Δ::FRT leu2Δ::FRT trp1Δ::FRT) (Schwarzmüller et al. 2014) and the ΔHTL GFP and ΔTL Δ*caglk07678g::HIS3* strains generated in the laboratory. Cells were rendered chemo-competent by treatment with lithium acetate (LiAc) as described in (Istel et al. 2015).

Flies

Drosophila melanogaster white A5001 (wA5001) strain was used as a wild type control and was compared to the MyD88 (Tauszig-Delamasure et al., 2002) immuno-compromised strain, which is deficient for the activation of the Toll pathway. Flies were generously provided by Dr. Dominique Ferrandon

IV. Culture media

Bacteria

Bacteria were grown at 37°C with agitation in Luria Bertani (LB) medium, composed of 1% (w/v) peptone, 0,5% (w/v) yeast extract, 1% (w/v) sodium chloride (NaCl) (plus 1,4% (w/v) agar for plates). When needed, 100 µg/mL ampicillin were added to the medium.

Yeasts

Yeasts were cultured at 30°C with agitation in complete, poor, or selective medium. YPD rich medium is composed of 2% (w/v) peptone, 1% (w/v) yeast extract, 2% (w/v) glucose (plus 1,4% (w/v) agar for plates). When needed, 400 µg/mL to 1 mg/mL Zeocin™ (Invivogen) was added. Nitrogen starvation medium was composed of 0,017% (w/v) yeast nitrogen base (YNB) without ammonium sulfate ($(NH_4)_2SO_4$) or amino acids, 2% (w/v) glucose and 0.025 % (w/v) $(NH_4)_2SO_4$. The selective medium Sc-His is composed of 0.067 % (w/v) YNB with $(NH_4)_2SO_4$ without amino acids, 2 % glucose, 0.0078 % (w/v) CSM -His (*Complete Supplement Mixture* – His) (plus 1.4 % (w/v) agar for plates)

Flies

Drosophila fly strains were raised at 25°C, under 14h of daylight and 60% humidity. They are fed a medium composed of 6.4 % (w/v) organic cornmeal (Priméal), 4.8 % (w/v) crystallized sugar (Tereos Syral), 1.2 % (w/v) yeast (Bio Springer), 0.48 % (w/v) agar (Sobigel), 0.42 % (w/v) nipagin (VWR Chemicals). Fly food was renewed every 8 days.

V. Plasmids

Yeast plasmids were generated from the vector plasmids pRS313, and pRS314. These plasmids include : an origin of replication from *E.coli,* the centromeric sequence *CEN6_ARS4* designed to keep a low plasmid copy number in the yeast, and the *LEU2* (pRS313*)* or *TRP1* (pRS314) gene for the selection of transformed yeasts. The plasmids derived from the pRS314 vector are described hereafter :

- pRS-*CgRPR1* : plasmid including the wild type locus *RPR1 from C. glabrata* (*RPR1Cg*). The cloned region comprises the sequence for the RPR1 RNA and the promoter and terminator sequences.
- pRS-promSc-*CgRPR1* : plasmid including the Cg*RPR1* locus under the *S. cerevisiae* promoter sequence.
- pRS-Sc*RPR1* : plasmid including the wild type locus *RPR1 from S. cerevisiae* (Sc*RPR1*), with promotor and terminator sequences.

The pRS315-Zeo (ZeoR/AmpR/CEN) centromeric plasmid was used to amplify the *Sh ble* gene for zeocin resistance. The pRS313-GFP (LEU/AmpR/CEN) centromeric plasmid was used to test the efficiency of synthetic promoters.

VI. Antibodies

For western blots, mouse primary antibody anti-HA-HRP (Roche), anti-GFP (Abcam) and anti-mouse-HRP (GE Healthcare) were diluted in a solution containing 1X PBS, 0,1 (v/v) Tween20 and 5% dry milk.

B. Methods

I. Yeast transformation

1 to 3 µg of DNA was transformed in the CBS138 *C. glabrata* strain with electroporation, as described in (Istel et al. 2015). Generation time for zeocin-selected cells was extended to 6h, according to . Transformed cell were plated on selective media YPD supplemented with 400 µg/mL (Alderton et al. 2006) to 1 mg/mL Zeocin™ or Sc -His.

II. Drosophila infection

5 to 7-day old female flies were selected and transferred to a yeast-free medium 48h prior to the infection, at 29°C. *C. glabrata* in stationary phase were diluted in 250 µL of fresh medium at an OD_{600nm} of 15. Flies were anesthetized with CO2 and pricked under the wing with a needle soaked with *C. glabrata.* Infected flies were placed at 29°C and their survival was monitored every 24h. Survival curves were analyzed with the software Prism 6 (GraphPad).

III.Sporulation and dissection of tetrads

A drop of diploid cells harvested in exponential phase of growth was placed on AcK agar medium and incubated 4 days at 30°C. The formed cells and asci were resuspended in 200 µL sterile water and the digestion of the asci wall was induced by the addition of 75µL of *Arthobacter luteus* lyticase (20 mg/mL Sigma-Aldrich®). After 15 min incubation at 25°C, 100 µL of the suspension were deposited on agar medium. The spores released from the asci but still grouped into tetrads were then collected and isolated using a micromanipulator (Singer ®).

IV. Plasmids amplification

5 µL of ligation reagent was added to 50 µL of DH5α chemo-competent bacteria. Cells are incubated on ice for 30 minutes then heat shocked at 42°C for 45 secondes, followed by 2 minutes on ice. 500 µL of LB medium were added to the transformed cells, which were then incubated with agitation at 37°C for 1h. Bacteria were plated on LB medium containing 100 µg/mL ampicillin. Plasmid DNA was extracted with the « GeneJET Plasmid Miniprep » kit from Thermo Scientific

V. Genomic DNA extraction

Yeasts were cultured overnight in the appropriate medium at 30°C and genomic DNA was extracted as described in (Harju et al. 2004) For optimal extraction, DNA was precipitated with isopropanol for 30 minutes at -80°C.

VI. RNA extraction

A single colony of *C. glabrata* was cultured in 45 mL of liquid medium with agitation, at 30°C, until it reaches exponential growth phase (OD_{600nm} = 3). Cells were centrifuged at 9000 g for 5 minutes and the pellet was re-suspended in 1 mL of « TRI Reagent® » (Sigma-Aldrich). The solution was transferred to a « Lysing Matrix Y » (MP Biomedicals) tube and vortexed for 5 minutes. The lysed cells were transferred in a fresh tube containing 200 µL of cold chloroform. After homogenization and incubation at room temperature for 15 minutes, tubes were centrifuged at 12 000 g for 15 minutes, at 4°C. RNAs were isolated following the manufacturer's instruction.

VII. Protein extraction

A single colony of *C. glabrata* was cultured in 45 mL of liquid medium with agitation, at 30°C, until it reaches exponential growth phase (OD_{600nm} = 3). Yeasts were centrifuged at 4200 g for 5 minutes and the pellet was re-suspended in 900 µL of lysis buffer (20 mM Tris-HCl [pH 7.2], 50 mM NaCl, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 0.5 % (v/v) NP40, 1 tablet « PierceTM Protease Inhibitor Tablets, EDTA- Free » (Thermo Scientific). The solution was transferred to a « Lysing Matrix Y » (MP Biomedicals) tube and the lysis was realized with the « FastPrep-24TM 5G » kit from MP Biomedicals. Two lysis cycles of 40 seconds followed by 5 minutes on ice were performed. Cells were then incubated on ice for 30 minutes and centrifuged at 99 000 g for 30 minutes. Supernatant was harvested and the concentration was determined with a Bradford assay.

VIII. PCR amplification

PCR amplification were performed from a plasmid (10 ng/µL), from genomic DNA (50 ng/µL) or from 1 µL of a re-suspended colony of *C. glabrata* diluted in 20 µL of sterile water with 0,5 µM of forward and reverse primers, 200 µM dNTPs, the Phusion High-Fidelity 1X buffer and 0.4 U of PhusionTM High-Fidelity DNA polymerase (Thermo Scientific). Amplification was performed according to the manufacturer's instruction, and hybridation temperatures were adapted to the primers used. In addition, a 3 minute-denaturation step was realized for colony PCRs. Amplification products were separated on an agarose gel (1,5% (w/v) agarose) by electrophoresis, in 1X TBE buffer. The gel was incubated in ethidium bromide and analyzed under UV light. PCR products were purified with the « Wizard® SV Gel and PCR Clean-Up System » kit (Promega).

IX. DNA fragment assembly by « Gibson Assembly® »

PCR fragments were assembled together or cloned in a plasmid with the « Gibson Assembly® Master Mix » kit (New England Biolabs®). Assembled fragments were amplified by PCR, and transformed into the CBS138 *C. glabrata* strain. Assembled plasmids were transformed in the DH5α chemo-competent bacteria.

X. Reverse transcription

1 µg of total RNA was retro-transcribed with the « Maxima First Strand cDNA Synthesis Kit for RT-qPCR with dsDNase » kit (Thermo Scientific). For each sample, a control condition without retro-transcriptase (RT-) and a control without RNA (No Template Control, NTC) were added.

XI. Quantitative PCR (qPCR)

RT-qPCR reaction mix was prepared in a total volume of 10 µL, from 12,5 ng cDNA and with the primers described in (mettre le numéro du tableau). The « Maxima SYBR Green/ROX qPCR Master Mix » kit (Thermo Scientific) was used following the manufacturer's instructions. An additional step was performed in order to calculate the melting temperature, by increasing the temperature by 0,5°C every 5 seconds, from 55 to 95°C. Measurements were automatically executed by the thermocycler « CFX96 TouchTM Real- Time PCR Detection System » (BioRad Laboratories) and analyzed with the CFX Manager (BioRad Laboratories) et Prism 6 (GraphPad) softwares.

XII. Western blot analysis

Protein extracts were boiled at 95°C for 5 minutes in loading buffer (0.015 % (w/v) bromophenol blue, 5 % (v/v) glycerol, 1 % SDS). 30 µg of proteins were loaded on a « NovexTM WedgeWellTM 10% Tris-Glycine » gel (Invitrogen) in a 1X TGS buffer and electrophoresis was run at 110V for 15 minutes, then 170V for 1h. Proteins were transferred on a nitrocellulose membrane with transfert buffer (1X TGS, 20% (v/v) EtOH) at 90V for 45 minutes. The membrane was incubated with agitation in milk for 20 minutes at room temperature, then for 2h in primary antibodies (anti-HA-HRP ; 1:5 000) or overnight, at 4 °C (anti-GFP ; 1:1 500). The membrane was washed 3 times with PBS-T (1X PBS, 0.1 % (v/v) Tween-20) for 10 minutes prior to the incubation with the secondary antibody anti-mouse-HRP (1:10 000 ; for GFP detection) for 4h. The membrane was washed 5 minutes in PBS-T and proteins were revealed by chemoluminescence with the « ECLTM Prime Western Blotting System » kit (GE Healthcare).

Chapter I. Characterization of the nuclear RNase P of *Candida glabrata.*

Context

Studies conducted on *Candida glabrata* were mainly driven by a need to understand its pathogenicity and to subsequently find adequate tools in order to fight this threat. The constant search for particularities within *C. glabrata* and related species, finally shed light on unexpected features reconsidering theories hitherto admitted. On a broader scale these findings have provided new keys to understand the biology of opportunistic pathogens, the evolution of fungi and even the molecular biology of eukaryotes.

In this context, I have attempted throughout my thesis to characterize one of the peculiarities of *C. glabrata* : its **remarkably large nuclear Ribonuclease P**. First identified in 2005, the RNA moiety of this Ribonuclease P (RNase P) contains three major insertions, conferring a total length about 3 times larger than the average eukaryotic RNase P RNA (Kachouri et al. 2005). This finding is in contradiction with the evolution of eukaryotic RNase P, which tend to restrict their RNA moiety and increase their protein complexity. This chapter reports my investigation of the role of these insertions in the biology of *C. glabrata* and is organized as follows:

i) An overview of the Ribonuclease P complex and its role within the three domains of life, which will focus mainly on eukaryotic RNase P.
ii) The results section, intentionally merged with the discussion to provide a better insight of the methodological choices made during this study.
iii) A conclusion summarizing each results and placing them in a broad scope

Overview

I. The Ribonuclease P

Ribonuclease P (RNase P) is an endonuclease enzyme whose main activity resides in the processing of the 5′-end of the precursor transfert RNA (pre-tRNA). This enzyme is widespread in the 3 domains of life (Archea, Bacteria and Eukarya), in the form of a ribonucleoprotein complex, consisting of an RNA component associated with one or more proteins depending on the organism (Evans et al. 2006). The catalytic activity is provided by the non-coding RNA subunit of the complex, which is by definition a ribozyme (**ribo**nucleotide en**zyme**). Although group I intron derived from *Tetrahymena* was the first RNA described as possessing catalytic activity (Kruger et al. 1982), the RNA subunit of the RNase P is in fact the first described ribozyme acting in *trans* (Guerrier-Takada et al. 1983). To date, RNase P RNA, and ribosomal RNA are the only RNAs described as universal true enzymes capable of performing multiple catalytic cycles in vivo. The strong structural homology of the RNase P RNA subunit across all domains of life suggests its presence within a universal common ancestor (Chen and Pace 1997). RNase P is even considered today as a remanence of the pre-biotic era, the « RNA world », in which chemical reactions were only catalyzed by RNA molecules (Altman and Kirsebom 1999).

I.1. Catalytic activity of the RNase P.

RNase P is a site-specific endoribonuclease, catalyzing the hydrolysis of a phosphodiester bond between two nucleotide of a substrate RNA, thereby generating 5' monophosphate and 3' hydroxyl ends (Kazakov and Altman 1991) (**Figure 3.1**). This reaction is carried out by the RNA component of the RNase P, whose catalytic domain (P4 domain, highly conserved in all RNase P) stabilizes a type 2 nucleophilic substitution (SN2) at the substrate cleaving site. This reaction is (suggested as carried out by coordination of a bivalent metal ion (Mg2+ for example) with a water molecule as a nucleophile. RNase P requires a certain concentration of divalent metal ion to ensure its catalytic activity, and is therefore described as a metalloenzyme. Directly involved in cleavage, divalent metal ions are also necessary for correct folding of the RNA subunit of RNase P, and substrate recognition (Beebe et al. 1996).

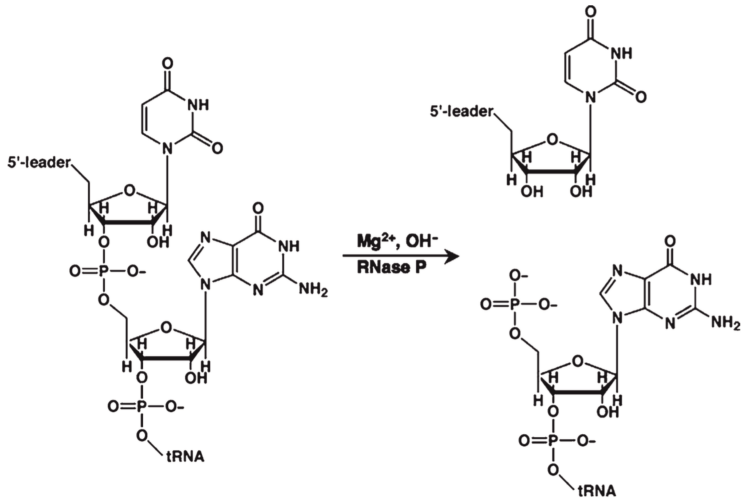

Figure 3.1 Hydrolysis of the phosphodiester bond between tRNA and the 5'-leader sequence by neutrophilic substitution. RNase P is a metalloenzyme using metallic divalent cation, such as Mg2+, to hydrolyse the phosphodiester bond between two ribonucleotides, releasing 5' monophosphate and 3' hydroxyl ends. From (Koutmou et al. 2010).

I.2. Substrate recognition.

Studies on the recognition of the RNA substrate by RNase P have focused mainly on the main RNA substrate, the premature tRNA (**Figure 3.2**). In the 3 domains of life, the tRNA is transcribed as a precursor with a 5' leader sequence processed by the RNase P and a 3' trailer sequence processed by the RNase E. In bacteria, both the RNA and protein subunits of RNase P contribute to the tRNA recognition, but the sole RNA subunit has been shown to be sufficient to induce a specific cleavage in vitro (Guerrier-Takada et al. 1983). The protein subunit interacts specifically with the 5' sequence leader (Zahler et al. 2003). The RNA of the RNase P directly recognizes the RCCA pattern; the nucleotides of the TΨC loop and the D loop; and finally the acceptor-stem (Kirsebom 2007). Concerning the eukaryotic RNase P, the precise mechanism of substrate recognition is less described compared to bacteria. However it was suggested that rather than specifically recognizing elements of the tRNA, recognition of the tRNA is based on a « measuring mechanism » of the size resulting from the stack of the T-domain and the acceptor-stem (Marquez et al. 2006). This mechanism therefore suggests a high potential to recognize other substrates than tRNA precursor.

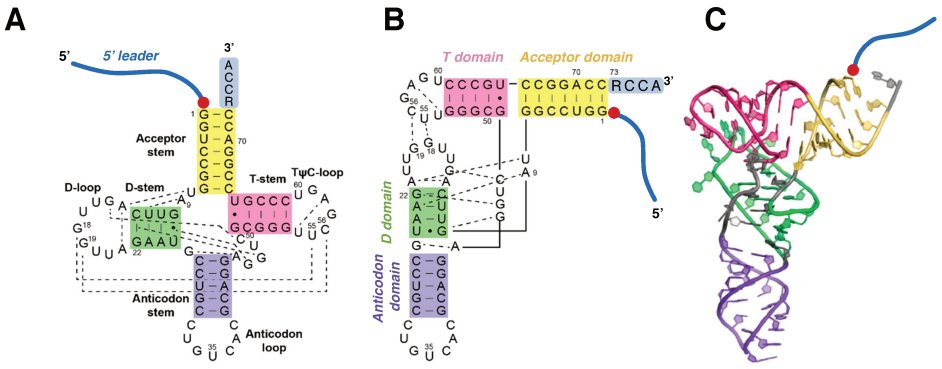

Figure 3.2 Structure of a consensus pre-tRNA presenting a 5' leader sequence.

A and **B** : secondary and tertiary structures of *B. subtilis* $tRNA^{Asp}$ with tertiary interactions represented by dashed lines. **C** : crystal structure of *E. coli* $tRNA^{Phe}$. RNase P cleavage site is represented by a red dot. Adapted from (Klemm et al. 2016)

I.3. Substrate diversity

Beyond tRNA substrate, RNase P is also involved in the recognition and cleavage of many other RNAs. Bacterial RNase P is involved in the cleavage of RNA substrate adopting tRNA-like structure as : transfert-messenger RNA (tmRNA) precursors; Turnip Yellow Mosaic Virus (TYMV) viral RNA, RNA from ColE1 plasmid; and long-nuclear retained RNA (Giegé et al. 1993, KoMine et al. 1994, Jung and Lee 1995, Wilusz et al. 2008). These data therefore indicate that the tertiary structure seems more important than the sequence itself. However, bacterial RNase P can also target substrates that do not exhibit a pre tRNA-like form such as the 4.5S RNA and bacteriophage Φ80-induced RNA (Bothwell et al. 1976, Peck-Miller and Altman 1991). Their recognition involve the protein subunit of the complex. Finally, bacterial RNase P is also involved in the maturation of polycistronic mRNA and bacteriophage RNA (Alifano et al. 1994, Hartmann et al. 1995). Concerning eukaryotes, temperature sensitive RNase P mutants in *S. cerevisiae* have revealed a panel of non-tRNA substrate cleaved by RNase P including: mRNAs encoding subunits of the three polymerase RNAs, C/D box snoRNA and intergenic regions transcripts (Marvin et al. 2011).

II. Evolution of the RNase P

RNase P diverged widely through evolution, whether in the sequence of the RNA subunit, protein composition or the type of substrates. Despite this diversity, conservation can be find at the RNA component structure of the RNase P, which core sequence and secondary structure exhibit a consensus form within all living organisms (**Figure 3.3**) (Ellis and Brown 2010). This structure contains five universally conserved regions (CR-I to CR-V). Interactions between CR-I and CR-V regions form the catalytic domain P4, the most conserved domain within all RNase P. The CR-II and CR-III regions have a universally conserved location in a loop formed between the P10 (or P11 according to the RNA) and P12 domains. Together the CR-I , CR-IV and CR-V regions are part of the catalytic domain (C-domain); CR-II and CR-III belong to the specificity domain (S-domain) involved in substrate recognition.

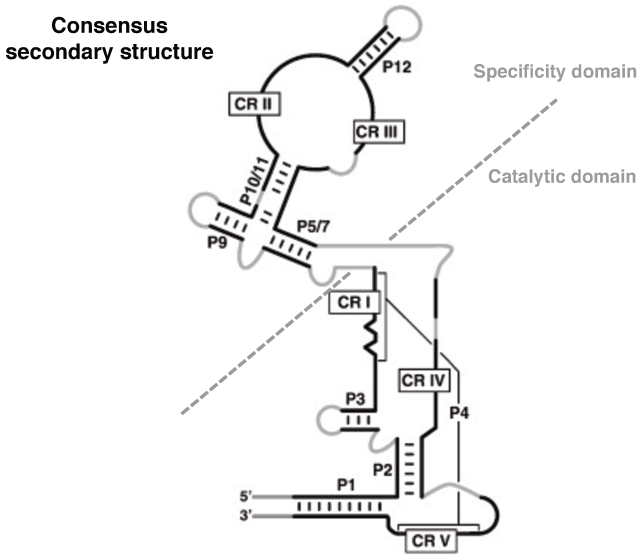

Figure 3.3. Consensus secondary structure of bacterial, archaeal and nuclear eukaryotic RNase P RNA. Adapted from (Ellis and Brown 2010).

II.1. Bacterial RNase P.

Bacterial RNase P is composed of an RNA subunit of about 400 nt , associated with a small 14 kDa protein, named C5 protein (Brown 1998).

A. RNA subunit

The secondary structure of bacterial RNase P RNA is highly conserved among species and can be divided into two categories (**Figure 3.4**) : Ancestor type (A-type) such as *Escherichia coli* RNase P RNA, and Bacillus type (B-type) such as *Bacillus subtilis* RNase P RNA. These two structures are very similar and have been described as exchangeable *in vivo* (Wegscheid et al. 2006). The type-A RNase P RNAs are the most common in bacteria. These RNAs have specific domains not found in the other organisms: domains P6, P13, P14, P16 and P17 (Haas and Brown 1998). These domains are crucial for RNase P since they create long-range interactions stabilizing the tertiary structure of the RNA. The P16 and P17 domains interact with the bulge between P7 and P5 forming the P6 helix. The P13 domain interacts tertiary with P12 and P14 interacts with P8 domain. B-type RNase P RNAs do not have P16/P17/P6 and P13/P14 domains but instead exhibit P5.1/P15.1 and P10.1 structures that seem to have the same structural stabilization role (Haas and Brown 1998).

B. Role of the bacterial RNase P protein subunit.

The C5 protein is highly conserved among bacterial species but does not appear to play the same roles within the holoenzyme. Although in both cases C5 protein is involved in the recognition of the 5' leader sequence of a pre-tRNA, in *E. coli* the protein stabilizes the secondary structure of the RNA subunit (Westhof et al. 1996) while in *B. subtilis* the protein increases the substrate specificity of the RNase P (Crary et al. 1998).

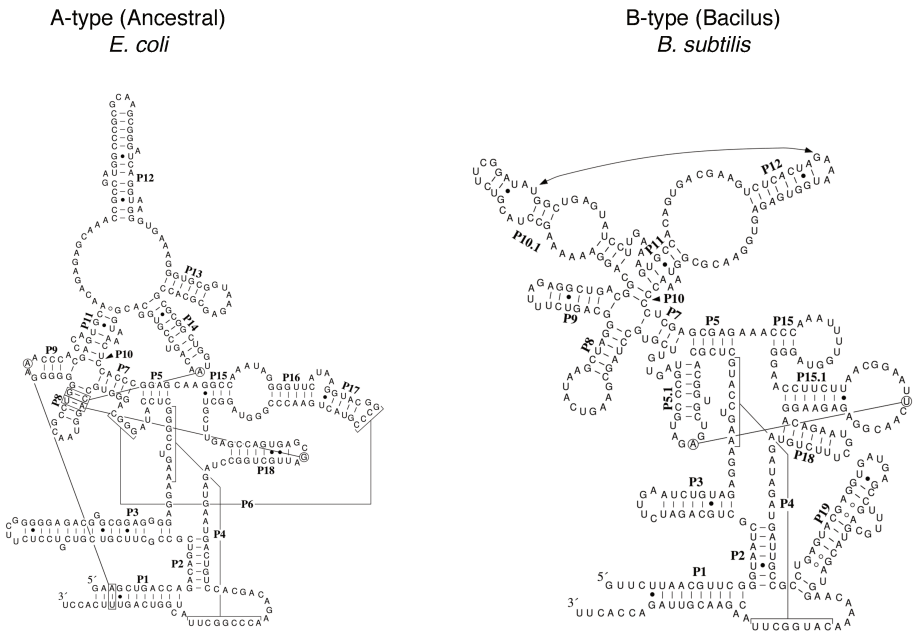

Figure 3.4 Secondary structure of the two classes of bacterial RNase P RNA. Adapted from (Ellis and Brown 2010).

II.2. Archaeal RNase P.

Archaea have an intermediate place in the evolution of life. Neither bacterial, nor eukaryotic, archaea still present similar characteristics with these two types of organisms. The archaeal RNase P is consistent with the intermediate nature of the organisms since protein subunits are close to the eukaryotic RNase P, and RNA moiety presents similar features to bacteria, with two main classes of RNase P RNA : Ancestral type (A-type) similar to the bacterial A-type; and the type-M RNase P RNA (figure 3.5) (Jarrous and Gopalan 2010). The M-type of archaea is closely similar to A-type but lacks the P8 and P16/P17/P6 domains. Compared to bacterial RNase P, most archaeal RNase P RNAs have lost complexity but have acquired a increased number of associated proteins. These proteins, which number varies between 2 and 5, have a strong homology with eukaryotic RNase P proteins (Jarrous and Gopalan 2010). The role of these proteins seems to be linked to the stabilization of the three-dimensional structure of the RNA subunit, in response to the absence of domains present in bacteria, and might be involved in substrate recognition.

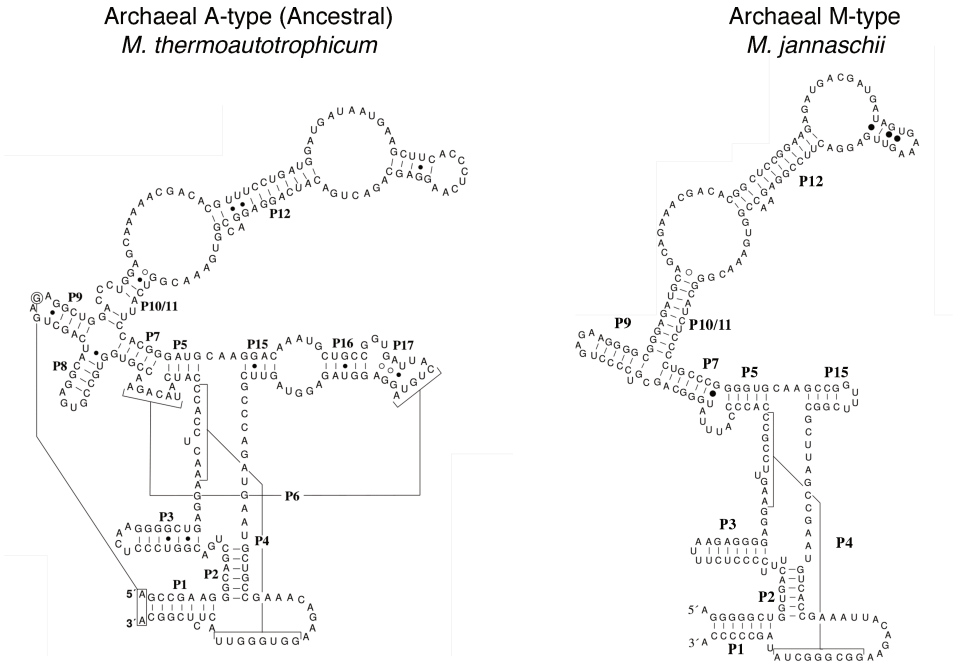

Figure 3.5 Secondary structure of the two classes of archaeal RNase P RNA. Adapted from (Ellis and Brown 2010).

II.3. Eukaryotic RNase P.

The RNase P activity was divided and localized according to the sub-cellular compartments within the eukaryotes. This partitioning gave rise to the emergence of various forms of RNase P which are specialized according to their localization. Eukaryotic cells present a nuclear RNase P and organellar RNase P. Moreover the nuclear RNase P shares 80% to 90% of its protein components with the nuclear Rnase MRP (historically named **M**itochondrial **R**Nase **P**) which is widespread in eukaryotes (Marquez et al. 2006). The RNase MRP RNA has a similar structure to RNase P, and sequence conservation indicates that it has evolved from RNase P. Nevertheless, RNase MRP has a different role from RNase P since it matures ribosomal RNAs but does not overlap with the specific substrate of the RNase P (Esakova and Krasilnikov 2010).

A. Nuclear RNase P.

Since this thesis focuses on yeasts, the RNase P of the yeast model *S. cerevisiae* will be mainly described here. The RNA component of *S. cerevisiae* RNase P (named RPR1 for **R**Nase **P** **R**NA) has a size of 369 nt. and contains all the domains universally found in the consensus structure (**Figure 3.6**) (Marquez et al. 2006). The domains for tertiary interaction stabilization commonly found in bacteria or archaea are absent in eukaryotic RNase P RNA. Some domains are located at the same position but they share no similarity of sequence, structure or function with bacterial counterparts. The nomenclature of these domains is often modified to avoid confusion and are called "eukaryal pair region", abbreviated as "Ep". The RNA component of the eukaryotic RNase P contains the eP8, eP9, eP15 and eP19 domains. The RNA component of *S. cerevisiae* RNase P, as well as all eukaryotic RNA of RNase P, has the particularity of presenting the P3 domain, essential for the docking of protein subunits (Perederina et al. 2007).

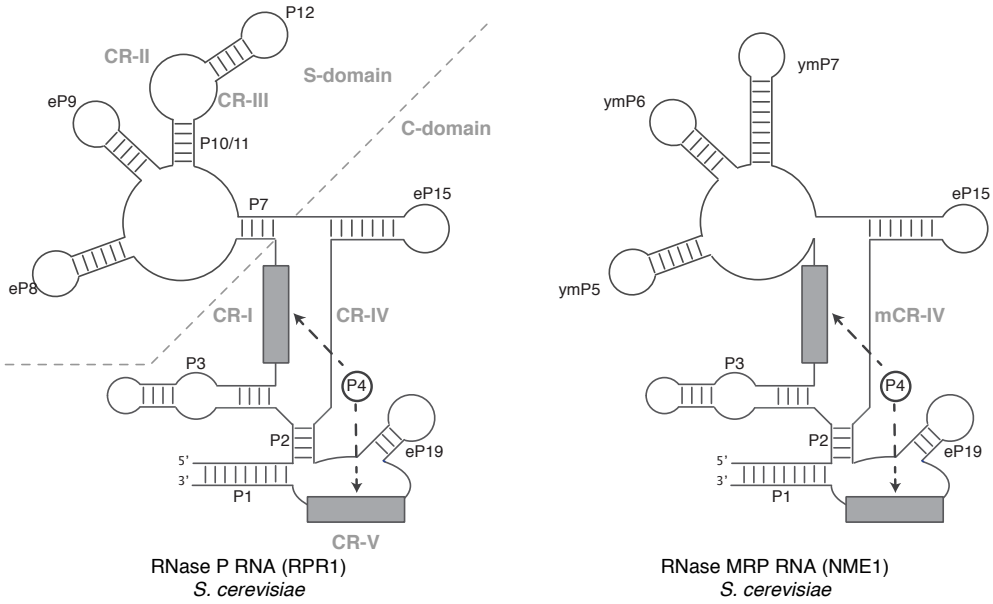

Figure 3.6 Secondary structure of the RNase P RNA and the RNase MRP RNA of *S. cerevisiae* . Adapted from (Ellis and Brown 2010)

B. Protein subunits of the nuclear RNase P.

The nuclear RNase P of *S. cerevisiae* is composed of nine proteins: Pop1, Pop3, Pop4, Pop5, Pop6, Pop7, Pop8, Rpp1 and Rpr2 (**Table 2**). Each of these proteins is essential for cell survival and were thus hypothesized as involved in essential maturation processes (Chamberlain et al. 1998). Only Pop1, Pop4, Pop6 and Pop7 proteins are directly bound to the RNase P RNA (Houser-Scott et al. 2002). Although the components of the nuclear RNase P of *S. cerevisiae* have been described almost two decades ago, the first *in vitro* reconstitution of the full holoenzyme was performed in 2018 (Perederina et al. 2018). This study showed that proteins form two sub-complexes: the Pop6/Pop7 complex and the Rpp1/Pop5/Pop8 complex. Moreover, only Pop1 and Pop6/Pop7 proteins are able to bind directly to RNA without the need of any other proteins. Concerning Pop4, it is actually not able to bind to RPR1 RNA without the presence of the Rpp1/Pop5/Pop8 sub-complex.

Table 2. RNA and protein moieties of the RNase P complex . Homologous protein genes are on the same line. Data taken from (Walker and Engelke 2006)

Yeast *S. cerevisiae*	**Mammalian** *H. sapiens*	**Archaeal** *P. horokoshii*	**Bacterial** *E. coli*
	RNA gene		
RPR1	*H1*	*PH rnpB*	*M1*
	Number of proteins		
9	10	4 or 5	1
Protein genes			
			C5 (rnpA)
POP1	*hPOP1*		
POP3	*RPP38*		
POP4	*RPP29*	*PH1771*	
POP5	*hPOP5*	*PH1581*	
POP6			
POP7	*RPP20*		
POP8			
RPP1	*RPP30*	*PH1877*	
RPR2	*RPP21*	*PH1601*	
	RPP40		
	RPP25		
	RPP14		

C. Role of the protein subunits.

Pop1 protein is required for the catalytic activity by contributing to the folding of RPR1 RNA and the substrate binding. The Pop6/Pop7 sub-complex has a purely structural role: the heterodimer binds to the strictly eukaryotic P3 domain and provides an interface to the Pop1 protein. The Pop8 protein allows the interaction between Pop1 and Rpp1/Pop5. The Rpp1/Pop5 sub-complex is required for catalysis and is bind to the vicinity of the catalytic core of the RPR1 RNA. In addition Rpp1/Pop5 allows Pop4 binding on RNA. Pop4 is bound to the substrate recognition domains but its presence is not essential to the precise cleavage of the pre-tRNA.

D. Functions of the eukaryotic nuclear RNase P.

As previously described, the nuclear RNase P is involved in cleavage of various substrates. Recent studies on human cells showed that RNase P directly interacts with various cellular mechanisms such as replication, DNA repair, and chromatin remodeling (Jarrous 2017). RNase P has been described as directly interacting with RNA transcription by RNA polymerase I and III (Reiner et al. 2008, Serruya et al. 2015). Concerning chromatin remodeling, protein subunits of the human RNase P (Rpr2, Pop4 and Pop1 homologous proteins) induce transcriptional silencing by repressing histone H3.3 recruitment (Newhart et al. 2016). These studies show evidence of the versatile function of RNase P, and its potential entanglement in all major intracellular mechanisms.

II.4.The diversity of eukaryotic RNase P.

The form of a ribonucleoprotein complex, which RNA subunit contains the catalytic activity, is universally conserved. However several exceptions have appeared through the evolution of eukaryotes, in the form of strictly proteinaceous RNase P, named PRORPs for **PRO**teinaceous **R**Nase **P** (Gobert et al. 2010). These proteinaceous RNase P were first discovered in human mitochondria (Holzmann et al. 2008) and plant organelles (Gobert et al. 2010). Furthermore, the RNase P activity of the PRORP has been described as substituting ribonucleoprotein complex in the nucleus of *Arabidopsis thaliana* and *Trypanosoma* (Gutmann et al. 2012, Taschner et al. 2012). The emergence of these protein forms seems logical according to the evolution of RNase P through the 3 domains of life. Indeed, as discussed above, the complexity of the RNA component has decreased in favor of an increasing protein complexity. Nevertheless, the conservation of RNase P as a large ribonucleoprotein complex in the vast majority of eukaryotes is currently enigmatic. The maintenance and biosynthesis of this large complex appears to be disproportionate to perform a simple hydrolysis of a phosphodiester bond; whereas a single protein as PRORP is able to do so. A study conducted in 2014 reinforced this reconsideration (Weber et al. 2014). This work reported a successful rescue of the RNase P RNA component deletion of *S. cerevisiae* by the PRORP3 protein from *A. thaliana*. This exchange showed no effect on the fitness of *S. cerevisiae* and surprisingly, it did not prevent the cleavage of alternative substrates by the nuclear RNase P. The substitution of RNase P even appeared to have a positive impact on yeast fitness under saline stress conditions.

Why RNase P is conserved as a ribonucleoprotein complex ?

This question does not have a clear answer yet. However some elements can be brought by changing the scope of the question as follows: What is the driving force behind the emergence of proteinaceous RNase P ? The major hypothesis involves the organelles and the high efficiency of the ribonucleoprotein version of the RNase P . Indeed RNase P in the form of a ribonucleoprotein complex are much more effective than PRORPs and recognize a very large number of substrates (Howard et al. 2013). The recognition of substrates alternative to tRNA seems to involve the protein subunits of the RNase P, via protein-protein interactions close to the substrate or by driving the subcellular localization of the complex. The majority of cellular RNAs are bound to proteins, which may be eventually discriminants to induce or prevent cleavage by RNase P. Compared to the nucleus, organelles express less RNA binding proteins and the high catalytic efficiency of the ribonucleoprotein version of RNase P could be a danger for organellar RNAs. An appropriate solution to this vulnerability would be an evolutive driving force for the emergence of PRORPs, but also for the maintenance of nuclear RNase P in ribonucleoprotein form. Indeed, unlike proteins, large nuclear RNAs are not imported into organelles (which could explains why mitochondria transcribe their own ribosomal RNA). Maintaining RNase P in the form of a large ribonucleoprotein complex could represent a way to contain the RNase P and its high catalytic efficiency outside the organelles. In conclusion, the maintenance of a ribonucleoprotein complex through the three domains of life and the emergence of PRORP in eukaryotes would be two sides of the same coin.

III. The nuclear RNase P of *C. glabrata*

The opportunistic pathogen *C. glabrata* has several genomic features, which have been described in the general introduction. Among these peculiarities, *C. glabrata* presents large insertions within several non-coding RNA genes, conferring unprecedented sizes to the transcripts. The first RNA identified with this characteristic in *C. glabrata* was RPR1, the RNA component of RNase P (Kachouri et al. 2005). This RNA acquired three new domains: **p7a** (230 nt.), **p7.1** (31nt.) and **eP8.1** (485 nt.) conferring to the RPR1 RNA a surprising length of 1149 nt. (**Figure 3.7**). Compared to *S. cerevisiae* and its 369 nt. RPR1 RNA, the counterpart of *C. glabrata* is almost three times larger. Initially considered as potentially intronic sequences, the experiments performed on RPR1 showed that these additional domains remain present in the active form of the RNase P holoenzyme (Kachouri et al. 2005). The RPR1 RNA of *C. glabrata* exhibits all the conserved domains found in eukaryotes and the additional domains are all inserted within the Specificity Domain (S-Domain). The presence of these insertions therefore does not seem to substitute essential domains of the RNase P RNA. Moreover, their location within the specificity domain seems to indicate a potential role in the recognition of RNA substrates.

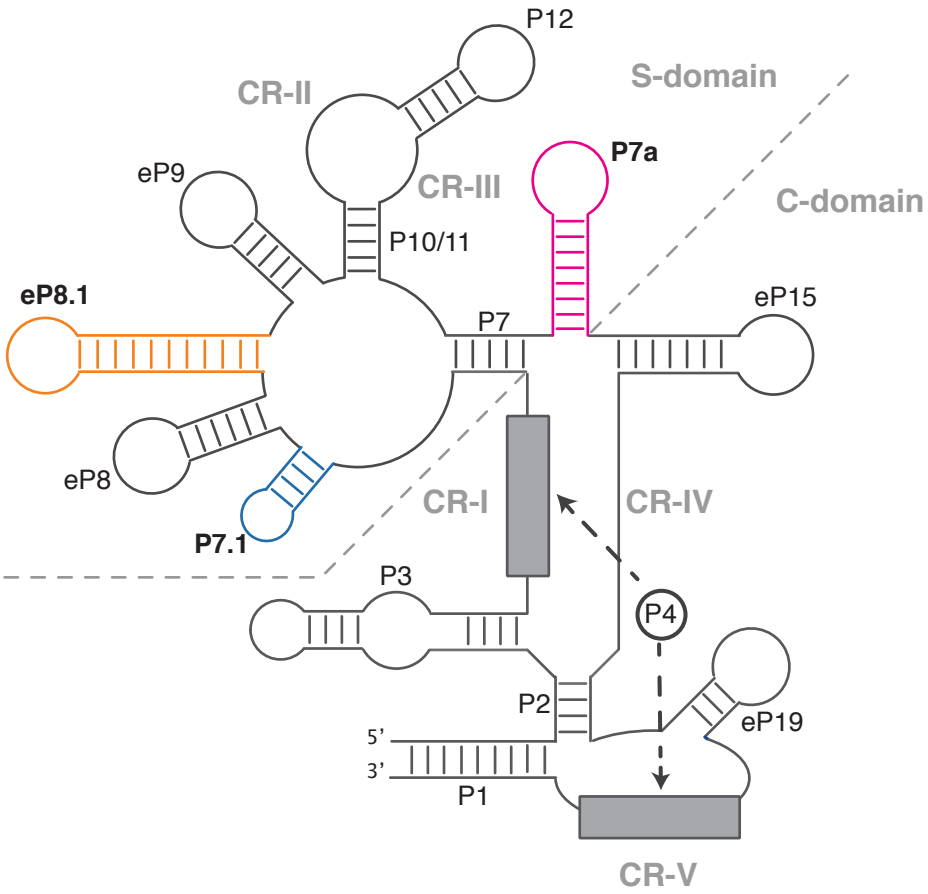

Figure 3.7 Schematic secondary structure of the RNase P RNA of *C. glabrata*. The structure is based on the secondary structure of *S. cerevisiae* and data from (Kachouri et al. 2005). Additional domains (P7a, P7.1 and eP8.1) are highlighted with specific color. For convenience this schematic view does not represent the accurate size of the additional domains. Domain **P7a** : 230 nt./ **P7.1** : 30 nt. / **eP8.1**: 485 nt.

Candida glabrata was the only known species with such a large RNase P RNA moiety, and the emergence of additional domains was hypothesized to confer somehow an advantage to the pathogenicity of the yeast. However, the genome sequencing of all species within the *Nakaseomyces* clade in 2013 (Gabaldón et al. 2013) has once again changed the situation: every *Nakaseomyces* species present a large RNA RPR1 (**Table 3**). The potential benefit of additional domains for the pathogenicity of *C. glabrata* was therefore excluded since even species described as "environmental" have the same particularity. Moreover, the largest RNase P RNA identified is found in *N. delphensis* (1368 nt.), species described as non-pathogenic.

The emergence of additional domains in the RNase P RNA moiety occurred therefore within the common ancestor of all *Nakaseomyces* species. This observation goes against the evolution of the eukaryotic RNase P which dynamics tend towards a shortening of the RNA moiety in favor of an increase complexity of the protein subunits. A simple question therefore emerges from this observation: Why such a large RNA moiety of RNase P has been conserved within the *Nakaseomyces*?

I have therefore attempted to provide some elements of an answer by focusing on the nuclear RNase P of *C. glabrata.*

Table 3. RNA moiety of the nuclear RNase P of *S.* cerevisiae and the *Nakaseomyces* clade.

Species	Gene ID	RPR1 length (nt)	Human pathogen
Saccharomyces cerevisiae	***RPR1***	**369**	**No**
Candida castelli	CACA0s21e03949r	607	No
Nakaseomyces bacillisporus	NABA0s02e00803r	936	No
Candida nivariensis	CANI0s02e03047r	1057	Yes
Candida bracarensis	CABR0s26e00792r	1068	Yes
Candida glabrata	**CAGL0L08044r**	**1149**	**Yes**
Nakaseomyces delphensis	NADE0s06e00726r	1368	No

Genes IDs and RNA length were extracted from the GRYC database (***gryc.inra.fr***)

Aim of the study

Throughout my thesis work, I have attempted to characterize the role of the additional domains within the RNA moiety of *C. glabrata* RNase P by answering two major questions:

1) Are these domains essential to the catalytic activity of the RNase P of *C. glabrata* ?

The additional domains could simply be implicated in structural stabilization, such as the peripheral helices of the bacterial RNase P. Their presence would therefore be essential for the catalytic activity of the RNase P holoenzyme. In contrast, if these domains provide only new non-canonic functions to the RNase P, the deletion of the additional domains will not prevent the holoenzyme from achieving the essential pre-tRNA maturation activity.

2) Does the presence of these domains modify the composition of the protein moiety RNase P ?

If the additional domains have a role in structural stabilization, then some of the holoenzyme proteins described in eukaryotes may be superfluous. On the other hand, if the emergence of additional domains is a constraint, the RNase P complex of *C. glabrata* could display new proteins which might either be involved in: i) maintaining a functional structure of the RNA despite the constraints induced by the presence of large insertions; or ii) providing new functions to the RNase P which could involve the holoenzyme in new non-canonical processes.

Results and discussion

A. Swapping the RNAse P RNA moiety *between C. glabrata* and *S. cerevisiae.*

In order to assess the biological role of a gene or a feature, the historical method used in molecular biology relies on a shutdown of its expression. However, in the case of the RPR1 RNA, its characterization/analysis faces two major challenges:

- i) *RPR1* is an essential gene, *C. glabrata* being strictly haploid the direct deletion is therefore not possible. In the case of an essential gene encoding a protein, indications concerning its function can be provided by inducing its overexpression under the dependence of a strong constitutive promoter. The non-coding RNA RPR1 is under the dependence of an RNA polymerase III promoter. Because the efficiencies of these promoters are not documented, the modulation of its expression is currently a difficult task.

- ii) The precise deletion of a domain is only possible if the structure of the given RNA has been resolved. Indeed, the function of an RNA depending strongly on its structure, its modification is thus risky and is very likely to generate artifactual results. The slightest deletion or unexpected insertion of a ribonucleotide could weaken, or create, a stem-loop and dramatically modify the overall structure and the RNA activity. The available structural data for *C. glabrata* RPR1 are solely based on structural alignments. Note that the positioning of additional domains may differ by several nucleotides depending on the alignment method and the investigator. These data therefore do not allow to have a precise structure of RPR1.

To identify the role of the additional domains of RNase P RNA, the method I used consisted of a swap of the *RPR1* gene of *C. glabrata* with a functional gene lacking additional domains. For this purpose we choose the *RPR1* gene of the model yeast *S. cerevisiae*. The swap was made in both ways: i) a substitution of the *S. cerevisiae* o *RPR1* with the version of *C. glabrata*; ii) a substitution of the *C. glabrata RPR1* gene with the version of *S. cerevisiae*.

The following hypotheses have been tested throughout these experiments :

1) If the additional domains of RPR1 are involved in the conservation of the essential activity of the RNase P, then the *RPR1* gene of *C. glabrata* could replace a version without additional domains. Consequently, *S. cerevisiae* would be able to survive by expressing the *RPR*1 gene of *C. glabrata.* The opposite case would suggest that *RPR1* needs *C. glabrata* specific cofactors to ensure the holoenzyme activity.

2) If the additional domains of RPR1 are only intended to provide additional functions to the RNase P, then *C. glabrata* is able to survive without their presence. Consequently, *C. glabrata* would be able to survive by expressing the *RPR*1 gene of *S. cerevisiae.* The opposite case would suggest that the additional domains of RPR1 are essential for a functional activity of the RNase P.

The anticipated results and their implication to the characterization of the role of the additional domains are summarized in the following table.

Table 4. Anticipated results and conclusion of the swap experiments of RPR1 gene in S. cerevisiae and C. glabrata.

Species	Survives with additional domains in RPR1	Survives without additional domains in RPR1	Potential role of the additional domains
C. glabrata	/	Yes	Provide new non-essential functions to the RNase P
S. cerevisiae	Yes	/	
C. glabrata	/	No	Involved in the conservation of essential activity
S. cerevisiae	Yes	/	
C. glabrata	/	Yes	Provide new non-essential functions to the RNase P **and** need specific cofactors to form an active structure
S. cerevisiae	No	/	
C. glabrata	/	No	Specific cofactors are essential for the RNase P activity
S. cerevisiae	No	/	

I. Replacement of the *S. cerevisiae RPR1* gene with the *RPR1* gene from *C. glabrata*.

The diploidy of *S. cerevisiae* is a major advantage for the investigation of essential processes. It enables notably the deletion of one of the two copies of an essential gene, without compromising the cell survival. For this purpose, the *RPR1* gene of *S. cerevisiae* (referred from now to *ScRPR1* gene) was disrupted within a diploid strain of *S. cerevisiae* named S2 strain (derived from the crossing of BY4719 and BY4738 strains), by the insertion of the genetic marker *URA3* within the *RPR1* locus. To determine if the *RPR1* gene of *S. cerevisiae* can possibly be exchanged by the *C. glabrata* version (referred from now to *CgRPR1*), complementation tests were performed (**Figure 4.1**). The S2 strain of *S. cerevisiae* has only one copy of the Sc*RPR1* gene (*RPR1*/*rpr1*Δ::*URA3*), it was first transformed with the pRS-CgRPR1 plasmid containing the *RPR1* gene of *C. glabrata*, under its own promoter. In order to select the transformed cells, the plasmid contains the genetic marker *TRP1*, which confers to the S2 strain a prototrophy to tryptophan. The transformed strain S2 thus expresses both versions of the *RPR1* gene. To generate cells containing only one copy of the *RPR1* gene, i.e. to form haploid cells, the meiosis of the S2 diploid strain was triggered on AcK medium (**Figure 4.1**). This meiosis event, called sporulation, led to the formation of 4 haploid cells contained in an ascus. Therefore, for each haploid cell two different cases can occur: i) the haploid cell includes the wild-type endogenous *ScRPR1* gene or ii) the haploid cell includes the *ScRPR1* gene disrupted by the *URA3* marker. Concerning the segregation of the plasmid pRS-CgRPR1, it comprises the centromeric sequence CEN-ARS enabling in most cases, the vector to be present in all haploid cells generated after meiosis. The tetrads obtained were dissected by micro-manipulation after digestion of the ascus wall by zymolyase and then placed on YPD medium.

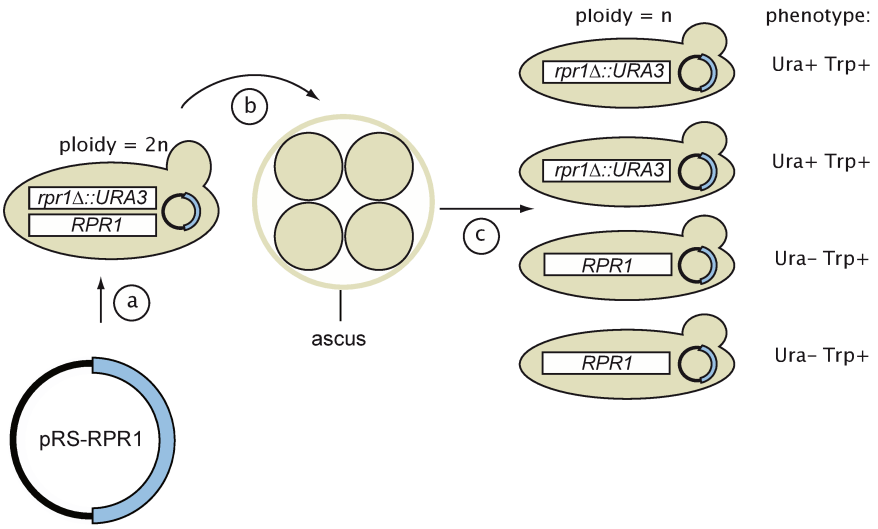

Figure 4.1 Functional complementation of the *RPR1* gene in *S. cerevisiae*.

Steps of the functional complementation of *RPR1* gene are: a) transformation of *S. cerevisiae* diploid strain containing 1 out 2 *RPR1* copies; b) induction of meiosis to obtain haploid cells; c) analysis of the viability of the haploid cells obtained. The viable cells are selected for their phenotype. Ura+ : cell containing the *RPR1* allele deleted. Ura- : cell containing the wild-type *RPR1*.

Concerning the complementation by the pRS-cgRPR1 plasmid, no tetrads resulting from a diploid transformed by the vector pRS-cgRPR1 were totally viable and only 2 out 4 spores within each tetrad were viable on rich medium (**Figure 4.2**). To determine their genotype, all the 256 spores obtained were streaked onto synthetic medium without uracil (SC-U). None of the 256 spores growing on YPD rich medium was able to grow on SC-U medium. These 256 spores revealed a Ura- phenotype, and consequently expressed the *scRPR1* wild-type gene.

In conclusion, the pRS-cgRPR1 construction was not sufficient to complement the loss of the endogenous *S. cerevisiae* locus.

Within the plasmid pRS-cgRPR1, the *CgRPR1* gene was placed under the dependence of its own promoter. In order to avoid the possible bias associated with the use of an exogenous promoter in *S. cerevisiae*, a new plasmid called pRS-promSc-CgRPR1 was designed. As indicated by its name, the *CgRPR1* gene has been placed downstream the endogenous *RPR1* promoter of *S. cerevisiae*. This plasmid was transformed within the S2 strain and the same complementation experiments described above were performed. Again, all tetrads obtained after meiosis comprised only 2 viable haploid cells out of 4. The phenotype of 256 viable spores was tested by transplanting on SC-U medium and as previously all spores showed a Ura- phenotype. Like the plasmid pRS-RPR1, these data indicate that the pRS-promSc-CgRPR1 plasmid is not likely to rescue the loss of the endogenous locus in *S. cerevisiae*.

The *RPR1* gene of *Candida glabrata* is then not able to substitute functionally its ortholog in *S. cerevisiae*.

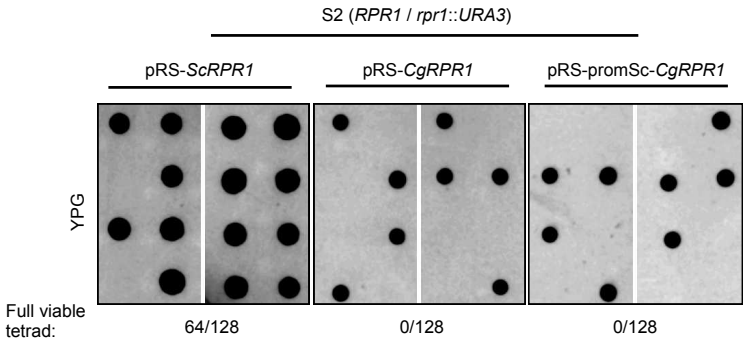

Figure 4.2. Viability test of meiosis products. After sporulation of the S2 strain transformed by vectors pRS-scRPR1, pRS-CgRPR1 and pRS-promSc-CgRPR1, the tetrads were dissected and separated on YPD. Through meiosis, the transformed vectors passed to the progeny by Mendelian segregation. For each S2 strain transformed, 128 tetrads were dissected and observed, and the number of complete tetrads was assessed. The presence of full viable tetrad indicates that the construction is able to complement the deletion of the endogenous *RPR1* locus. The pRS-CgRPR1 and pRS-promSc-CgRPR1 were not able complement the deletion of the *ScRPR1* gene.

II. Replacement of the *C. glabrata RPR1* gene with the *RPR1* gene from *S. cerevisiae*.

In order to confirm the inability of exchanging the *RPR1* gene, the same experiment was performed in the genome of *C. glabrata*. As *C. glabrata* does not present a diploid form, the exchange of the essential gene *RPR1* was carried out in two distinct stages, namely: i) the integration of the *ScRPR1* gene into the genome of *C. glabrata* followed by ii) the disruption of the endogenous locus by a selection marker.

a. Integration of *ScRPR1* into the genome of *C. glabrata*.

The integration site selected for the *ScRPR1* gene is located in a gene desert within the B chromosome (the detection of this site and its advantages are discussed in Chapter 2). The integration cassette contains the *ScRPR1* gene under the endogenous promoter of *CgRPR1* to avoid a possible transcription issue. In addition, the terminating sequence of the *S. cerevisiae* $tRNA^{Tyr2}$ terminator (tTY2), described as a functional terminator in *C. glabrata*, has been placed downstream of the *ScRPR1* sequence. To ensure the integration of the cassette by homologous recombination into the B chromosome, 210 bp homology arms were fused upstream and downstream of the *ScRPR1* gene (**Figure 4.3.A)**.

The integrative cassette does not have a marker to positively select a recombination event. It was therefore essential to increase the efficacy of homologous recombination, known to be low in *C. glabrata*. The choice was therefore to use the CRISPR-Cas9 tool developed in the laboratory (Enkler et al 2016).

The CRISPR system consists of two components: a single guide RNA (sgRNA) and a CRISPR-associated endonuclease (Cas protein). The sgRNA is a synthetic RNA composed of : a scaffold for the binding to Cas protein and a 20 nucleotide spacer that defines the genomic site to be cleave. The appearance of a double-strand break induces the recruitment of DNA repair system components that will significantly increase the probability of an homologous recombination (Doudna and Charpentier 2014).

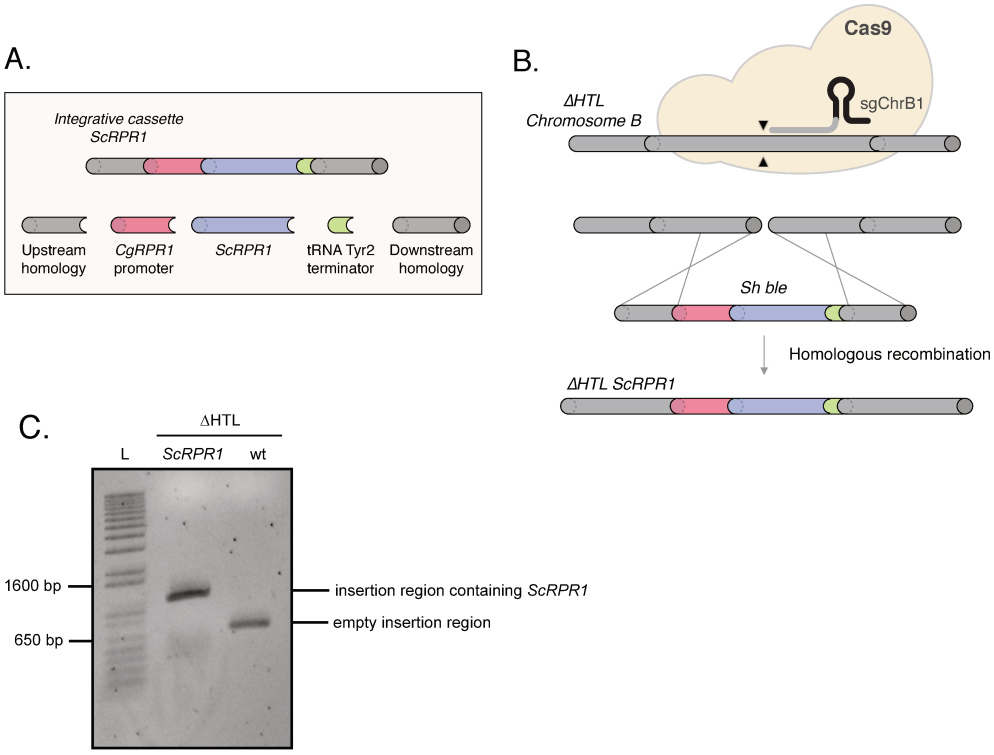

Figure 4.3. Integration of the *ScRPR1* gene in the genome of *C. glabrata*.

(**A**) The integrative cassette ScRPR1 is composed of 210 bp homology arms, the *RPR1* gene of *S. cerevisiae* under the promoter of *C. glabrata* and the $tRNA^{Tyr2}$ terminator. (**B**) The integrative cassette, plasmids enabling the expression of the *Cas9* protein and the single guide RNA sgCHrB1 were co-transformed in the ΔHTL strain. The integrative cassette enabling the expression of the *ScRPR1* gene was inserted into the chromosome B of *C. glabrata* by homologous recombination, enhanced by a *Cas9* induced double-strand break at the insertion site. (**C**) The sequence corresponding to the insertion site was amplified by PCR from colonies having grown on SC-TL medium. The expected size is 1178 bp for an insertion event, otherwise the expected size is 608 bp.

The guide RNA used for this purpose, sgChrB1, has been designed using the online CASTING tools (http://charn-ibmc.u-strasbg.fr:8080/casting.html) and cloned into the pRS315-sgChrB1 plasmid, containing the selectable marker *LEU2* . The Cas9 protein was provided by the plasmid pRS314-Cas9 containing the selectable marker *TRP1*.

The integration of *ScRPR1* into the genome of *C. glabrata* was performed in the triple auxotrophic strain ΔHTL by co-transformation of the plasmid pRS315-sgChrB1, pRS314-Cas9 and the *ScRPR1* cassette (**Figure 4.3.B)**. After transformation the cells were streaked on SC-TL selective medium. The integration of the *ScRPR1* cassette has been assayed by PCR amplification directly on the colonies. A total of 48 colonies were verified by PCR and only one showed successful integration (**Figure 4.3.C**). The PCR fragment obtained was sequenced and confirmed the integration of the *ScRPR1* cassette into the chromosome B of *C. glabrata*.

The expression of the *ScRPR1* gene was checked by RT-PCR on the total RNA extracted from the obtained strain ΔHTL *ScRPR1* (**Figure 4.4**). As a control, the RT-PCR experiment was performed on the endogenous locus *CgRPR1*. These experiments validated the co-expression of the two versions of the RPR1 RNA within *C. glabrata*.

b. Disruption of the endogenous locus.

The second step was the disruption of the endogenous locus *cgRPR1* by an integrative cassette. For this purpose the CRISPR-Cas9 tools were used again. As previously, the single guide RNA was designed using the CASTING tool, synthesized and cloned within the plasmid pRS315-sgRPR1. The *HIS3* gene was the first choice as a *CgRPR1* disruption cassette. Indeed the integration of the *HIS3* marker in the autotrophic strain ΔHTL can easily be selected by streaking on a medium without histidine. Like all cassettes designed to be integrated into a genome by homologous recombination, homologous sequences in the vicinity of the insertion site must be fused upstream and downstream of the marker sequence. Unfortunately, the homologous sequences at the insertion site in the *RPR1* gene could never be fused to the *HIS3* cassette.

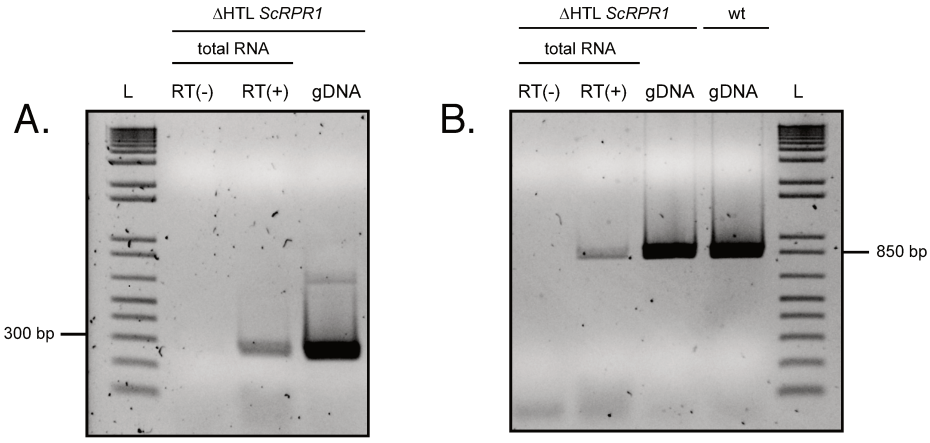

Figure 4.4. Co-expression of two *RPR1* gene in *C. glabrata*. Sc*RPR1* and *CgRPR1* gene expression were tested by RT-PCR. The sequence of interest was amplified at cDNA level by PCR. Positive control : same PCR experiments were performed on genomic DNA (gDNA lane). Negative control : same PCR experiments were performed on RNA sample that had not undergone retro-transcription, confirming that the amplification were not due to DNA contamination. (**A**) Amplification of the *ScRPR1* cDNA confirmed the expression *ScRPR1* in the ΔHTL *ScRPR1* strain. (B) Amplification of the *CgRPR1* cDNA confirmed the feasibility of the method.

This problem seemed to be due to the primers used for the fusion of the multiple fragments composing the cassette. These primers which size varies from 30 bp to 50 bp contains overhang sequences at the 5' ends, allowing the fusion with adjacent fragment. The primers probably formed a strong secondary structure preventing the optimal conduct of the PCR. To overcome this problem, the disruption cassette has been changed to the *Sh Ble* marker, a gene conferring resistance to the Zeocin antibiotic (Alderton et al. 2006). The *Sh ble* cassette was amplified from pRS315-Zeocin plasmid and flanked by homology sequences of 360 nt. upstream and downstream the insertion site (**Figure 4.5**). The disruption of the locus *CgRPR1* was performed in the strain ΔHTL *ScRPR1* previously generated, by co-transformation with the pRS315-sgRPR1 plasmid, pRS314-Cas9 plasmid and *Sh ble* cassette.

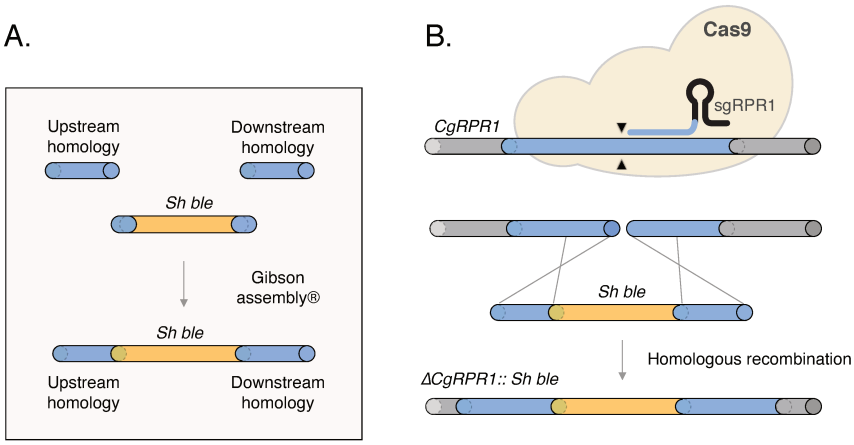

Figure 4.5. Inactivation of the *CgRPR1* gene. (**A**) The DNA fragments constituting the disruption cassette of *CgRPR1* were fused by the "Gibson Assembly®" method. (**B**) The disruption cassette, the plasmids enabling the expression of the *Cas9* protein and the single guide RNA sgRPR1 were co-transformed in the ΔHTL-*ScRPR1* strain. The disruption cassette of cgRPR1 was inserted into the genome by homologous recombination, enhanced by a *Cas9* induced double-strand break at the insertion site.

After transformation, cells were streaked on SC-TL selective medium supplemented by Zeocin. The correct integration of the *Sh Ble* marker has been assayed by PCR amplification directly on the colonies, using the same primers as for the amplification of the complete *Sh Ble* cassette. A total of 112 colonies were tested by PCR (**Figure 4.6**) and only two different cases occurred: i) a profile similar to the wild-type, meaning that integration did not happen (Clone 2 **Figure 4.6**); and ii) a surprising profile, combining a band size assuming insertion of the *Sh Ble* cassette « simultaneously » with a profile equivalent to the wild-type (Clone 1 and 3 **Figure 4.6**). The DNA fragments obtained were purified and sequenced. The sequence analysis showed that the fragments present the complete sequence, namely the *Sh Ble* gene flanked by homology arms. It was then hypothesized that the presence of these two profiles was due to contamination. According to this hypothesis, colonies tested by PCR might not have been monoclonal but resulted from a mix of wild-type cells, taking advantage of the zeocin resistance of the intended mutant. In order to test this hypothesis, the cells were successively re-streaked on YPD medium supplemented by increasing concentrations of zeocin. Following these re-streaking, all colonies were tested again by PCR. The results were exactly the same as previously, namely the presence of the two profiles, disruption and non-disruption of the *CgRPR1* gene, within the same cell.

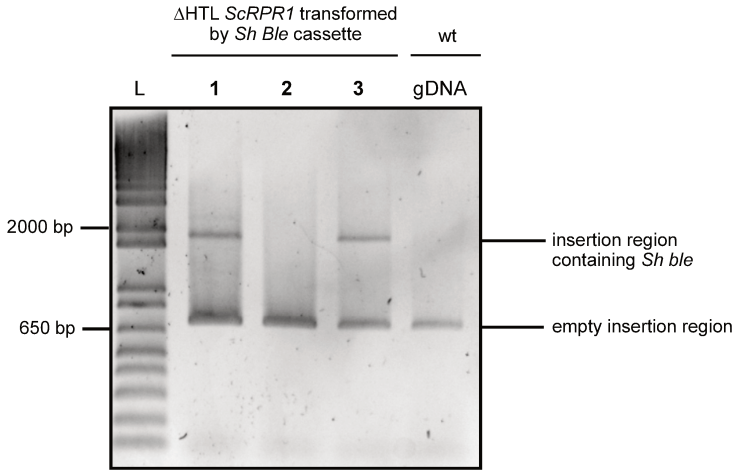

Figure 4.6. Random integration of the *Sh Ble* cassette in the genome of *C. glabrata*. A representative sample of 3 out of 112 colonies assayed is presented here. The sequence corresponding to the insertion site was amplified by PCR from colonies having grown on SC-TL medium + Zeocin™, after the transformation of the *Sh Ble* cassette. The same primers as for the amplification of the complete *Sh Ble* cassette were used for this experiment. The expected size was 1888 bp for an insertion event, otherwise the expected size was 717 bp. The presence of two profiles in the same lane indicates a random integration of the *Sh Ble* cassette.

Note that the primers used for these PCR experiments were the same as previously, namely those used to amplify the insertion cassette before transformation. Based on this observation, a second hypothesis was put forward to try to explain the combination of insertion profiles: the disruption cassette may have been randomly integrated into a locus that is not the location of *CgRPR1*. This hypothesis made sense since the primers used to verify the insertion of the *Sh ble* disruption cassette did not take into account the genomic context outside the homology arms.

New PCR experiments were therefore performed on the 112 original colonies, using primers binding outside the homology arms of the disruption cassette. For all the tested clones, profiles obtained were similar to the wild-type meaning that the integration of the *Sh Ble* cassette did not occur. These results therefore confirm that the disruption cassette was not specifically integrated in the *CgRPR1* locus but in an other location in the genome.

We suggested that if this complete 1888 nt. cassette was able to be inserted non-specifically, the homology arms flanking the ends of the *Sh ble* gene might be involved. The specificity of the homology arms of the disruption cassette within the genome of *C. glabrata* was tested using the BLAST-N tool available on the GRYC database (http://gryc.inra.fr/index.php?page=blast). Surprisingly no homology other than the *CgRPR1* locus was detected, even with the most permissive BLAST-N parameters.

How the cassette can be randomly inserted into the genome of *C. glabrata* ?

The use of the CRISPR-Cas9 tool has been suspected as being potentially involved in this event. Indeed, the major limitation of the CRISPR-Cas9 system lies in its propensity to induce non-specific double-strand breaks in the genome, called off-targets (Fu, Foden et al. 2013). An off-target could therefore have possibly allowed the integration of the *Sh Ble* disruption cassette.

In order to test this hypothesis, a new *CgRPR1* disruption experiment was performed without the use of the CRISPR-Cas9 tool. The strain ΔHTL *ScRPR1* was therefore transformed only with the *Sh Ble* disruption cassette and the transformants were spread on YPD zeocin supplemented medium. The insertion of the cassette was again tested by PCR using the original primers annealed to the ends of the homology arms (**Figure 4.6**). The results of this analysis were the same as previously : both profiles were simultaneously present. The *Sh Ble* disruption cassette has been non-specifically integrated. According to these results, *Cas9* is not involved in the random integration. The origin of this phenomenon has not yet been elucidated. Nevertheless, taken together, these experiments never showed a disruption of the *CgRPR1* gene. Analysis of more than 200 clones with no disruption event therefore suggests that the *ScRPR1* gene is not able to substitute *C. glabrata* version of RPR1.

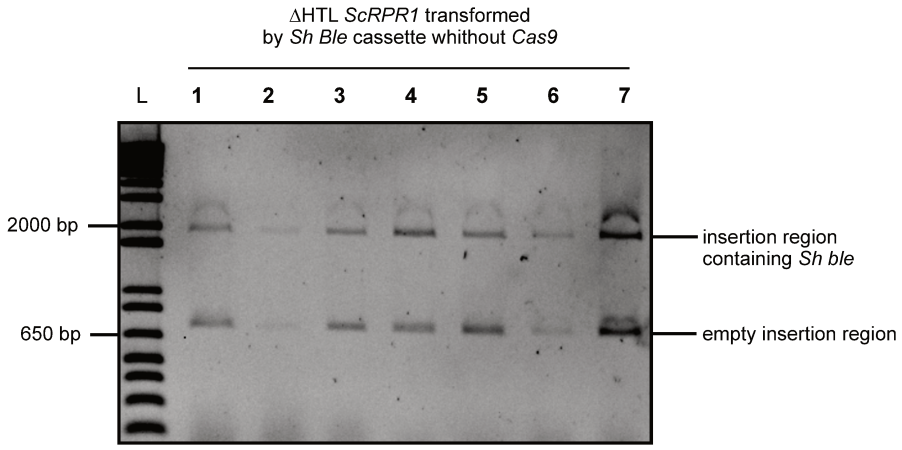

Figure 4.6. Random integration of the *Sh Ble* occurred without the use of the CRIPSR-Cas9 system. A representative sample of 5 out of 80 colonies assayed is presented here. The sequence corresponding to the insertion site was amplified by PCR from colonies having grown on YPD medium + Zeocin. The two profiles in the same lane indicates a random integration of the *Sh Ble* cassette. The presence of two profiles in the same lane indicates a random integration of the *Sh Ble* cassette.

Why can the *ScRPR1* and *CgRPR1* genes not be substituted ?

Previous experiments have not resulted in any *RPR1* gene substitution event between *C. glabrata* and *S. cerevisiae*. In conclusion, the *CgRPR1* and *ScRPR1* genes are described as non-interchangeable. These results do not provide significant clues about the role of additional domains of RPR1. Nevertheless, these experiments have validated the hypothesis that specific *C. glabrata* cofactors are essential for the RNase P activity. These results then support two assumptions:

Postulate **A**) CgRPR1 RNA needs cofactors only found in *C. glabrata.*

Postulate **B**) CgRPR1 RNA cofactors are not able to recognize ScRPR1.

B. Characterization of the protein moiety of the RNase P complex of *C. glabrata*

As described previously in the introduction, the nuclear RNase P is a ribonucleoprotein complex composed of an RNA associated with nine proteins. Each protein of the complex is essential to the activity of the RNAse P (Chamberlain et al. 1998). Therefore the hypothetical cofactors of CgRPR1 could refer to its protein partners. To understand postulates A and B, two situations are conceivable:

1) The nine RNase P proteins co-evolved with *C. glabrata* RPR1 RNA and are strictly specific to it. This hypothesis was tested *in silico* by an analysis of the protein sequences.

2) *C. glabrata* RPR1 RNA is associated with one or more additional proteins, i.e. other than the nine proteins described in yeasts. In *S. cerevisiae* these additional proteins may either be absent from the genome or be present but unable to bind to the RNase P complex. This hypothesis was tested experimentally by purification and characterization of the *C. glabrata* RNase P complex.

I. *In silico* analysis of the predicted protein moiety of *C. glabrata* RNase P.

The previous data suggested that the RNase P protein subunits are involved in the fact that *RPR1* genes are not exchangeable between *C. glabrata* and *S. cerevisiae*. Indeed, the proteins would be strictly species-specific and be unable to recognize an heterologous RNA moiety. RNase P protein subunit therefore probably co-evolved with the respective RPR1 RNAs. Moreover, although *C. glabrata* is considered to be phylogenetically close to *S. cerevisiae*, the evolutionary distance between these two species is as large as that between humans and zebrafish (Gabaldón et al. 2013) . According to this divergence, the *RPR1* genes exchange between *S. cerevisiae* and *C. glabrata* would seem to be a naive approach. However, even if the nucleotide sequence of *S. cerevisiae* and *C. glabrata RPR1* genes of seems globally divergent, the secondary structure adopted by these RNAs are very similar, and a strong sequence identity is conserved within the conserved eukaryotic domains (Kachouri et al. 2005). These sequence discrepancies therefore do not seem to be an obstacle to protein recognition of RNA. Another fundamental fact seems to go in this direction: the sharing of eight out of nine9 RNase P proteins with the RNase MRP. Indeed, even if the between the RNA moieties of RNase P and MRP within the same cell, yet the protein subunits of RNase P are capable of binding to the RNA of RNase MRP. Despite sequence divergence, secondary and tertiary structures are conserved between RNase P and MRP (Li et al. 2002, Hipp et al. 2011). Such conservation shows that the binding of protein to the RNA moiety is tolerant to nucleotide sequence divergence. Note that the RNA moiety of *C. glabrata* RNase MRP (*NME1* RNA) has a size equivalent to its counterpart in *S. cerevisiae* (respectively 337 nt. vs 340 nt.). Nevertheless the emergence of additional domains within RPR1 could be a driving force for several proteins to diverge strongly. The hypothesis was tested by searching major differences between *S. cerevisiae* and *C. glabrata* RNase P proteins.

Similarity between *S. cerevisiae* and *C. glabrata* RNase P protein moieties.

Each RNase P protein sequence from *S. cerevisiae* and their counterparts in *C. glabrata* were extracted from the GRYC database. The sequences were first retrieved via their annotation, using the protein names in *S. cerevisiae* as query. First point, all RNase P proteins described in *S. cerevisiae* have an homolog in *C. glabrata.* It indicates that the sequences divergence is restricted, since it did not prevent their annotations. The protein sequences obtained were then aligned pairwise via the T-COFFEE software (Notredame et al. 2000). Sequence identity and their similarity ,via the BLOSUM62 substitution matrix, were calculated and reported in the following table (**Table 5**).

Table 5. ***RNase P protein sequences identity and similarity between C. glabrata and S. cerevisiae.***

		Compared to S. cerevisiae	
Protein	***Gene ID***	***Identity (%)***	***Similarity (%)***
Pop1*	CAGL0K05049g	48,5	67,7
Pop3	CAGL0I03014g	21,3	41,1
Pop4*	CAGL0I09152g	51,6	72,1
Pop5	CAGL0M11836g	46,4	69,4
Pop6*	CAGL0K03993g	35,8	57
Pop7*	CAGL0M06039g	45,4	65,2
Pop8	CAGL0J04334g	45,1	69,2
Rpp1	CAGL0D03740g	64	77,4
Rpr2	CAGL0K08712g	24,6	38

In *S. cerevisiae,* protein marked with an asterisk (*) are described as interacting directly with the RNase P RNA moiety (Perederina et al. 2018). Pairwise alignments were performed with T-COFFEE software (Notredame et al. 2000).

In comparison with *S. cerevisiae*, all proteins except Rpp1 and Pop4 have a protein sequence identity below 50%. This value is lower than the mean identity value (54%) between all orthologous protein-coding genes of *S. cerevisiae* and *C. glabrata* (Gabaldón et al. 2013). It is important to note that while sequence identity is a substantial consideration, sequence similarity is a more significant signal of conservation of protein function (Pearson 2013). The most divergent proteins are Pop3 (41.1% of similarity with *S. cerevisiae*) and Rpr2 (38% of similarity with *S. cerevisiae*). Very little information is available in the literature about the role of these proteins and their interaction within the RNase P holoenzyme. However the case of Rpr2 is interesting since it is the only protein not shared with the MRP RNase. The divergence of RPR2 may therefore be due to the fact that RNA moiety of RNase MRP does not force its conservation, unlike other proteins. Proteins showing the highest percentage similarity with *S. cerevisiae* are Rpp1 (77.41%) and Pop4 (72.1%). The Rpp1 protein has been described in *S. cerevisiae* as located in the vicinity of the catalytic center of RPR1 (Perederina et al. 2018). This domain is highly conserved between *S. cerevisiae* and *C. glabrata* (Kachouri et al. 2005), strong similarity between the orthologs is therefore consistent. This theory also applies to the Pop4 protein since it interacts directly with the conserved CR-III region of RPR1 RNA, which secondary structure and sequence is strictly conserved between *C. glabrata* and *S. cerevisiae* (Kachouri et al. 2005). Concerning other proteins in direct interaction with the RNA moiety, namely Pop1, Pop6 and Pop7, their similarity with their orthologs in *S. cerevisiae* does not indicate a convincing signal of co-evolution with additional domains. All these proteins showed a sequence similarity greater than 65%, except for Pop6 which has 57% similarity with its ortholog in *S. cerevisiae*.

Similarity between *Nakaseomyces* species and *C. glabrata* RNase P protein moieties.

To confirm the absence of co-evolution signals of RNase P protein partners in *C. glabrata*, the protein sequences were aligned in pairwise with all orthologs present in the *Nakaseomyces* clade. All RNase P of this clade have a large RNA moiety ranging from 607 nt. for *N. castelli*, to 1368 nt. for *N. delphensis*. The protein sequences were retrieved from the GRYC database, using the annotation of the protein as query, otherwise via the BLASTp program. Pairwise alignments were performed via the T-COFFEE program and their similarity with the *C. glabrata* sequences were calculated via the BLOSUM62 substitution matrix (**Figure 4.7**). For all proteins, the percentage of similarity with *C. glabrata* was the highest in the species from the « *glabrata* group », which contains *C. bracariensis*, *C. nivariensis* and *N. delphensis (Gabaldón et al. 2013)*. These species also have the largest RNase P RNA moiety of the clade: 1149 nt. in *C. glabrata*, 1057 nt. in *C. nivariensis*, 1068 nt. in *C. bracariensis* and 1268 nt. in *N. delphensis*. Such as the previous comparison with *S. cerevisiae*, proteins with the lowest percentage of similarity in this group are Pop3 and Rpr2. Overall, the strong similarity between the protein subunits within the « *glabrata* group » is consistent with the phylogenetic proximity of these species (Gabaldón et al. 2013). Analysis of the RNase P protein sequences from other species of the *Nakaseomyces* clade also points in this direction, since the percentages of similaritiy between *N. bacilisporus*, *C. castelli* and *C. glabrata* are lower than within the « *glabrata* group ». Despite the sharing of a characteristic large RNA moiety of RNase P within *Nakaseomyces*, mismatch between the sequences of the RNase P protein subunits are eventually consistent with their phylogenetic distance. These data confirmed the absence of a strong protein conservation signal in relation to the emergence of additional domains. Note that these results are only preliminary, since the search for a co-evolution signal between the emergence of additional domains and protein subunits sequences can be performed more finely.

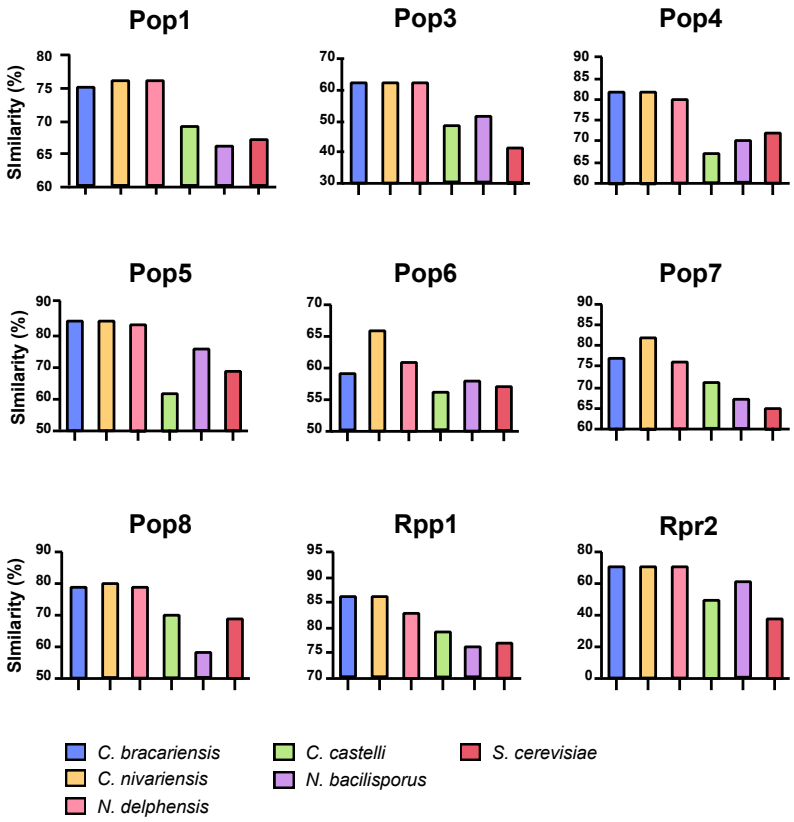

Figure 4.7. RNase P protein sequences similarity between *C. glabrata* and *Nakaseomyces* species. Pairwise alignments of each RNase P protein between each *Nakaseomyces* species and *C. glabrata* were performed (40 pairwise alignments) with T-COFFEE software (Notredame et al. 2000). Similarity rates were extracted from the 40 pairwise alignments and represented as bar charts. In comparison, data from pairwise alignments between *S. cerevisiae* and *C. glabrata* were also reported on each graph.

The most accurate method would be to analyze the precise interactions between the subunits of *S. cerevisiae* RNase P holoenzyme and to compare them with *C. glabrata* and other *Nakaseomyces* species. Unfortunately, no complete 3D structure of the nuclear fungal RNase P complex has been described yet. Nevertheless, structural data from isolated domains are available as is the case for the P3 domain associated with the Pop6/Pop7 protein of *S. cerevisiae* RNase P (Perederina et al. 2010).

Analysis of the P3 domain of RPR1 in *C. glabrata*.

As mentioned in the introduction, the P3 domain of RPR1 is a strictly eukaryotic domain. This domain has a helix-loop-helix structure and consists of a platform on which the proteins Pop1, Pop6 and Pop7 are bound. The P3 domain has a major stabilizing role for RNA, replacing specific peripheral domains of the bacterial RNase P RNA moiety. The P3 domain appears to be a key element in the shortening of RNA moiety eukaryotic in favor to an increased protein complexity (Perederina et al. 2010). The analysis of this domain is interesting for the case of the RNase P of *C. glabrata,* since its RNA moiety goes in the opposite direction of the eukaryotic RNase P evolution.

Secondary structure of the RPR1 P3 domain in *C. glabrata*.

The P3 domain sequence was extracted from the structural alignment data of the original publication describing the additional domains of *C. glabrata* RPR1 RNA (Kachouri et al. 2005). The secondary structure was predicted using RNAfold software (Lorenz et al. 2011). The predicted structure was compared to the *S. cerevisiae* counterpart, and the strictly conserved nucleotides were mapped on both structures (**Figure 4.8**). Comparison of the two P3 domains secondary structures showed a global similarity but presented nevertheless several notable differences. The proximal helix (first stem on the right) of the P3 domain of *C. glabrata* contains the eukaryotic consensus sequence of the P3 domain (5’ ACCUG-CAGGU 3’, highlighted in pink) but this helix is longer than the *S. cerevisiae* helix. Consequently, P3 domain inner loop of *C. glabrata* is smaller than *S. cerevisiae* (12 nt. vs 19 nt.).

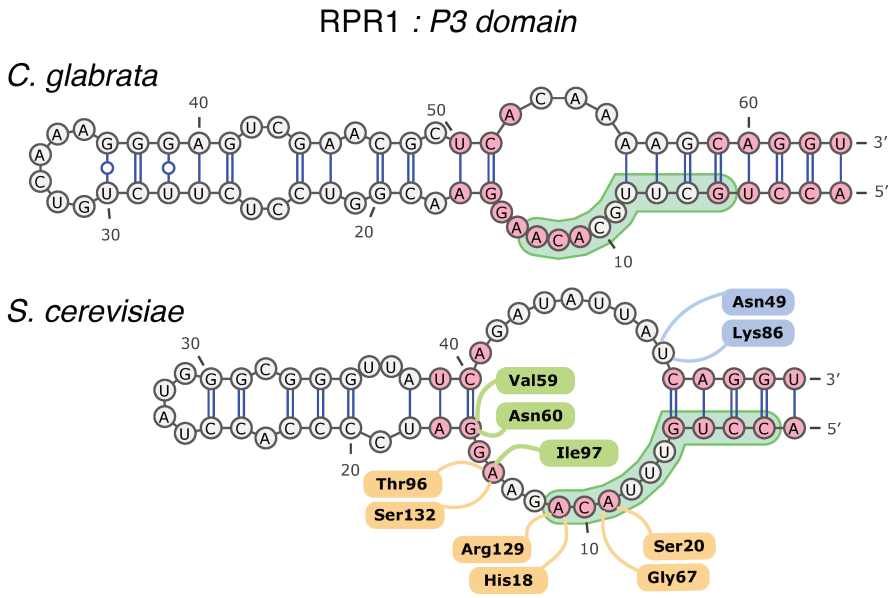

Figure 4.8 Comparison of the RPR1 P3 domain of *S. cerevisiae* and *C. glabrata*. Secondary structure of the RPR1 P3 domain of *S. cerevisiae* and *C. glabrata*. The structure of *S. cerevisiae* was extracted from (Perederina et al. 2010). The secondary structure of the P3 domain of *C. glabrata* was predicted using RNAfold algorithm (Lorenz et al. 2011).

Pink nucleotide : conserved RNA nucleotide in *S. cerevisiae* and *C. glabrata* P3 domain. Light green shape : nucleotides protected by Pop6/Pop7 binding (Perederina et al. 2007). Amino acid boxes: residue of Pop6 and Pop7 in direct interaction with the nucleotide of P3 RPR1 domain in *S. cerevisiae* (Green: Pop6 interaction with a conserved nucleotide in *C. glabrata;* Yellow : Pop7 interaction with a conserved nucleotide in *C. glabrata*, Blue: Pop7 interaction with a non-conserved nucleotide in C. *glabrata*). Data extracted from the crystal structure of the P3 domain - Pop6/7 complex (Perederina et al. 2010).

P3 domain interaction with Pop6/Pop7 proteins.

Direct interactions between Pop6/Pop7 amino acids and the *S. cerevisiae P3* domain nucleotides were mapped on the secondary structures (**Figure 4.8**). In comparison to *S. cerevisiae*, all nucleotides interacting with Pop6/Pop7 proteins have a conserved nature and position in *C. glabrata*. One exception is the nucleotide U49 from the P3 domain of *S. cerevisiae*, interacting with the Pop7 residues Asn49 and Lys86, which is absent from the P3 domain inner loop of *C. glabrata*. The protein sequence alignments of Pop6 and Pop7 between *S. cerevisiae* and *C. glabrata* were analyzed with respect to these interactions (**Figure 4.9**). Concerning Pop6, the residues Val59 and Ile97 interacting directly with the P3 domain of *S. cerevisiae* are strictly identical in *C. glabrata* and the residue Asn60, a polar amino acid, is substituted in *C. glabrata* by a serine, also a polar amino acid. Concerning Pop7, residues interacting directly with the P3 domain nucleotides are conserved in *C. glabrata*, except for the His18 residue of *S. cerevisiae* which has been substituted by a leucine. These data confirmed that the *C. glabrata* P3 domain and its interaction with Pop6/Pop7 proteins is similar to *S. cerevisiae*. Such similarity validates the hypothesis that the additional domains of RPR1 are not maintained to compensate the absence or the non-functionality of the P3 domain. Therefore, the roles of the additional domains of RPR1 are probably not strictly structural. To conclude, these analyses support the hypothesis that the additional RNA RPR1 domains of *C. glabrata* induce the presence of new protein partners within the RNase P holoenzyme.

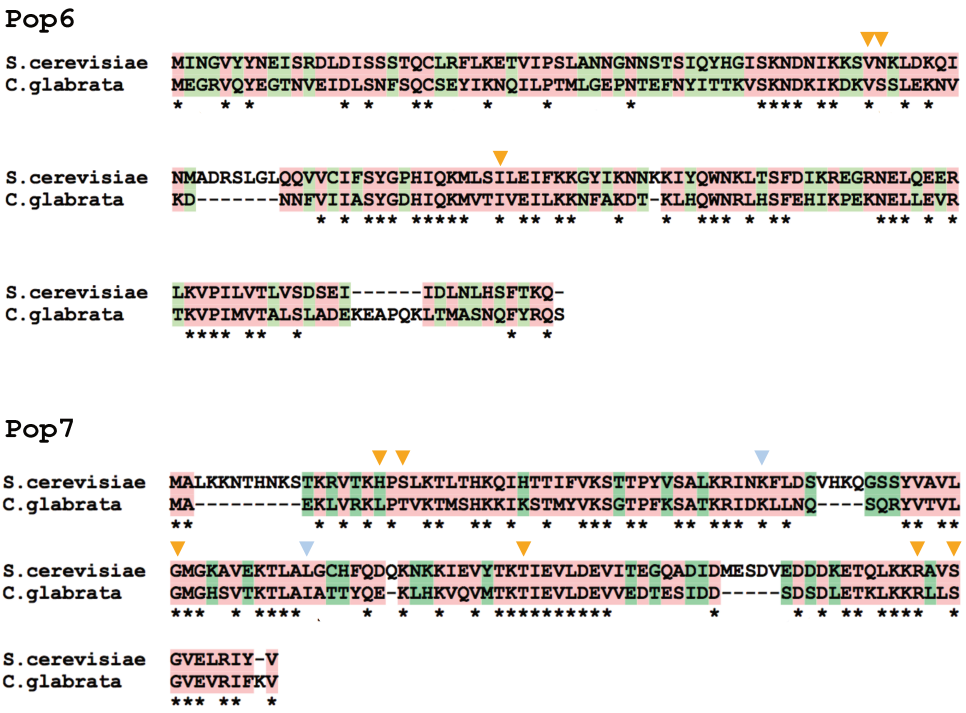

Figure 4.9 Protein alignments of Pop6 and Pop7, the heterodimer binding the P3 domain of RPR1. Alignment of the Pop6 and Pop7 protein sequences of *S. cerevisiae* and *C. glabrata* performed by T-Coffee tool (Notredame et al. 2000). Arrow indicates residue of Pop7 in direct interaction with RNA nucleotide of P3 RPR1 domain (see **Figure 4.8**); Yellow arrow: interaction with a conserved RNA nucleotide in *C. glabrata;* Blue arrow: interaction with RNA nucleotide non conserved in *C. glabrata*.

Pink boxes: amino acid with strong similar properties; Green boxes : amino acid with no similar properties. An asterisk (*) indicates strictly identical residue.

II. Purification of the RNase P complex of *C. glabrata*.

As previously discussed, the presence of additional domains within the RPR1 RNA could influence the proteins of the RNase P complex. This impact could result in a modification of the nature and the number of proteins associated to the RPR1 RNA, in respect to the yeast complexes already described (Chamberlain et al. 1998). Thus the RNase P complex of *C. glabrata* could display new proteins which might either be involved in :

- i) Maintaining a functional structure of the RPR1 RNA despite the constraints induced by the presence of large insertions.
- ii) Bringing new functions to the RNase P which may allow the cleavage of new types of RNA.

These hypotheses were tested by a purification of *C. glabrata* RNase P holoenzyme and a subsequent analysis of the protein moiety by mass spectrometry.

Purification of the RNase P complex of *C. glabrata*

The complex has been purified by affinity chromatography. Regarding the bait choice, two alternatives were suggested: i) using RPR1 RNA as bait by inserting an aptamer within its structure; or ii) using a protein belonging to the complex as a bait. As mentioned above, modifying a non-coding RNA with un-resolved structure is risky. The choice was therefore made to use a protein as bait for affinity chromatography.

Rpr2 protein as bait for affinity chromatography of the RNase P complex

The Rpr2 protein is the only protein in the complex described as exclusive to the RNase P (Chamberlain et al. 1998). Indeed, the eight remaining proteins are shared with the RNase MRP complex (Salinas et al. 2005). To date, there are no available antibodies specific to the Rpr2 protein of *C. glabrata*. It was therefore necessary to insert an epitope into the protein to perform an affinity chromatography. The protein was tagged by three repeats of the HA epitope (3XHA), derived from human influenza virus hemagglutinin . The HA epitope has the advantage of being small (1.1 kDa) and is known to have no or very little influence on the folding of the labelled protein (Kimple et al 2013). Regarding the position of the 3XHA-tag, the choice was made to locate the insertion at the C-terminal end of the Rpr2 protein. In order to determine if the C-terminal region was not folded inside the overall structure of the protein, the structure of the Rpr2 protein was predicted *in silico*. The prediction was performed using the Phyre2 tool, which method is based on multiple structural alignments (Kelley et al. 2015). The predicted structure seemed to indicate that the C-terminal end was not folded inside the protein. Nevertheless, a linker sequence with a neutral charge (composed of a succession of glycine and serine amino acids) was placed upstream of the 3XHA-tag. Using a protein linker is known for increasing the accessibility of the tag and reduces a possible impact on the structure of the given protein (Sabourin et al. 2007).

Integration of the 3XHA-tag in the coding sequence of *Rpr2*

The insertion of the 3XHA-tag into the genome was performed by homologous recombination, enhanced by a double-strand break produced by the CRISPR-Cas9 system, at the insertion site. Since the insertion locate at the C-terminal end of the protein, the sequence recognized by the RNA guide had to be as close as possible to the stop codon. Using the CASTING tool, a potential cutting site was detected 78 nt. upstream of the stop codon. The RNA guide appropriate to this site, named sgRPR2, was synthesized and then cloned within the plasmid pRS315-sgRPR2 to ensure its expression. The insertion site of the cassette containing the 3x-HA tag is located 26 codons upstream of the stop codon. It was then necessary for the integration cassette to re-provide this sequence, located after the cutting site, to allow an expression of the full-length Rpr2 protein. The incorporation of this sequence of 26 codons in the integration cassette had to fulfill three criteria: i) removing the endogenous stop codon to relocate it downstream of the HA tag sequence, in order to allow the expression of a chimeric protein; ii) preserving the protein sequence while clearing any homology with the endogenous genomic sequence, which could lead to homologous recombination; iii) clearing any homology with the genomic sequence recognized by the sgRPR2 guide RNA, in order to avoid a subsequent cut by Cas9, which repair would induce a mutation. The DNA sequence of the 26 codons to be re-provided has been entirely modified by silent mutation, to maintain the protein sequence unchanged. The 3X-HA tag integration cassette within the Rpr2 sequence consisted of: i) 200 bp homology arms that start at the Cas9-induced double-strand break; ii) the sequence modified by silent mutation of the 26 last codons of Rpr2; ii) the 3X-HA tag associated with the neutral linker. (**Figure 4.10 A and B**)

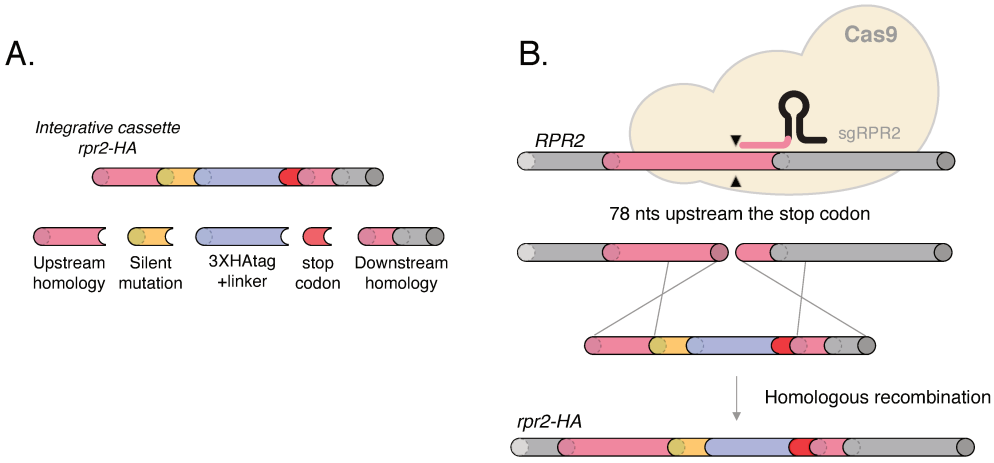

Figure 4.10. Labelling of the Rpr2 protein by a 3X-HA epitope. (**A**) The integrative cassette *rpr2-HA* is composed of 200 bp homology arms, a sequence modified by silent mutations re-providing the 26 codons after the insertion site, the 3X-HA tag downstream a neutral linker, and the re-located stop codon. (**B**) The integrative cassette, the plasmids enabling the expression of the *Cas9* protein and the sgRPR2 single guide RNA were co-transformed in the ΔHTL strain. The integrative cassette enabling the tagging of the Rpr2 protein was inserted into the C-terminal end of the *RPR2* gene by homologous recombination, enhanced by a *Cas9* induced double-strand break at the insertion site.

Tagging of the Rpr2 protein was performed in a ΔHTL strain, by co-transformation with the pRS315-sgRPR2 plasmid, pRS314-Cas9 plasmid and the integration cassette. The transformed cells were streaked on SC-TL selective medium. The correct integration of the Rpr2-HA cassette has been assayed by PCR amplification directly on the colonies. A total of 56 colonies were tested by PCR and 3 colonies showed successful integration (**Figure 4.11 A**). The PCR fragments obtained were sequenced and confirmed the integration of the 3X-HA cassette into the Rpr2 protein. Subsequently, the expression of the Rpr2-HA protein in the three generated strains was confirmed by western-blot, using an antibody directed against the HA epitope (**Figure 4.11B**).

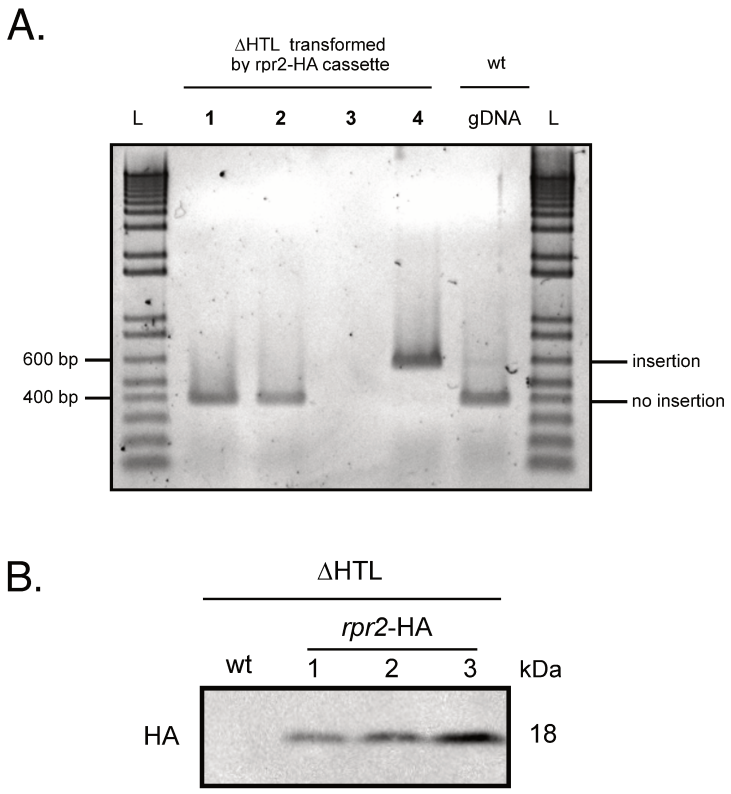

Figure 4.11 Expression of the Rpr2-HA protein. A) Representative sample of 4 out of 48 colonies assayed is presented here. The sequence corresponding to the insertion site was amplified by PCR from colonies having grown on SC-TL medium. The expected size is 707 bp for an insertion event, otherwise the expected size is 399 bp. **B)** The three independent ΔHTL *rpr2*-HA clones were grown in rich liquid medium at 30°C to exponential phase. After total protein extraction, expression of the Rpr2-HA protein was assayed by western-blot using monoclonal @HA antibody.

Co-immunoprecipitation experiments of the Rpr2-HA protein and its partners were performed on the total protein extracts of the three generated strains ΔHTL *rpr2-HA*, which are considered as biological replicates. These experiments were performed using the commercial "Myltenyi µMACS HA isolation kit", which includes anti HA antibodies coupled to magnetic beads. After passing the protein extracts through the purification columns, the eluates obtained were analyzed by nanoscale liquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS). These mass spectrometry experiments were performed in collaboration with the proteomic platform of IBMC, Strasbourg, FRANCE. As a control, all co-immunoprecipitation experiments followed by mass spectrometry were performed with the parental ΔHTL strain. A total of 5 independent experiments were conducted on the three biological replicates. Only 3 out of 5 experiments were successful, meaning that the bait Rpr2-HA was detected in mass spectrometry analysis. Criteria used to consider an identified protein as "validated" are the following: i) the protein should not be detected in the control sample; ii) the protein should be present in all biological replicates; iii) the protein should be present in all independent experiments. The results of the three independent experiments were reported in **Table 6**.

Table 6. Validated proteins co-precipitating with rpr2-HA, identified by nano LC-MS/MS

Validated protein	*Gene ID*	*Number of spectra assigned (mean)*
Pop1	CAGL0K05049g	85,67
Rpp1	CAGL0D03740g	57,67
Pop4	CAGL0I09152g	16,33
Rpr2	CAGL0K08712g	14,33
Pop6	CAGL0K03993g	14
Pop5	CAGL0M11836g	12,33
Pop7	CAGL0M06039g	11,0
Pop3	CAGL0I03014g	7,67
Pop8	CAGL0J04334g	6,00
Rcl1*	CAGL0L12078g	5

** Proteins marked with an asterisk, are not part of the RNase P complex described in S. cerevisiae*

First observation, all eight putative RNase P proteins subunits have been co-purified with Rpr2. These data validate the annotation of the nine *C. glabrata* RNase P protein-coding genes. Moreover, no specific proteins of the RNase MRP, namely Snm1 and Rmp1, have been purified which indicates that the proteins co-purified with Rpr2 are well associated with RNase P. This result is fundamental since it confirms *in silico* analyses: the presence of additional domains within RPR1 does not replace the loss of protein partners. On the contrary, a new protein has been detected within the complex: the Rcl1 protein. In *S. cerevisiae* this protein has been described as involved in ribosomal RNA maturation (Horn et al. 2011). More specifically, the Rcl1 protein is responsible for the co-transcriptional cleavage of the 35S precursor RNA at the A2 site, thus releasing the 18S ribosomal RNA, component of the small ribosomal subunit (Horn et al. 2011). Co-purification of the Rcl1 protein with Rpr2 is so far unpublished and no interactome data in *S. cerevisiae* has reported their interaction yet. The function of this protein is similar to RNase MRP, although no protein from this complex has been found after purification.

Is Rcl1 an additional protein of *C. glabrata* RNase P ?

This is the first time that Rcl1 protein has been co-purified with Rpr2 protein. But is its presence linked to the additional domains of the RPR1 RNA ? This hypothesis was tested by analyzing the sequence of this protein in *S. cerevisiae* and in all *Nakaseomyces* (**Figure 4.12**). Interestingly this protein is strongly conserved among all these species, with for example a sequence identity of 84% between *S. cerevisiae* and *C. glabrata.* No discrepancies correlate with the presence of additional domains in the RPR1 RNA. Based on these data, Rcl1 does not appear to be a constitutive part of the RNase P holoenzyme. This hypothesis is consistent since the Rcl1 protein is self-sufficient to carry out its catalytic activity (Horn et al. 2011).This interaction would therefore appear to be specific to Rpr2 and does not involve the additional domains of RPR1.

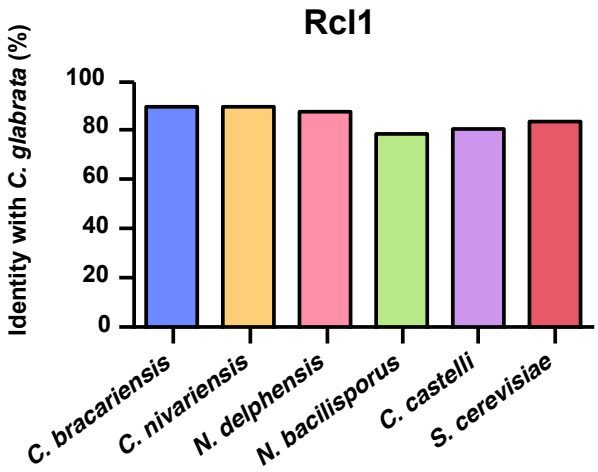

Figure 4.12. Rcl1 protein sequence identity between *C. glabrata*, *Nakaseomyces* species and *S. cerevisiae*. Pairwise alignments of Rcl1 protein sequence between *S. cerevisiae*, *Nakaseomyces* species and *C. glabrata* were performed with T-COFFEE software (Notredame et al. 2000). Identity rates were extracted from the 6 pairwise alignments and represented as bar charts.

This result is nevertheless very interesting because the interaction with Rcl1 could allow Rpr2 to locate the RNase P holoenzyme at ribosomal RNA transcript sites. This theory is consistent with publications that have described the location of RNase P at transcription sites by RNA polymerase I and III (Reiner et al. 2008, Serruya et al. 2015). Interaction between Rpr2 and Rcl1 would therefore be the missing link between RNase P and its participation in transcriptional processes.

Conclusion and perspectives

The work performed during my thesis on the characterization of the nuclear RNase P of *Candida glabrata* brought new findings about the RNA moitety and its interractions with protein partners. Moreover this study provided new insight about the evolution of the non-coding RNA RPR1.

RPR1 genes are not exchangeable between *S. cerevisiae* and *C. glabrata.*

The different experiments I have performed provided evidence that RPR1 genes are not exchangeable between *S. cerevisiae* and *C. glabrata*. The RNase P protein subunits of *C. glabrata* were supposed to be strictly specific to the endogenous RPR1 RNA. Bioinformatic analyses eventually demonstrated the opposite, that is, that there were no major differences between the protein subunits of *C. glabrata* and *S. cerevisiae*. We should note that these analyses only provided a global insight on the possible co-evolution of *C. glabrata* proteins with the RPR1 RNA. Moreover, even if the structural data analysis of the P3 domain did not show major differences between C. glabrata and S. cerevisiae, these results do not exclude a divergence of RPR1 protein partners. Indeed, protein-protein interactions are also important for holoenzyme assembly. The unsuccessful exchange of RNase P RNA moiety is probably due to defective interactions between protein subunits. Testing this hypothesis is for the moment a difficult task, since the available structural data are not yet sufficient to analyze protein-protein interactions within the RNase P. The analysis of protein-protein interactions between S. cerevisiae and C. glabrata RNase P sub-protein units could nevertheless be performed by electrophoretic mobility shift assay (EMSA) or Yeast-Two-Hybrid experiments for example. Each protein of *S. cerevisiae* and *C. glabrata* would therefore have to be expressed, purified, and brought into contact with each other. The effort required for such experiments would be disproportionate to the biological question that will be answered. The RNase P RNA moiety exchange

experiments were conducted to answer an overarching question: "What role does the additional domains of RPR1 play in *C. glabrata* biology? "» . Unfortunately, the strategy used was inconclusive, most likely because of the large number of RPR1 RNA co-factors within eukaryotic cells. A new strategy is under consideration and will be based on a more straightforward method: the substitution of the *C. glabrata* RNase P RNA moiety by a single proteinaceous RNase P protein (PRORP). This experiment has already been performed in *S. cerevisiae* (Weber et al. 2014) and could most probably give new indications regarding the role of additional domains in *C. glabrata*. The most direct technique would obviously consist in the deletion of the additional domains within the *RPR1* gene of *C. glabrata*. This method would nonetheless require the establishment of a precise secondary structure of the RPR1 RNA, and no exploitable structural data is available yet. Chemical probing and SHAPE (Selective 2'-hydroxyl acylation analyzed by primer extension) experiments were performed before my arrival at the laboratory, unfortunately without success. The presence of large additional domains within this RNA might thus prevent the resolution of its structure.

The link between additional domains of RPR1 and the appearance of new proteins within the holoenzyme in *C. glabrata*.

Purification experiments of the RNase P protein subunits of *C. glabrata* revealed the presence of a new protein: Rcl1. Even so, it is difficult to state whether the presence of Rcl1 is linked to the emergence of RPR1 RNA additional domains. The strong sequence conservation rate of this protein between *S. cerevisiae* and *C. glabrata* suggests that the additional domains of RPR1 have no influence on this interaction. Is this interaction specific to *C. glabrata* ? Considering several studies describing the implication of nuclear RNase P in transcription processes, this interaction is likely to be preserved in eukaryotes. The late discovery of this interaction between Rpr2 and Rcl1 is probably due to improved mass spectrometry techniques. Indeed, a large number of publications on nuclear RNase P indicates that the low abundance of this

enzyme is a major obstacle to its study. I encountered the same problem throughout my thesis, especially during the Rpr2 labelling experiments. The first strategy adopted, which was not presented in this manuscript, was to express the chimeric protein Rpr2-HA on a plasmid. However, the amount of purified protein was too low to be analyzed by mass spectrometry, presumably because of a competitive effect between the endogenous Rpr2 protein and the HA tagged version. The insertion of the HA epitope directly into the genomic Rpr2 locus solved this problem. Given the emergence of new analytical technology such as nanoLC-MS/MS (used in this study), it would be wise to perform new purification experiments of *S. cerevisiae* RNase P protein subunits.

The co-immunoprecipitation experiments performed during this thesis were carried out under native conditions. Repeating these experiments after chemical or UV cross-linking could allow the detection of new proteins interacting only transiently with the RNase P. Given the large size of the additional domains, it is very likely that proteins are directly associated with them, but my thesis work has not been able to demonstrate it. Again, the precise description of the RPR1 RNA structure will allow us to integrate an aptamer, thereby directly allowing the purification and analysis of the holoenzyme. The purified holoenzyme would also enable the identification of alternative RNase P substrates, by sequencing co-purified RNAs.

Emergence and conservation of large RNase P RNA moieties within the *Nakaseomyces.*

As previously mentioned, the emergence of additional domains within the RPR1 RNA goes against the evolution of eukaryotic RNase P. These insertions appeared suddenly in the last common ancestor of *Nakaseomyces* species and were more or less maintained during the evolution of this clade. Preliminary bioinformatics comparison of the RPR1 sequences among *Nakaseomyces* indicated a great diversity in the nature and position of additional domains (not shown in this manuscript). This divergence suggests that these areas are not essential for the function of the RNase P and that no selection pressure has maintained them. How did these domains emerged ? The

favored hypothesis would implicate errors of the DNA polymerase in the appearance of multiple duplications within the *RPR1* gene. Acquisition of these additional sequences being neutral, the domains could diverge freely throughout the evolution. The presence of duplication signals was investigated within additional domains of *C. glabrata* but no significant results were obtained. This potential mechanism appears to be consistent with the « Constructive Neutral Evolution theory" ((Stoltzfus 2012) recently proposed to explain the conservation of Rnase P RNA moiety within eukaryotes ((Gopalan et al. 2018)

The study of *C. glabrata* RNase P during my thesis was a double challenge. First, the characterization of a ribonucleoprotein complex is difficult when the majority of its properties, such as the structure of the RNA moiety or the nature of the protein partners, are unknown. Eukaryotic RNAse P with its structures and functions remain enigmatic, despite the large number of publications on this topic. The second challenge was intrinsic to the nature of *C. glabrata*. Even if this yeast is phylogenetically close to *S. cerevisiae*, it is not yet recognized as a yeast model for molecular biology studies, despite its great potential. Indeed, *C. glabrata* lacks a large number of tools, such as epitope-tagged protein library for instance, and the manipulation of its genome is much more complicated and time consuming than in *S. cerevisiae*. Even with the help of the CRISPR-Cas9 technology, both the low homologous recombination efficiency and the lack of mating conditions make the study of *C. glabrata* molecular mechanisms challenging.

Chapter II. Improvement of genome editing tools in *C. glabrata.*

Context

One of the topic of my host laboratory relies in the identification of *C. glabrata* loci involved in a systemic infection of a host. To identify these genes, the strategy is based on the use *Drosophila melanogaster* as an animal model of infection. This model has many advantages, in particular ease of handling, short generation time and highly studied innate immunity (Brunke et al. 2015). Infection experiments were performed on the *MyD88* strain (Myeloid Differenciation primary response 88) an immunodeficient strain of *D. melanogaster* (Tauszig-Delamasure et al. 2002). This strain presents a mutated effector of the Toll pathway, which inhibits the production of drosomycin, an antimicrobial peptide. The immune system of these flies mutant is no longer able to control the proliferation of *C. glabrata* and thus quickly succumb to infection (Quintin *et al*., 2013). Before my arrival in the laboratory, a transcriptomic analysis of *C. glabrata* was performed during the *D. melanogaster* infection process. This analysis led to the identification of several genes differentially expressed during systemic infection of *D. melanogaster*. Within this cohort, a particular interest has been shown in genes with unknown functions. To study these genes, the CRISPR-Cas9 system was developed in *C. glabrata* (Enkler et al. 2016). To assess the impact of a given gene during the infection process, the following workflow has been set-up (**Figure 5.1**) : i) genes are disrupted in the triple auxotrophic strain ΔHTL, by insertion of the *HIS3* marker into the coding sequence enhanced by a Cas9 double-strand-break; ii) immunodeficient flies are infected with loss-of-function mutants and their survival is monitored. In comparison with a parental ΔTL strain, if infected flies survival showed a significative shift, the gene was considered as potentially involved during infection of the fly and was labelled as candidate gene .

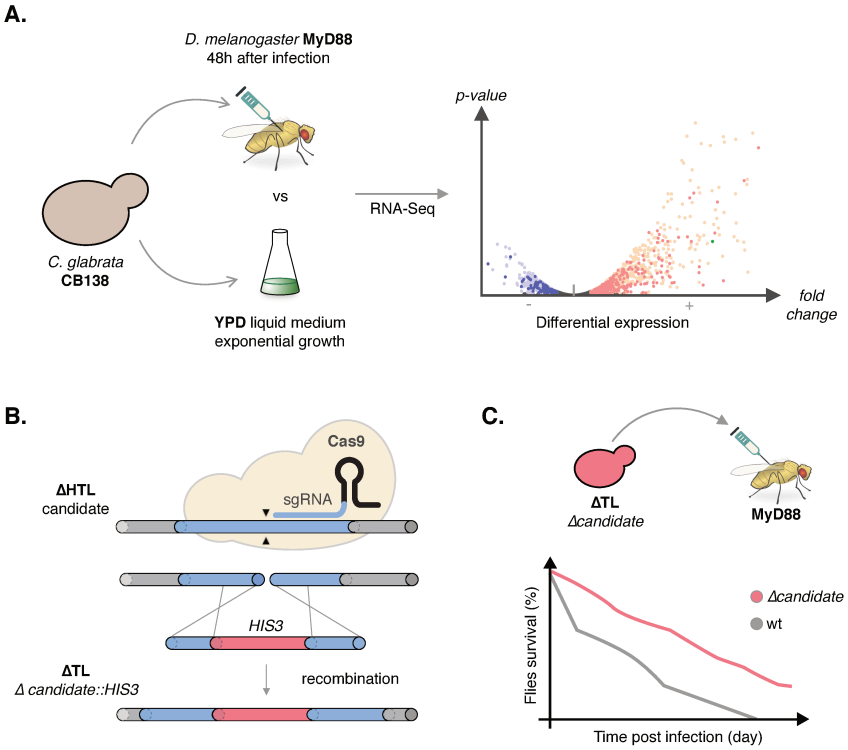

Figure 5.1 Study of specific *C. glabrata* genes differentially expressed during *D. melanogaster* infection. A) After 48h infection of *D. melanogaster* by the reference strain CBS138, total RNAs of *C. glabrata* were extracted, sequenced and compared to RNA expressed during exponential growth in YPD liquid medium. **B**) The study was first focused on over-expressed genes without homologs in *S. cerevisiae*. To decipher their role during infection, loss-of-function mutants were generated using CRISPR-Cas9 technology. In a auxotrophic ΔHTL strain, selected genes were disrupted by inserting the *HIS3* marker into the open reading frame (ORF). **C)** Immunodeficient flies (*MyD88*) were infected by the loss-of-function mutants and their survival was monitored. In comparison with a reference strain ΔTL, if infected flies survival showed a significative shift, the gene was considered as potentially involved during infection of the fly and was labelled as candidate gene.

Throughout my thesis, I participated in the characterization of one of these candidate genes.The *CAGL0K07678g* gene belongs to the 10% most overexpressed genes during the infection of *D. melanogaster*. Following the workflow, the *CAGL0K07678g* gene was disrupted within the ΔHTL strain via the insertion of an *HIS3* marker and its impact on virulence was assayed by infection of immunodeficient flies (**Figure 5.2**). As a control, infections were carried in the same time on immunocompetent flies, wild-type *Drosophila* A5001. The fly survival analysis showed that the deletion of the *CAGL0K07678g* gene increased significantly the survival of infected flies. This result therefore indicated that *CAGL0K07678g* gene is somehow likely involved in the virulence of *C. glabrata*. In order to characterize the function of this gene, we first analyzed its expression under multiple conditions. In particular, we have demonstrated that the expression of *CAGL0K07678g* is induced after a shift in a low nitrogen environment (nitrogen starvation medium). This stress is the same as *C. glabrata* undergoes in the phagosome, and we have confirmed the *CAGL0K07678g* overexpression during human macrophages infection. To bring new information concerning the role of *CAGL0K07678g*, we wanted to analyze the behavior of the mutant Δ*caglk07678g::HIS3* in nitrogen starvation condition. Surprisingly, we faced a major problem: the auxotrophic parental *C. glabrata* strains ΔHTL and ΔTL, are not capable of growing in a nitrogen starvation medium.

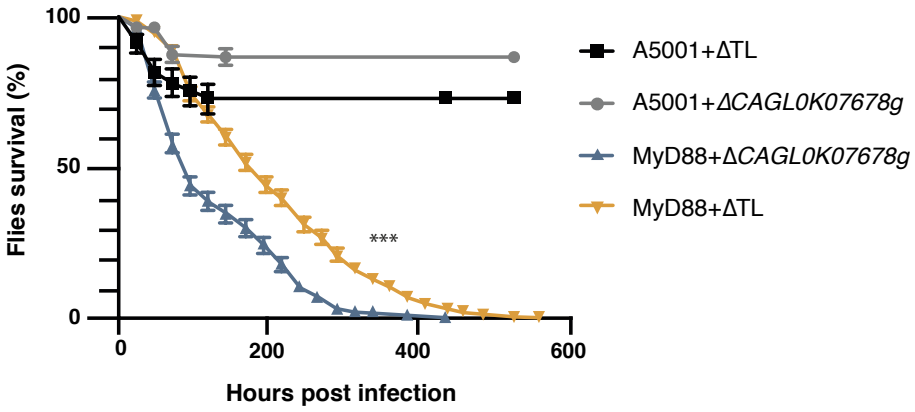

Figure 5.2 Deletion of CAGL0K07678g in ΔHTL strain impedes *C. glabrata* virulence in *D. melanogaster*. (**A**) Survival curves of A5001(wt) and MyD88 flies infected by ΔTL or Δ*cagl0k07678g::HIS3* strains. Two statistical tests (Mantel-Cox and Gehan-Breslow-Wilcoxon) were performed using Prism software (GraphPad). **** *p*value<0.0001.

Following this observation, I disrupted the *CAGL0K07678g* gene within the reference strain CBS138/ATCC2001, which has been shown to grow in nitrogen starvation medium. To disrupt this gene in the prototrophic strain CBS138, I used the strategy developed during the RNase P study, namely the integration of the selectable marker *Sh Ble*, conferring resistance to the antibiotic Zeocin. Following the same procedure as described in Chapter I, I designed an integrative cassette flanked by 500bp homologous sequences upstream/downstream the insertion site. After transformation, cells were streaked on YPD supplemented by Zeocin. The correct integration of the *Sh ble* marker has been assayed by PCR amplification directly on the colonies, using the same primers as for the amplification of the complete *Sh ble* cassette **(Figure 5.3**)Several independent mutants Δ*caglk07678g::ShBle were* thus obtained. To confirm the results obtained with the mutant Δ*caglk07678g::HIS3* in the ΔHTL strain, I performed flies infection experiments with the mutants obtained in the reference strain CBS138 (**Figure 5.4**). Surprisingly, the results obtained were contradictory with the previous ones: the disruption of gene *CAGL0K07678g* in the reference strain CBS138 does not significantly increase the survival of infected flies. This observation therefore challenges the direct involvement of *CAGL0K07678g* in the virulence of *C. glabrata.* Moreover all, this observation questions the use of the auxotrophic strains ΔHTL and ΔTL, for the study of *C. glabrata* pathogenicity. The decreased virulence of the Δ*caglk07678g::HIS3* mutant within the strain ΔHTL would be a combinatorial effect due to auxotrophies.

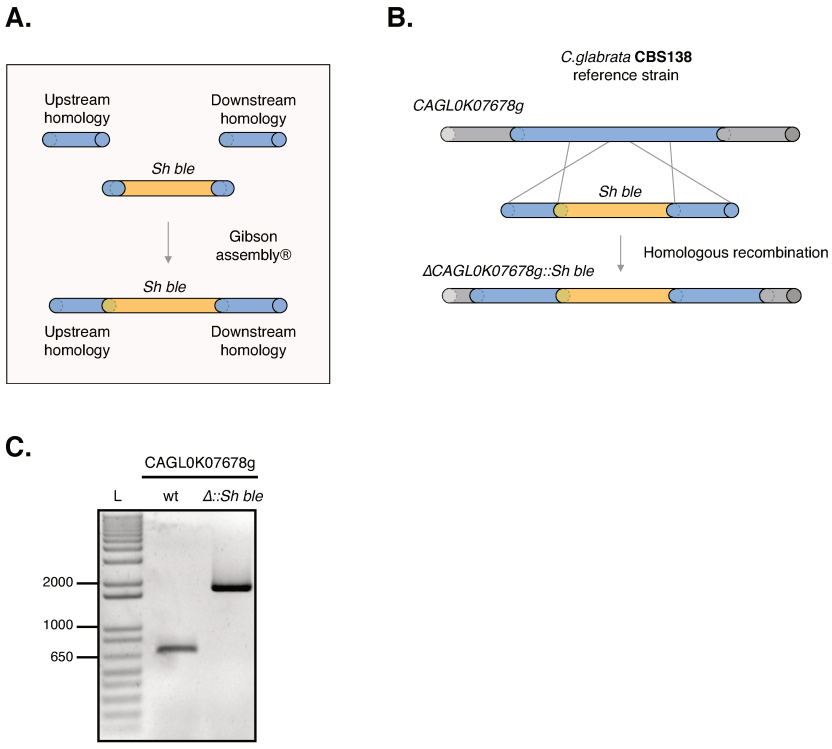

Figure 5.3 Deletion of *CAGL0K07678g* in the reference strain CBS138. (**A**) The DNA fragments constituting the disruption cassette of *CAGL0K07678g* were fused by the "Gibson Assembly®" method. (**B**) The disruption cassette was transformed in CBS138 reference strain and inserted in the *CAGL0K07678g* ORF by homologous recombination. (**C**) A representative sample of 1 out of 36 colonies assayed is presented here. Sequence corresponding to the insertion site was amplified by PCR from colonies having grown on YPD medium + Zeocin™, after the transformation of the *Sh Ble* cassette. The same primers as for the amplification of the complete disruption cassette were used for this experiment. The expected size was 1840 bp for an insertion event, otherwise the expected size was 668 bp.

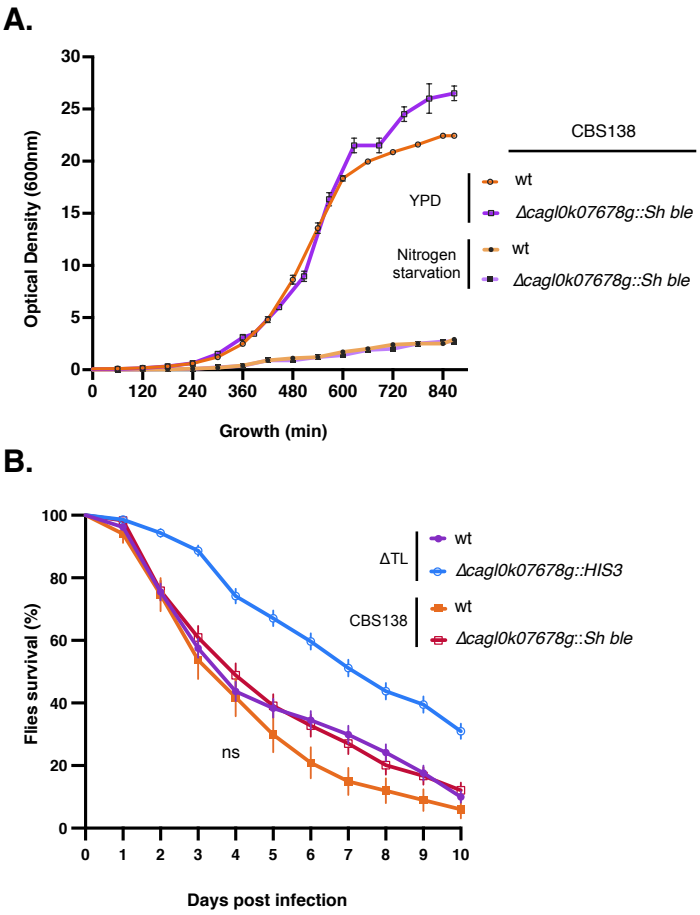

Figure 5.4 Deletion of CAGL0K07678g in the CBS138 strain has no significative impact on *C. glabrata* virulence in *D. melanogaster*. (**A**) Growth were monitored for CBS138 Δ*cagl0k07678g*::*Shble* and CBS138 wild-type strains in rich (YPD) and nitrogen-starvation media by measuring optical density at 600 nm length-wave. (**B**) Survival curves of MyD88 infected by ΔTL, ΔTL Δ*cagl0k07678g::HIS3*, CBS138 and CBS138 Δ*cagl0k07678g::Shble*. Two log-rank statistical tests (Mantel-Cox and Gehan-Breslow-Wilcoxon) were performed using Prism software (GraphPad).

Aim of the study

Auxotrophic *C. glabrata* strains have been described as suitable for the study of their pathogenicity (Jacobsen et al. 2010, Brunke et al. 2015). However our data suggested the opposite : the use of the triple auxotrophic ΔHTL strain induced a bias in the study of *C. glabrata.*

This observation was the starting point of the study presented in this chapter. It was essential to develop new tools to modify the genome of *C. glabrata* without the use of auxotrophies. Throughout my thesis work, I have set up new methods allowing the genome editing of *C. glabrata* prototrophic strains.

Results and discussion

The gene disruption method by the integration of a selectable marker into the coding sequence is a conventional technique. This method was notably used in *C. glabrata* for the creation of loss-of-function mutants collection, by the integration of the *NAT* **(N-acetyl** transferase) cassette (Schwarzmüller et al. 2014) However, information provided by the disruption of a gene is generally not sufficient to understand its function. Given genes need to be modified more finely, for instance by modulating their expression or by the generation of a chimeric protein. The integration of complete exogenous sequences within a genome also enables to provide new insights about the biology of an organism. The integration of genes such as GFP (**G**reen **F**luorescent **P**rotein) and its derivatives; or heterologous enzymes (such as PRORP protein) is a useful way to decipher molecular processes within a cell. Concerning the studies focused on *C. glabrata,* the expression of mutated gene is much more supported by plasmids than by direct genome modification. Two facts explain likely this situation: i) the wide range of plasmids designed for *S. cerevisiae* are functional within *C. glabrata;* ii) compared to *S.* cerevisiae, homologous recombination is weakly effective for *C. glabrata* and requires long homology arms(Cormack and Falkow 1999) Transient expression via a plasmid may not be the best method. Plasmids must be maintained within the cell by selection pressure, so the environment in which *C. glabrata* interacts must be completely controllable. Unfortunately, environmental control is not always possible, especially for host infection experiments.Moreover, the plasmid maintenance within *C. glabrata* is often possible via auxotrophy of the recipient strain. As previously demonstrated, auxotrophies can eventually provide significant bias in the study of *C. glabrata.* In this context, I have designed new tools to overcome the current limitations regarding genome editing of a prototrophic strain of *C. glabrata.*

A. Protein labelling in the reference strain CBS138

The protein labelling with an epitope is a very convenient method because it provides the ability to monitor the expression of a given protein, to determine its location and to characterize its partners. Moreover, this method technique useful for the study of *C. glabrata* since very few antibodies specific to this species are available. Studies focusing ont *C. glabrata* used mostly plasmid carried chimeric proteins. I have developed a method for the integration of an epitope at the C-terminal end of a protein, positively selectable in a prototrophic strain of *C. glabrata*. This work focused mainly on the integration of the 3X-HA epitope associated with a neutral linker, previously used in the RNase P study. Concerning the positive selection of an integration event, I chose the *Sh ble* cassette, conferring resistance to the Zeocin™ antibiotic.

I. Design of the integrative cassette

To select positively the integration of the 3X-HA epitope at the C-terminal end of a protein-coding gene, the integrative cassette had to fulfill four criteria : i) be as small as possible, in order to avoid the emergence of point mutations during PCR amplification of the cassette; ii) be flanked by homology arms of sufficient size to allow homologous recombination; iii) removing the endogenous stop codon to relocate it downstream of the HA tag sequence, in order to allow the expression of a chimeric protein; iv) simultaneously bring the *Sh Ble* resistance cassette, while preventing the transcription of an aberrant RNA.

Therefore this integrative cassette had to contain a RNA polymerase II terminating sequence, between the 3X-HA epitope stop codon and the *Sh ble* marker promoter. The *CYC1* gene terminator sequence from *S.* cerevisiae is mostly constructs. However the large size of this sequence (367 nt.) present a disadvantage for the design of a short integrative cassette. Therefore integrated a short synthetic terminator, designed from the works of Guo and colleagues (Guo and Sherman 1996) which size is only 39 nt. As proof of concept, the protein encoded by *CAGL0K07678g* gene was chosen to be labelled by the 3XHA-*ShBle* cassette **(Figure 5.5**). The fragment of the 3XHA-*ShBle* integrative cassette, containing the 255 nt. upstream homology sequence; the neutral protein linker; the 3X-HA epitope; the synthetic terminator and a 20 nt. homologous sequence with the *Sh ble* cassette promoter, has been chemically synthesized (by Integrative DNA Technologies, Inc). The other components of the cassette were amplified by PCR and assembled with the Gibson Assembly® method (**Figure 5.5A**).

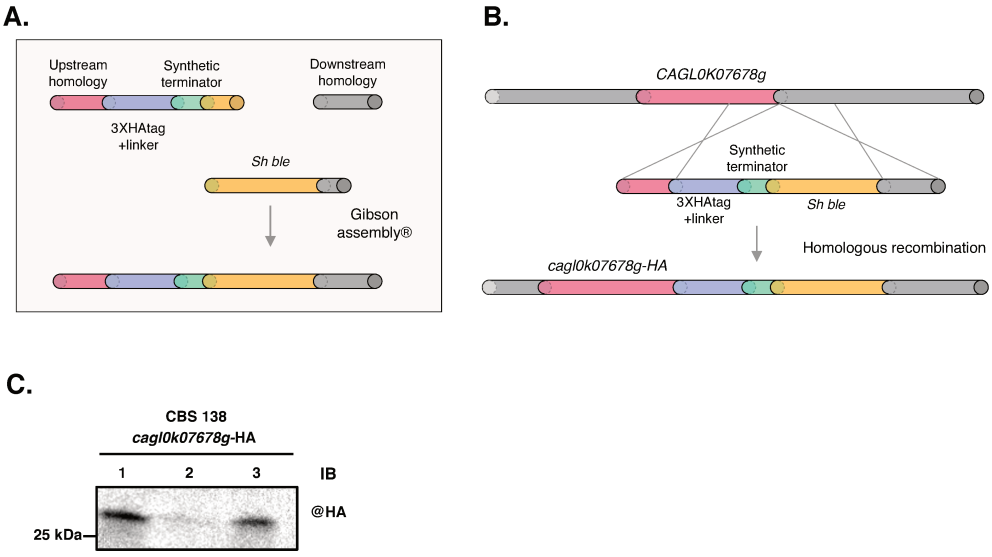

Figure 5.5. Tagging of the CAGL0K07678g protein by a 3X-HA epitope in a CBS138 strain. (**A**) The integrative cassette Cagl0k07678g-HA is composed of left and right homology arms of respectively 255 bp and 274 bp; the 3X-HA tag downstream a neutral linker; the synthetic terminator from (Guo and Sherman 1996) and the *Sh ble* selectable marker conferring the resistance to Zeocin™. (**B**) The integrative cassette enabling the tagging of the Cagl0k07678g protein was transformed in the CBS138 reference strain, and inserted into the C-terminal domain by homologous recombination. (**C**) **Expression of the *cagl0K7678g*-HA fusion protein.** The three independent *cagl0K7678g*-HA mutants were grown in rich liquid medium at 30°C to exponential phase, cells were then pelleted and transferred in liquid nitrogen-starvation medium for 6h. After total protein extraction, the expression of the *cagl0K7678g*-HA protein was assayed by Western Blot using monoclonal @HA antibody. IB: Immune blotting.

The labelling of the Cagl0K07678 protein was performed in a CBS138 strain, by transformation of the 3X-HA-*ShBle* integrative cassette. The transformed cells were streaked on Zeocin supplemented YPD medium. The correct integration of the 3X-HA-*Shble* cassette has been assayed by PCR amplification directly on the colonies. A total of 32 colonies were tested by PCR and 3 colonies showed successful integration. The PCR fragments obtained were sequenced and confirmed the integration of the 3X-HA cassette into the Cagl0k07678 protein. Subsequently, the expression of the Cagl0k07678-HA protein in the three generated strains was confirmed by western-blot, using an antibody directed against the HA epitope **(Figure 5.5.C**). The clone 2 appeared to have a weaker expression than the others, however the poor quality of the load controls did not give insight about this divergent expression.

These data have proved that the 3X-HA-*ShBle* integration cassette is functional. The development of this cassette is therefore transposable to virtually any protein of *C. glabrata*, and allowed an effective protein labelling without the need for auxotrophies.

B. Integration of a full gene into the genome of the reference strain CBS138.

The integration of a full gene sequence (promoter + coding sequence + terminator) within a genome is a commonly used method to restore the function of a deleted gene, in order to confirm the phenotype of a mutant. Moreover, the integration of a gene within the genome is also used to bring new functions to a given organism, such as the constitutive expression of a fluorescent protein. Concerning *C. glabrata,* the integration of a fluorescent protein would be very useful to follow the evolution of fungal cells during the infection process of a host. Data obtained for the integration of 3X-HA epitope, within the CBS138 reference strain, suggest that such method can be applicable for the integration of a complete gene. Nevertheless, the application of this integration method faces a major constraint, the size of the integrative cassette. Indeed the average eukaryotic genes size is about 1300 nt. ((Xu et al. 2006) Since the *Sh ble* marker has a size of 1172 nt., fusion of this marker with a full gene would give the integrative cassette a size of about 3000 nt., including the homology arm. A PCR amplification of a such fragment would likely lead point mutation emergence. It was essential to reduce the size of this cassette. Using a synthetic terminator was not sufficient for this reduction, so I focused on promoter sequences. Indeed, the promoter sequence most commonly used in exogenous protein expression is the promoter of the gene encoding the **G**lyceraldehyde-3-**p**hosphate **d**ehydrogenase (*GPD*). The sequence of this constitutive promoter 655 nt. long. To cope with this constraint, I have developed new synthetic promoters, based on the work of Redden and Alper (Redden and Alper 2015) whose promoter sequence size varies between 80 nt. and 130 nt., at least 5 times shorter than the GPD promoter. These short promoters developed for *S. cerevisiae*, provided a transcriptional activity comparable to constitutive promoters. Moreover, due to their synthetic nature, they have no homology with genomic sequences of *C. glabrata,* thus preventing non-specific recombination events.

I. Development of short synthetic promoter for *C. glabrata*.

The synthetic promoters were built following the indications of Redden and colleagues (Redden and Alper 2015). Their promoter were composed of a synthetic upstream activating sequence (UAS); a 30 bp. neutral AT-rich sequence to distance the UAS upstream the TATA box; a 30 bp. synthetic core promoter and a transcription start site (**Figure 5.6.A**). The first synthetic promoter designed for *C. glabrata* showed a similar structure. Concerning the modifications made for *C. glabrata* : UAS C was chosen; an A/T rich spacer of 30 bp and an A/T rich transcription start site (TSS) of 19 bp were randomly generated (**Figure 5.6.B**).

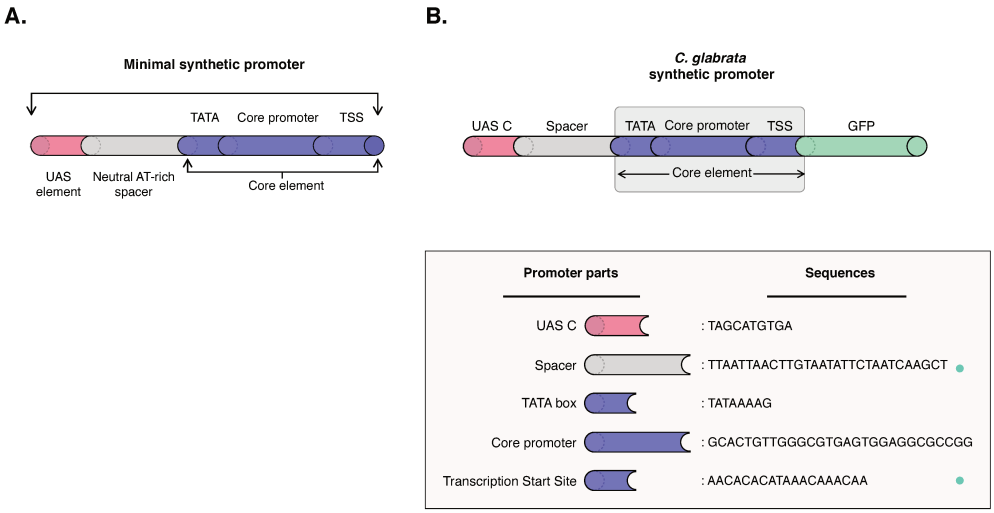

Figure 5.6. Development of short synthetic promoter for *C. glabrata*. (**A**) The synthetic promoters were built following the indications from Redden and colleagues (Redden and Alper 2015). Their promoter were composed of a synthetic upstream activating element (UAS); a 30bp neutral AT-rich sequence to distance the UAS upstream the TATA box; a 30bp synthetic core promoter and a transcription start site. (**B**) The first synthetic promoter designed for *C. glabrata* presented a similar structure. Concerning the modifications made for *C. glabrata* (green dots): UAS C was chosen; an A/T rich spacer of 30 bp and an A/T rich transcription start site (TSS) of 19 bp were randomly generated.

Integration of the GFP coding sequence within the reference strain CBS138

The insertion of the GFP coding sequence was chosen as a proof of concept (**Figure 5.7**). The first step was to amplify by PCR both homology arms fragments. The upstream homology arm was amplified by overhang primers containing a 20 nt. sequence homologous to the synthetic promoter and a 20 nt. sequence homologous to the *GFP* sequence The downstream homology arm was amplified by overhang primers containing a 20 nt. sequence homologous to the *Sh ble*. Both homology arms were assembled with the other fragments by the "Gibson Assembly®" method (New England Biolabs®). The CBS138 *C. glabrata* reference strain transformed by this cassette **(Fig. 5.7B)** and gene integration was confirmed by PCR **(Fig. 5.7C**).

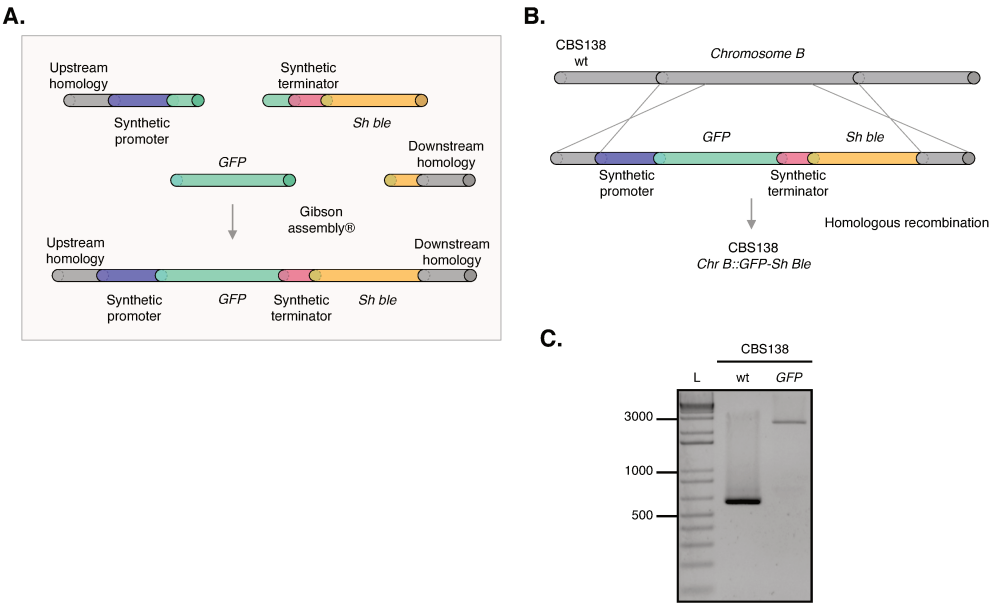

Figure 5.7 Integration of the GFP cassette in the CBS138 reference strain. (**A**) The integrative cassette promoting the expression of the Gfp protein is composed of left and right homology arms of respectively 502 bp and 523 bp; the coding sequence of *GFP* gene; the synthetic terminator from the work of Guo and colleagues (Guo and Sherman 1996) and the *Sh ble* selectable marker conferring the resistance to Zeocin™. (**B**) The GFP - *Sh ble* cassette was transformed in the CBS138 reference strain and was inserted into the chromosome B. (**C**) A representative sample of 1 out of 36 colonies assayed is presented here. The sequence corresponding to the insertion site was amplified by PCR from colonies having grown on YPD medium + Zeocin™. The expected size was 2650 bp for an insertion event, otherwise the expected size was 607 bp.

Expression of the GFP protein.

The GFP protein expression can be visualized by microscopy, however, observation of the transformed cells showed no fluorescence signal. To provide a finer detection of the GFP protein, a western blot analysis was performed **(Fig. 5.8A**). As control, these experiments were also performed in a ΔHTL strain transformed with the plasmid pRS313-GFP, allowing the expression of the GFP gene under the GPD promoter. Western-blot data showed the absence of GFP protein in the samples. It was hypothesized that the absence of GFP signal could result from the absence of messenger RNA expression (mRNA). An RT-PCR analysis was performed (**Figure 5.8.B**) and showed the expression of the GFP mRNA. An RT-qPCR experiment supported the efficiency of the synthetic promoter **(Figure 5.8 C**). These data indicated that the problem certainly lay in the translation stage of the GFP mRNA

Why is the GFP protein not functionally translated ?

RT-PCR experiments were performed with a forward primer annealed to the ATG codon and showed that the GFP mRNA was correctly transcribed. However the protein was not functional. I suggested that the transcription of a non-functional GFP was due to the start position of the transcription, which might be slightly downstream start codon.

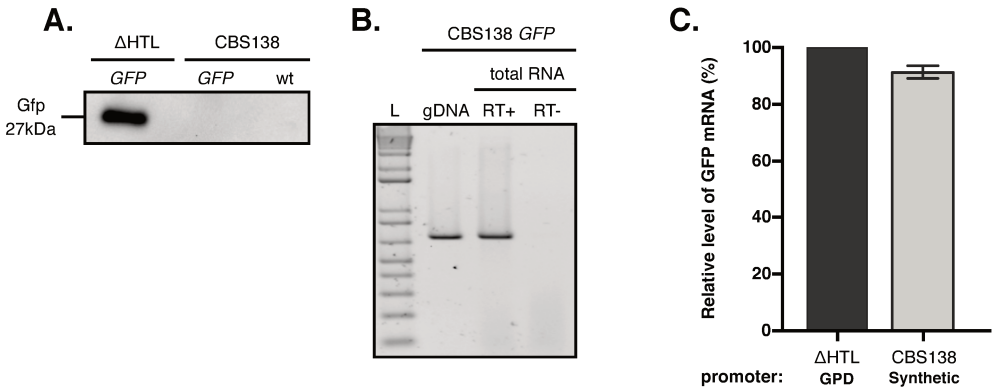

Figure 5.8 Expression of the GFP protein. (**A**) GFP protein was not functionally synthesized. Three independent CBS138 *GFP* colonies were grown in rich liquid medium at 30°C to exponential phase. After total protein extraction, the expression of the GFP protein was assayed by western blot using monoclonal anti-GFP antibody. As a control same experiments were performed on the wild-type CBS138 strain and on the ΔHTL GFP strain, which contains the *GFP* coding sequence under the constitutive GPD promoter. (**B and C**) **The *C. glabrata* synthetic promoter induced the transcription of the *GFP* gene**. (**B**) Expression of the *GFP* mRNA was evaluated by RT-PCR. The sequences of interest were amplified at cDNA level by PCR. Positive control : same PCR experiments were performed on genomic DNA (gDNA lane). Negative control : same PCR experiments were performed on RNA sample that had not undergone retro-transcription (RT-). (**C**) Expression of *GFP* mRNA was measured by RT-qPCR and the results were analyzed using Prism software (GraphPad).

II. Development of three new synthetic promoters for *C. glabrata*.

To test this hypothesis, I built three new synthetic promoters **(Figure 5.9**), which differ from the original promoter.

- Synthetic Promoter A: To distance the TATAbox from the start codon, a 30 bp A-rich sequence was inserted between the TATA Box and the transcription initiation site (TSS).
- Synthetic Promoter B: Throughout my thesis work I performed multiple PCR with short primers. The shortest primer I was able to use was 13 nt. long. The forward primer used to amplify the GFP coding sequence was 19 nt. long. Hence, I created a small spacer of 6 random nucleotides and I placed it upstream the TSS to relocate it away from the initiation codon.
- Synthetic Promoter C: This promoter was designed after preliminary tests on Synthetic Promoter A and B. For this promoter I assessed the impact of three different UAS (Upsteam Activation Sequences) combination on the efficiency of the Synthetic Promoter A.

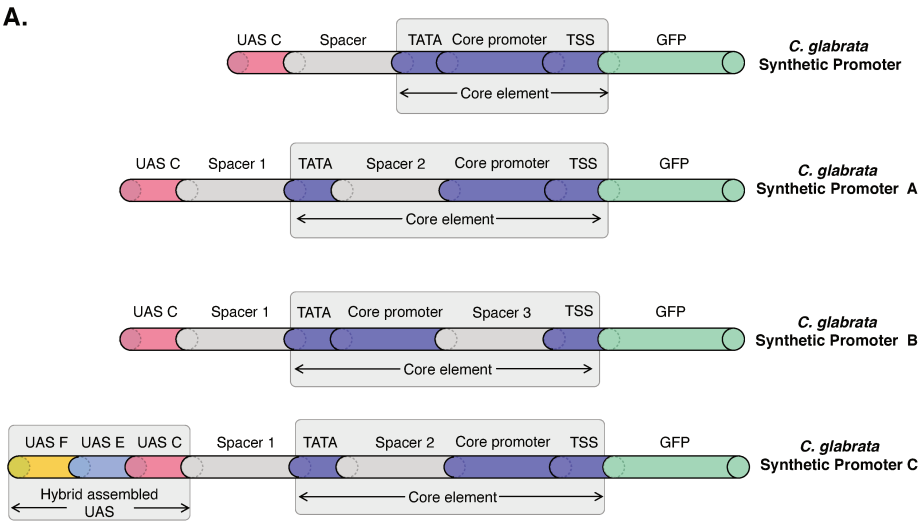

Figure 5.9. Improvement of short synthetic promoter for *C. glabrata*. The original synthetic promoter of *C. glabrata* was modified to allow the synthesis of a functional protein. **Promoter A**: a 30 bp A-rich sequence (spacer 2) was inserted upstream of the core promoter to distance it from the TATA box. **Promoter B**: a small 6 bp sequence (spacer 3) was inserted upstream of the TSS. **Promoter C**: based on promoter A, UAS F and E, described by Redden and colleagues (Redden and Alper 2015) were combined with UAS C.

For convenience and to reduce costs, the experiments assessing the new synthetic promoters efficiencies were performed within the ΔHTL strain. The three synthetic promoters were cloned upstream the GFP coding sequence in the pRS313-GFP plasmid. The transformed strains fluorescence was visualized by confocal microscopy (LSM700, Zeiss ; **Fig. 5.10A**). The expression of the GFP protein was thus confirmed in all strains.

I performed RT-qPCR experiments to quantify the efficiencies of the synthetic promoters, and to compare them with the GPD promoter **(Fig. 5.10B**). None of the synthetic promoter tested was effective as the GPD promoter. However their efficiencies are still remarkable given their short size. These data validated the hypothesis that the initiation codon was too close to the « Core Element ». Indeed, increasing the distance between the core promoter and the TATA-box (Synthetic Promoter A) had the same effect as increasing the distance between the TSS and the core promoter (Synthetic Promoter B). Both modifications provided a consistent expression of the GFP protein. Moreover the TATA-box likely was to close to the initiation codon, increasing the distance with only 6 nt. solved the problem.

The combination of three UAS within the Synthetic Promoter C, provided a 10% increase of efficiency compared to the Synthetic Promoter A. The synthetic promoter B was the most effective promoter. The presence of a 30 nt. spacer between the TATA-box and the core promoter therefore seemed to decrease the efficiency of the Synthetic Promoter A and C.

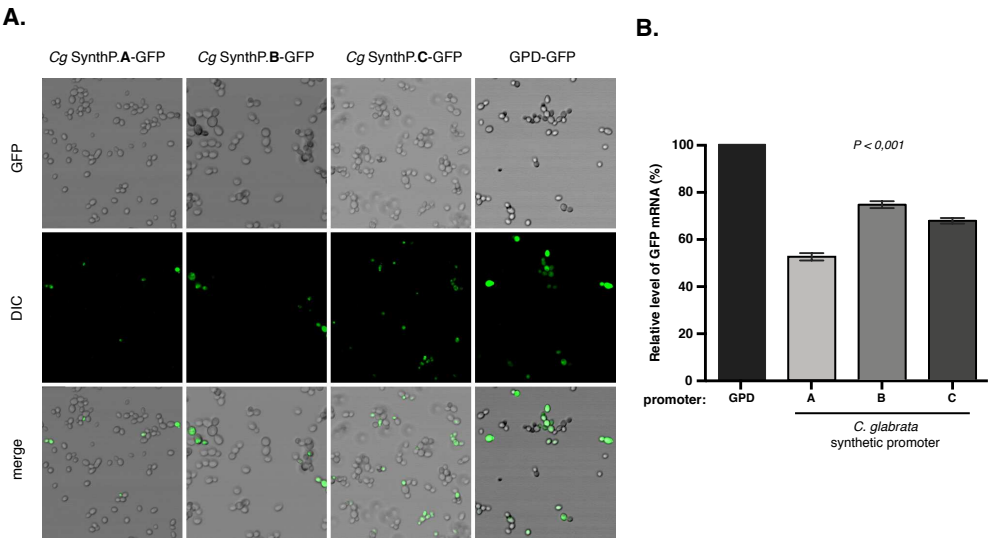

Figure 5.10. The three new synthetic promoters for *C. glabrata* provide an efficient expression of the GFP protein. (A) Translation efficiency of the Gfp protein was visualized by confocal microscopy (LSM 700, Zeiss). Images were processed using ZEN (Zeiss) and ImageJ software. (B) The *GFP* mRNA expression was measured by RT-qPCR and the analysis of the results was performed using Prism software (GraphPad). Unpaired t test with Welch's correction were performed to test the significant differences between each promoter efficiency.

Conclusion and perspectives

The work performed during my thesis on the improvement of genome editing tools for *Candida glabrata* brought new findings that may be useful to overcome the limitations encountered in its study. Moreover this work provided new insight about the use of auxotrophic strain of *C. glabrata*.

Auxotrophies may represent a bias in the study of the virulence of *C. glabrata*.

The development of auxotrophic *C. glabrata* strains was a major step forward, allowing the range of tools available for *C.glabrata* to be extended. Such strains allowed the use of plasmids commonly used in *S. cerevisiae.* Auxotrophies did not appear to interfere with the biology of *C. glabrata* (Schwarzmüller et al. 2014), unlike deletion of the *URA3* marker in *C. albicans* which is known to affect its virulence (Brand et al. 2004). However recent studies have highlighted the profound impact of auxotrophies on yeast transcriptome, proteome and metabolome, despite their supplementation in culture media (Alam et al. 2016). Moreover our data suggested that auxotrophies may have a combinatorial role with a loss-of-function mutant, and could lead to misinterpretations about the virulence. My work obviously does not challenge all former studies performed with the ΔHTL strain. Rather, it highlights the precautions we must take in reaching hasty conclusions. It seems however reasonable to abandon the study of autotrophic strains in favour of wild-type strains. This is even more consistent in view of recent results from Carreté and colleagues (Carreté et al. 2018) whose highlighted the high diversity between different *C. glabrata* isolates. Studies on virulence must take this high diversity into account, and the development of new genetic manipulation tools applicable to clinical isolates is now necessary.

Towards a return to the conventional homologous recombination?

Many efforts have been made by the scientific community working on *C. glabrata* to increase the efficacy of its homologous recombination. This is notably the case of my host laboratory, whose implemented the CRISPR-Cas9 system into *C. glabrata* (Enkler et al. 2016). The advantages of this system are undeniable, and this tool was necessary to modify the *RPR2* gene, since all previous attempts were unsuccessful. This does not prevent to take a critical look at this technology. Indeed, even if the Cas9-induced double-strand-break significantly increases the recombination efficiency, its implementation requires eventually a lot of effort. The limited availability of cleavage site in the genome and the need to sufficiently express Cas9 protein and the single guide RNA are major constraints in the application of this system. In addition, constitutive expression of the Cas9 protein appears to be toxic to *C. glabrata* (Enkler et al. 2016). CRISPR-Cas9 system appeared promising for the study of *C. glabrata* but its use is eventually disproportionate for the current studies on *C. glabrata*. Increasing the efficiency of the homologous recombination was the preliminary part, it is now necessary to move on to the second step: the generation of a mutant collection with an equivalent quality to that available in *S. cerevisiae* or even in *D. melanogaster*.

I have attempted through my thesis work to contribute to this shift. The lack of effective tools is frustrating given the high potential of *C. glabrata* (frustration I personally experienced by studying its nuclear RNase P) and the community working on *C. glabrata* should no longer waste time by constantly « reinventing the wheel ». The new method I have set up tends to go in this direction. Since the design of a new positively selectable integration cassette is not revolutionary, it may take into account for the first time the constraints allowing its scalability. I tried to create elements as short as possible, to reduce strongly the probability of point mutation inherent to PCR amplification, but also in a concern of financial cost. By combining a synthetic promoter and terminator, the size of an expression cassette can be reduced by almost 700 nt. Taking as a reference the prices charged by Integrative DNA Technologies, this size reduction would save 150€ for each complete synthesis of an expression cassette.

This kind of cassette could be the key element in the development of a large-scale protein-labelled mutant collection. Most importantly, this technique might enable us to manipulate easily clinical isolates genomes. Improvements are nevertheless necessary, such as the use of a recyclable selection marker, or the shift from antibiotic resistance marker to a dominant nutritional marker such as the *Aspergillus nidulans* acetamidase (Fu et al. 2016).

RESUMÉ EN FRANÇAIS

Introduction

Depuis la fin des années 1980 l'incidence des infections fongiques est en nette augmentation et est aujourd'hui une cause majeure de morbidité et de mortalité. Cette émergence est en corrélation avec l'augmentation du nombre de patients immunodéprimés, ainsi qu'avec l'évolution des pratiques médicales et chirurgicales, notamment l'utilisation croissante de techniques invasives et de radiothérapie.

Candida glabrata est une levure pathogène opportuniste appartenant au clade des *Nakaseomyceses*. Cette levure fut longtemps considérée comme un simple champignon saprophyte commensal pour les humains. Ce n'est qu'à la fin des années 1980 que *Candida glabrata* a été décrite comme responsable d'infection fongique systémique chez des patients immunodéprimés (Just et al. 1989). *Candida glabrata* a alors rapidement été considérée comme un pathogène émergent (Hazen 1995). Aujourd'hui, *Candida glabrata* est la deuxième cause de candidémie en Europe et en Amérique du Nord (Guinée 2014). Le taux de mortalité lié aux infections par *Candida glabrata* est l'un des plus élevés parmi les espèces de *Candida*: 40% des infections invasives par *C. glabrata* conduisent au décès du patient (Tortorano et al. 2006). Une hospitalisation en unité de soins intensifs est le principal facteur de risque d'une infection par *C. glabrata.* Les caractéristiques présentées par *C. glabrata* font davantage référence à un colonisateur opportuniste plutôt qu'à un pathogène hautement virulent. De plus, *C. glabrata* ne semble pas présenter des caractéristiques claires résultant de l'adaptation à l'humain. Contrairement à *Candida albicans*, *C. glabrata* induit très peu de dommages aux tissus colonisés. Sa stratégie d'infection réside principalement dans la furtivité et l'évasion passive des cellules phagocytaires. Néanmoins, le fort taux de mortalité induit par ce pathogène ainsi que sa resistance accrue aux antifongiques font de *C. glabrata* une menace sérieuse pour la santé publique. L'équipe « Evolution des ARN non-codants chez les levures », dirigée par le Dr. Fabrice Jossinet, s'intéresse de près aux différents mécanismes moléculaires mis en jeu par *C. glabrata* lors de l'infection. La thématique majeure du laboratoire d'accueil réside dans l'étude des particularités génomiques de *C. glabrata* et leur possible implication dans la pathogénicité de cette levure.

Objectifs des travaux de thèse

C. glabrata est considérée comme phylogénétiquement très proche de la levure modèle *Saccharomyces cerevisiae*. Pourtant, le génome de *C. glabrata* présente de nombreuses particularités telles qu'une plasticité accrue, l'émergence de gènes codants pour de nouvelles protéines et l'apparition de nouveaux domaines structuraux au sein d'ARN non-codants ubiquitaires. Ces particularités sont partagés à des niveaux différents avec toutes les espèces du clade des *Nakaseomyces*. C'est dans ce contexte que s'est inséré mon travail de thèse, qui est subdivisé en deux parties distinctes. J'ai dans un premier temps étudié le complexe ribonucléoprotéique atypique de la Ribonucléase P (RNase P) nucléaire de *C. glabrata,* en caractérisant le rôle de sa sous-unité ARN ainsi que la nature de ses partenaires protéiques (Chapitre I). Dans un second temps, j'ai amélioré les outils d'édition du génome de *C. glabrata* existants, afin de faciliter les etudes expérimentales portant sur les particularités génomique de cette levure (Chapitre II).

Chapitre I : Caractérisation de la RNase P nucléaire de *C. glabrata*

La RNase P est une endoribonucléase essentielle impliquée dans la maturation de différents ARN et notamment des ARN de transfert (ARNt). Elle est présente sous la forme d'un complexe ribonucléoprotéique composé d'un ARN non-codant, appelée RPR1 chez les levures, détenant l'activité catalytique. La présence de cette enzyme sous la forme ribonucléoprotéique est conservée dans les trois domaines du vivant. Néanmoins, la taille et la complexité de la sous-unité ARN a diminué au cours de l'évolution en faveur de l'acquisition d'un nombre accru de partenaires protéiques. Si chez la grande majorité des eucaryotes l'ARN RPR1 présente une taille de l'ordre de 400 nucléotides, chez *C. glabrata* cet ARN non-codant possède une taille surprenante de 1149 nucléotides (**Figure 1**). L'augmentation de la taille de la sous-unité ARN de la RNase P va à l'encontre de l'évolution du complexe observée chez les eucaryotes. Cette taille est due à la présence de trois domaines additionnels, P8.1, P7.1 et p7a, qui ne semblent cependant pas modifier le coeur structural conservé chez les eucaryotes. L'emergence de ces domaines pourrait donc hypothétiquement impliquer la RNase P de *C. glabrata* dans de nouvelles fonctions.

Mon travail de thèse a consisté à caractériser le rôle des domaines additionnels de l'ARN RPR1 de *C. glabrata*. Ces insertions soulèvent deux questions majeures auxquelles j'ai tenté de répondre par des approches expérimentales et bio-informatiques :

- 1) Les domaines additionnels au sein de l'ARN RPR1 sont-ils essentiels à l'activité catalytique de la RNase P de *C. glabrata* ?
- 2) La présence de ces domaines modifie-t-elle la composition de la fraction protéique RNase P ?

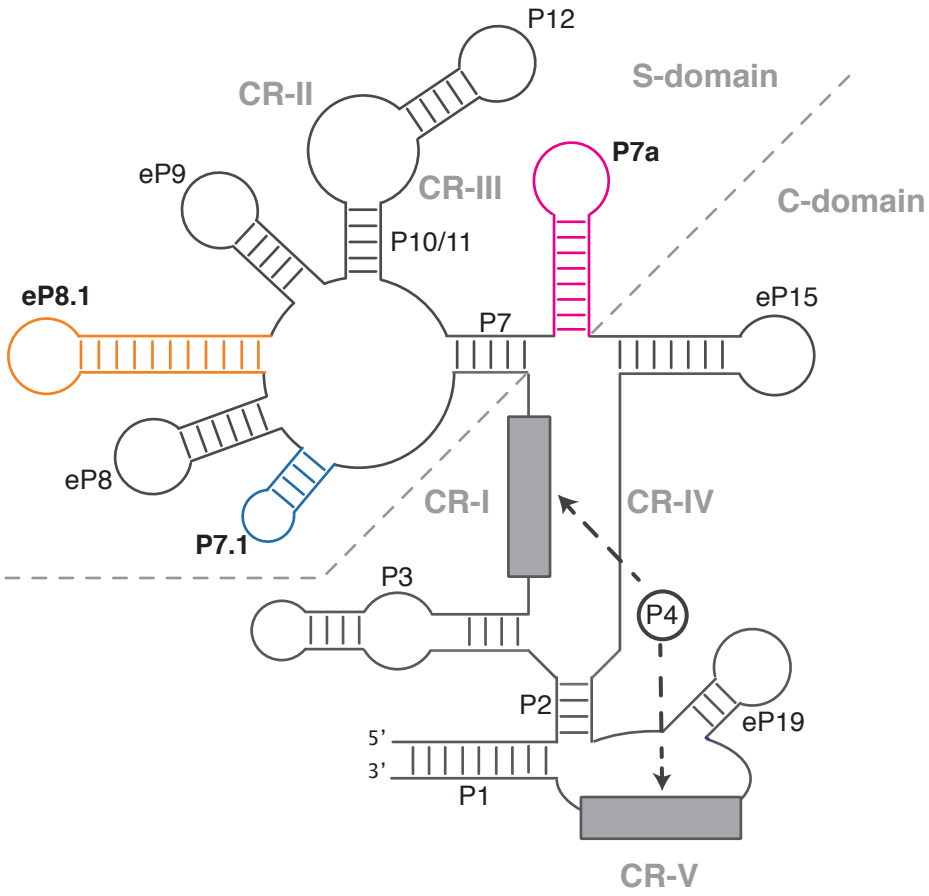

Figure 1. Structure secondaire de l'ARN RPR1 de *C. glabrata*. Cette représentation est basée sur la structure secondaire de *S. cerevisiae*. Chaque domaines additionnels (P7a, P7.1 et eP8.1) est mis en évidence par l'utilisation d'une couleur spécifique. Pour des raisons de commodité, cette vue schématique ne représente pas la taille exacte des domaines supplémentaires. Domaine **P7a** : 230 nt./ **P7.1** : 30 nt. / **eP8.**1 : 485 nt. CR : « Conserved Region » , région conservée parmi toutes les RNase P. S-Domain : Domaine de Spécificité ; C-domain : Domaine Catalytique.

1. Rôle des domaines additionnels de RPR1 dans l'activité catalytique de la RNase P de *C. glabrata.*

Les domaines additionnels pourraient simplement être impliqués dans la stabilisation structurelle. Leur présence serait donc essentielle pour l'activité catalytique de la RNase P. En revanche, si ces domaines ne fournissent que de nouvelles fonctions non canoniques à la RNase P, la suppression des domaines supplémentaires n'empêchera pas l'holoenzyme d'effectuer son activité essentielle de maturation pré-ARNt. Afin de tester cette hypothèse, la méthode de travail mise en place a consisté en l'échange de l'ARN RPR1 de *C. glabrata* par une version de RPR1 ne possédant pas de domaines additionnels. Le choix s'est porté sur l'échange du gène RPR1 de *C. glabrata* par la version de *S. cerevisiae* très proche phylogénétiquement. L'échange a été réalisé dans les deux sens par trans-complémentation fonctionnelle à savoir : un échange de l'ARN RPR1 au sein de *S. cerevisiae* par la version de *C. glabrata*; et l'échange de l'ARN RPR1 au sein de *C. glabrata* par la version de *S. cerevisiae.*

1.1 Remplacement du gène RPR1 de *S. cerevisiae* par le gene *RPR1* de *C. glabrata*

La diploïdie de *S. cerevisiae* est un avantage majeur pour l'étude des gènes essentiels, permettant ainsi la délétion de l'une des deux copies d'un gène essentiel, sans compromettre la survie de la cellule. A cet effet, le gène *RPR1* de *S. cerevisiae* a été délété au sein d'une souche diploïde de *S. cerevisiae* nommée souche S2. Pour déterminer si le gène *RPR1* de *S. cerevisiae* pouvait être échangé par la version *C. glabrata*, des tests de trans-complémentation ont été réalisés (**Figure 2**). L'analyse phénotypique des cellules obtenues a prouvé que toutes les cellules viables possédaient uniquement la version endogène de *RPR1*. Aucun événement de trans-complémentation fonctionnelle n'a donc été mis en évidence, indiquant que la levure *S. cerevisiae* n'est pas capable de survire avec le gène *RPR1* de *C. glabrata.* Ces résultats ont permis démontrer indirectement que l'ARN RPR1 de *C. glabrata* a besoin de co-facteur spécifiques pour permettre l'activité du complexe de la RNase P. Ces co-facteurs pouvant être soit totalement absent au sein de *S. cerevisiae* ; soit présents mais incapables de reconnaitre efficacement l'ARN RPR1 de *C. glabrata.*

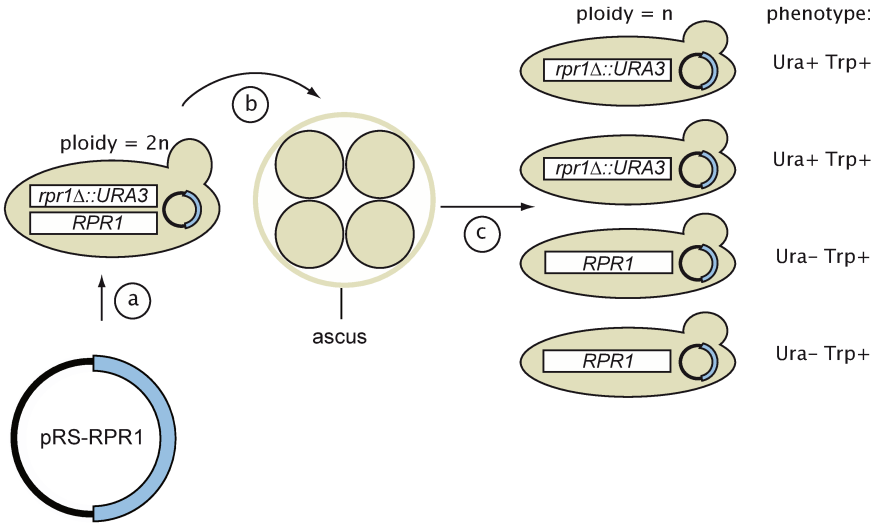

Figure 2. Trans-complémentation fonctionnelle du gène *RPR1* au sein de *S. cerevisiae* par le gène *RPR1* de *C. glabrata*.

Les étapes de la trans-complémentation fonctionnelle du gène *RPR*1 sont : a) la transformation de la souche diploïde de *S. cerevisiae* contenant 1 seule copie du gène *RPR1* ; b) l'induction de la méiose pour obtenir des cellules haploïdes ; c) l'analyse de la viabilité des cellules haploïdes obtenues. Les cellules viables sont sélectionnées pour leur phénotype. Ura+ : cellule contenant l'allèle *RPR1* de *C. glabrata*.

1.2 Remplacement du gène *RPR1* de *C. glabrata* par le gene *RPR1* de *S. cerevisiae*

Contrairement à *S. cerevisiae, C. glabrata* est strictement haploïde empêchant une délétion directe d'un gène essentiel tel que *RPR1*. L'échange des RNAse P a donc été réalisé en deux étapes. Dans un premier temps une construction permettant l'expression du gène *RPR1* de *S*. *cerevisiae* a été insérée dans le génome de *C*. *glabrata* (**Figure 3**). Cette étape a résulté en la génération d'une souche unique de *C*. *glabrata* exprimant deux gènes *RPR1* d'origines différentes. La seconde étape a consisté en la délétion de la version RPR1 endogène de *C*. *glabrata* pour permettre l'expression exclusive de la gène *RPR1* de *S*. *cerevisiae*. La cassette de disruption utilisée contenait le gène *Sh Ble,* conférant une resistance à une molécule antibiotique, la zeocine (**Figure 4**). Toutes les cellules résistantes à la zéocine obtenues résultaient d'une recombinaison homologue aspécifique, prouvant que la délétion du gène *RPR1* de *C. glabrata* n'est pas possible. Ce résultat indique donc que l'ARN RPR1 de *C*. *glabrata* n'est pas échangeable avec celui de *S*. *cerevisiae*.

Les données obtenues suite aux expériences de trans-complementation n'ont pas apporté de réponse concrète quant au rôle des domaines additionnels de RPR1. Néanmoins, ces résultats ont confirmé l'hypothèse selon laquelle des co-facteurs de l'ARN RPR1, essentiels à l'activité de la RNase P, sont strictement spécifiques à *C. glabrata.*

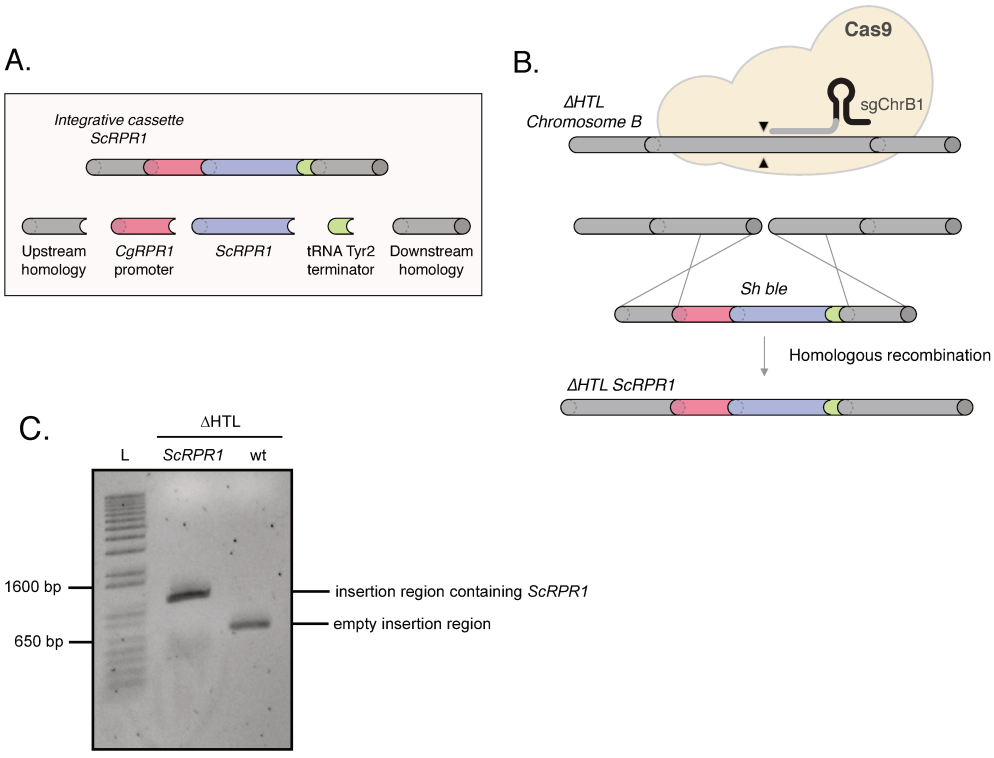

Figure 3. Integration du gène *RPR1* de *S. cerevisiae* au sein du génome de *C. glabrata.*

(**A**) La cassette intégrative est composée de bras d'homologie de 210 nucléotides, du gène *RPR1* de *S. cerevisiae* sous la dépendance du promoteur de *C. glabrata*, et du terminateur de transcription tRNATyr2. (**B**) La cassette intégrative, les plasmides permettant l'expression de la protéine Cas9 et l'ARN guide ont été co-transformés dans la souche ΔHTL. La cassette intégrative permettant l'expression du gène RPR1 de *S. cerevisiae* a été insérée dans le chromosome B de *C. glabrata* par recombinaison homologue, après une coupure double-brin induite par Cas9 au niveau du site d'insertion. (C) La séquence correspondant au site d'insertion a été amplifiée par PCR directement à partir de colonies ayant poussé sur un milieu privé de Threonine et de Leucine. La taille attendue est de 1178 pb pour un événement d'insertion, dans le cas contraire la taille attendue est de 608 pb.

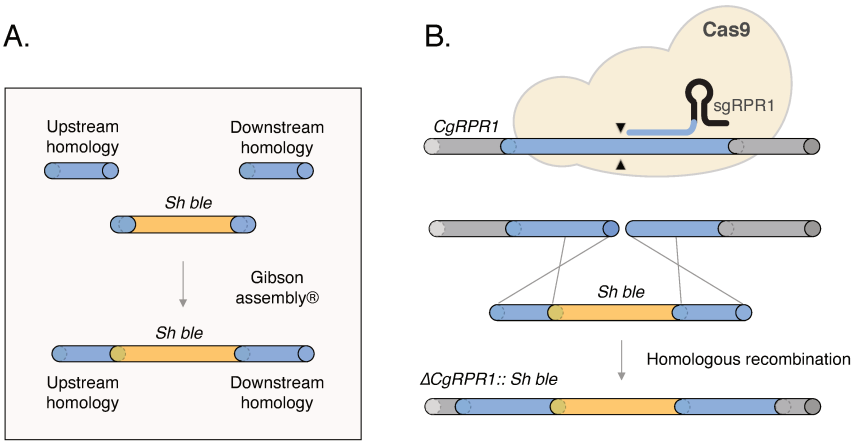

Figure 4. Inactivation du gène *RPR1* au sein de *C. glabrata*. (**A**) The DNA fragments constituting the disruption cassette of *CgRPR1* were fused by the "Gibson Assembly®" method. (**B**) The disruption cassette, the plasmids enabling the expression of the *Cas9* protein and the single guide RNA sgRPR1 were co-transformed in the ΔHTL-*ScRPR1* strain. The disruption cassette of cgRPR1 was inserted into the genome by homologous recombination, enhanced by a *Cas9* induced double-strand break at the insertion site.

1.3. Analyses bio-informatiques des protéines associées à l'ARN RPR1

Les séquences des protéines putatives du complexe de la RNase P de *C. glabrata* (Pop1, Pop3, Pop4, Pop5, Pop6, Pop7, Pop8, Rpp1, et Rpr2) ont été analysées bio-informatiquement. Les séquences de chaque protéine décrite comme interagissant avec l'ARN RPR1 de *S. cerevisiae* ont été comparées avec les homologues présents chez *C. glabrata* ainsi que chez toutes les espèces du clade des *Nakaseomyces* (**Figure 5**). De manière surprenante, la conservation de séquence et la similarité des résidus entre chaque homologue n'ont montré aucun signal fort de co-évolution des protéines avec l'ARN RPR1 chez *C. glabrata*. Ces résultats tendent à prouver que l'incapacité d'échanger les genes RPR1 entre *S. cerevisiae* et *C. glabrata* n'est visiblement pas due à une divergence des protéines du complexe, puisqu'elles présentent toutes les caractéristiques permettant une liaison aux deux versions de l'ARN RPR1.

Ces données semblent renforcer l'hypothèse selon laquelle la présence de domaines additionnels au sein de l'ARN RPR1 de *C. glabrata* pourrait influencer la composition protéique du complexe de la RNase P. Cet impact pourrait entraîner une modification de la nature et du nombre de protéines associées à l'ARN RPR1, par rapport aux complexes de levure déjà décrits (Chamberlain et al. 1998). Ainsi, le complexe RNase P de *C. glabrata* pourrait présenter de nouvelles protéines pouvant être impliquées dans : i) le maintien d'une structure fonctionnelle de l'ARN RPR1 malgré les contraintes induites par la présence de grandes insertions; ou ii) l'apport de nouvelles fonctions à la RNase P, pouvant permettre le clivage de nouveaux types d'ARN. Ces hypothèses ont été testées par une purification de l'holoenzyme RNase P de *C. glabrata* et suivie d'une analyse de la fraction protéique par spectrométrie de masse.

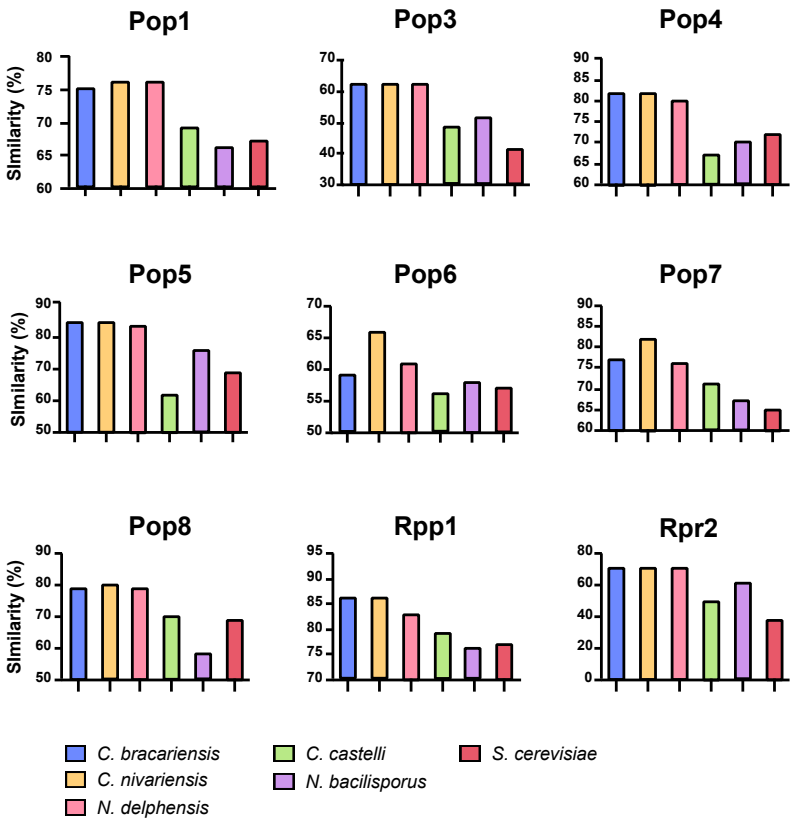

Figure 5. Similarité de séquence des proteins associées à l'ARN RPR1 entre *C. glabrata*, les espèces *Nakaseomyces* et *S. cerevisiae*. Des alignements par paires de chaque protéine ont été réalisés avec le logiciel T-COFFEE (Notredame, Higgins et al. 2000). Les taux de similarité ont été extraits des 40 alignements par paire générés et ont été représentés sous forme de diagramme.

2. Caractérisation de la composition protéique du complexe de la RNase P de *C. glabrata*.

Saccharomyces cerevisiae est l'organisme modèle le plus proche phylogénétiquement de *Candida glabrata*. Chez cette levure le complexe de la RNase P contient 9 protéines (POP1, POP3, POP4, POP5, POP6, POP7, POP8, RPP1, et RPR2) codées par des gènes dont les homologues sont retrouvés dans le génome de *C. glabrata*. En utilisant la technologie CRISPR-Cas9, mise au point dans *C. glabrata* par le laboratoire d'accueil, j'ai tout d'abord intégré directement dans le génome une étiquette HA (human influenza hemaglutinine) par recombinaison homologue à l'extrémité C terminale de la protéine Rpr2. Cette protéine a été sélectionnée car elle est exclusive à la RNase P contrairement aux autres protéines de la RNase P. Cette protéine étiquetée a permis de purifier le complexe par immunoprecipitation et d'analyser sa composition, en collaboration avec la plateforme de protéomique de l'IBMC, Strasbourg. Les huit sous-unités protéiques putatives de la RNase P ont été co-purifiées avec Rpr2. Ces données valident l'annotation des 9 gènes codant pour la protéine RNase P de *C. glabrata*. De plus, aucune protéine spécifique de la RNase MRP, à savoir Snm1 et Rmp1, n'a été purifiée, indiquant que les protéines co-purifiées avec Rpr2 sont bien spécifiquement associées à la RNase P. Ce résultat est fondamental puisqu'il a confirmé les analyses *in silico* : la présence de domaines supplémentaires au sein de RPR1 ne remplace pas la perte de partenaires protéiques. Au contraire, une nouvelle protéine a été détectée dans le complexe : la protéine Rcl1. Chez *S. cerevisiae*, cette protéine a été décrite comme étant impliquée dans la maturation de l'ARN ribosomal (Horn et al. 2011). La co-purification de la protéine Rcl1 avec Rpr2 n'a pas encore été publiée et aucune donnée d'interaction chez *S. cerevisiae* n'a encore rapporté leur interaction. La fonction de cette protéine est similaire à celle de la RNase MRP, bien qu'aucune protéine de cette enzyme n'ait été trouvée après purification. Afin de déterminer le lien entre la présence de cette protéine au sein de la RNase P de *C.* glabrata et la présence de domaines additionnels au sein de l'ARN RPR1 la séquence de la protéine Rcl1 a été analysée chez *S. cerevisiae* et toutes les espèces du clade des *Nakaseomyces* (**Figure 6**). La séquence protéique de Rcl1 est fortement conservée parmi toutes ces espèces et aucune divergence de séquence ne corrèlait avec la présence de domaines additionnels au sein de l'ARN RPR1.

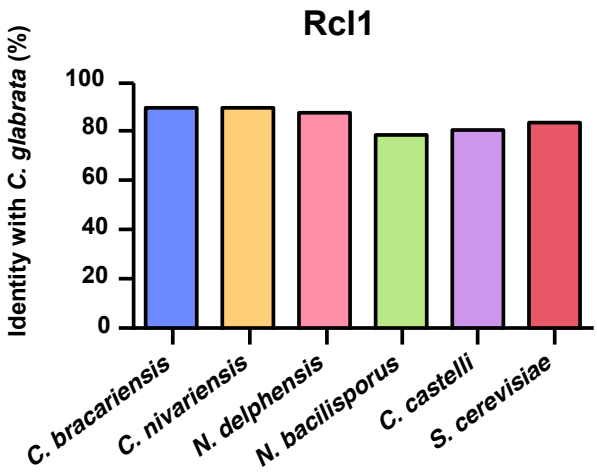

Figure 6. Identité de séquence de la protéine Rcl1 entre *C. glabrata*, les espèces du clade des *Nakaseomyces* et *S. cerevisiae*. Les alignements par paires ont été réalisés avec le logiciel T-COFFEE (Notredame, Higgins et al. 2000). Les taux d'identité strictes ont été extraits des 6 alignements par paires réalisés et ont été représentés sous forme de diagrammes.

Chapitre II : Amélioration des outils d'édition du génome de *C. glabrata*

La délétion d'un gène par intégration d'un marqueur de sélection au sein de sa séquence codante est une technique conventionnelle, couramment utilisée dans la génétique des levures. Pour *C. glabrata* cette méthode a notamment permis la création d'une collection de mutants « perte de fonction », par l'intégration de la cassette NAT (N-acétyl transférase) (Schwarzmüller et al. 2014). Les informations fournies par la délétion d'un gène ne sont cependant pas suffisantes pour comprendre sa fonction. Les gènes étudiés doivent être modifiés plus finement, par exemple en modulant leur expression ou par la génération d'une protéine chimérique. L'intégration de séquences exogènes complètes au sein d'un génome permet d'apporter de nouvelles connaissances sur la biologie d'un organisme. L'intégration de gènes entiers tels que la GFP (Green Fluorescent Protein) par exemple, est un moyen efficace pour décrypter des processus moléculaires au sein d'une cellule. Concernant les études portant sur *C. glabrata*, l'expression de gènes mutés est majoritairement apporté par des plasmides plutôt que par la modification directe de son génome. L'inconvénient principal dans l'utilisation de plasmides réside dans leur maintien au sein de la cellule transformée. Ce maintien est généralement possible par la présence au sein du plasmide d'un gène d'une voie de biosynthèse absent au sein d'une souche auxotrophe. Cependant de plus en plus de publications mettent en garde quant à l'utilisation des souches autotrophes dans l'étude des levures, pouvant apporter de nombreux biais dans l'interpretation des expériences menées. Ces problèmes ont été observés au cours de cette thèse en analysant le rôle d'un gène aux fonctions inconnues, le gène *CAGL0K07678g*, dans la virulence de *C. glabrata* au cours de l'infection systémique du modèle animal *Drosophila melanogaster.* L'inactivation de *CAGL0K07678g* dans une souche auxotrophe de *C. glabrata* a laissé pensé que ce gène était directement impliqué dans l'infection systémique d'un hôte. Cependant, la reconduite de cette expérience d' inactivation dans une souche de référence a confirmé le contraire, à savoir que l'inactivation de ce gène ne diminué pas significativement la

virulence de *C. glabrata* lors de l'infection systémique de *D. melanogaster*. La souche auxotrophe de *C. glabrata* induit donc un biais conséquent dans la caractérisation de la fonction des gènes cette levure. Dans ce contexte, j'ai conçu de nouveaux outils permettant de surmonter les limitations actuelles concernant l'édition du génome d'une souche prototrophique de *C. glabrata.* J'ai améliorer les méthodes d'édition du génome majoritairement utilisés dans l'étude de *C. glabrata* permettant i) la disruption, ii) l'étiquetage par un épitope et iii) l'intégration de potentiellement n'importe quel gène au sein du génome de *C. glabrata*. Cette méthode repose uniquement sur le système de recombinaison homologue sans induction d'une coupure double brin préalable. Dans un premier temps j'ai mis en place une méthode rapide de construction de cassette de disruption basée sur l'assemblage de bras d'homologie d'un gène ciblé avec le gène conférant la resistance à la zeocine. L'insertion de cette cassette dans une séquence codante inactive le gène d'intérêt et peut facilement être sélectionnée sur un milieu supplémenté en zeocine. Dans un second temps j'ai conçu une cassette d'étiquetage de l'extrémité C terminale des protéines par l'etiquette HA (human influenza hemaglutinine) fusionnée au gène conférant la resistance à la zeocine (**Figure 7**) . Les événements d'étiquetage d'une protéine par l'épitope HA peuvent êtres rapidement sélectionnés sur un milieu supplémenté en antibiotique. Cet epitope permettra par exemple la réalisation d'expériences d'immunoprecipitation afin d'analyser les partenaires d'une protéine donnée. Ces méthodes ont été testées sur le gène *CAGL0K07678g* , qui a pu être inactivé et étiqueté pour la première fois dans une souche sauvage de *C. glabrata*.

Sur le même principe, j'ai mis en place une méthode d'intégration de gène entier par fusion avec le gène de résistance à la zeocine. Un gène nécessite un promoteur et un terminateur de transcription afin d'être exprimé. Pour que cette méthode d'intégration de gène soit la plus pratique possible ces éléments se devaient être de petite taille. J'ai donc conçu 3 nouveaux promoteurs et un terminateur synthétiques présentant une taille de l'ordre de 4 fois inférieure à ceux utilisés couramment chez *C. glabrata* (**Figure 8 et 9**). Cette méthode d'intégration de gène a été mise au point et testée de manière positive avec la protéine fluorescente GFP (**Figure 10**) .

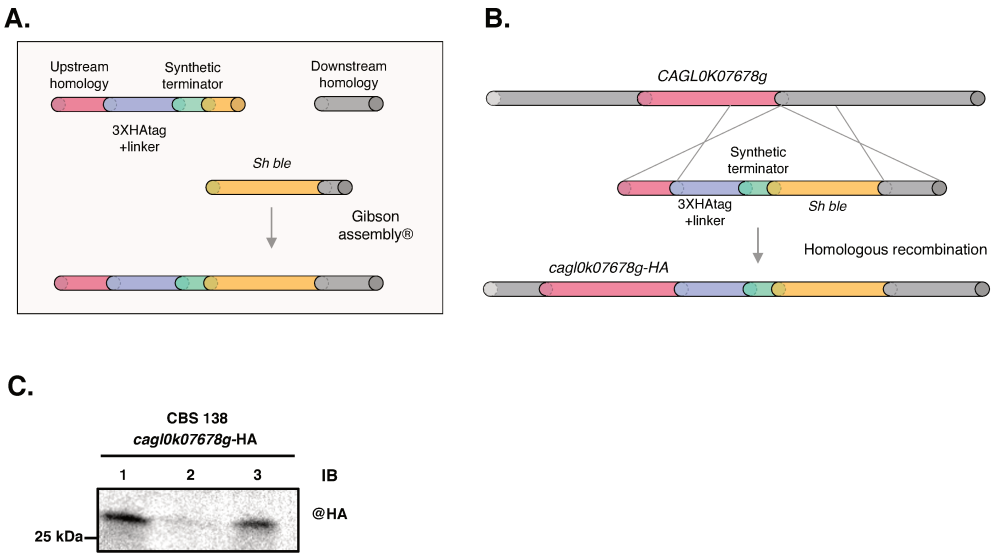

Figure 7. Etiquetage de la protéine CAGL0K07678g par un épitope HA dans la souche de référence CBS138 de *C. glabrata*. (**A**) La cassette intégrative est composée de bras d'homologie gauche et droit de respectivement 255 pb et 274 pb ; l'épitoge 3X-HA en aval d'un espaceur neutre ; le terminateur synthétique de et le marqueur de sélection *Sh ble* conférant la résistance à la zéocine. (**B**) La cassette intégrative permettant le marquage de la protéine Cagl0k07678g a été transformée dans la souche de référence CBS138 et insérée dans le domaine C-terminal par recombinaison homologue. (**C**) **Expression de la protéine de fusion cagl0K7678g-HA**. Les trois mutants indépendants cagl0K7678g-HA ont été cultivés à 30°C dans un milieu liquide riche jusqu'à la phase exponentielle de croissance, les cellules ont ensuite été centrifugées et transférées dans un milieu liquide pauvre en nitrogène pendant 6h. Après extraction des protéines totales, l'expression de la protéine cagl0K7678g-HA a été analysée par Western Blot en utilisant l'anticorps monoclonal dirigé contre l'épitoge HA.

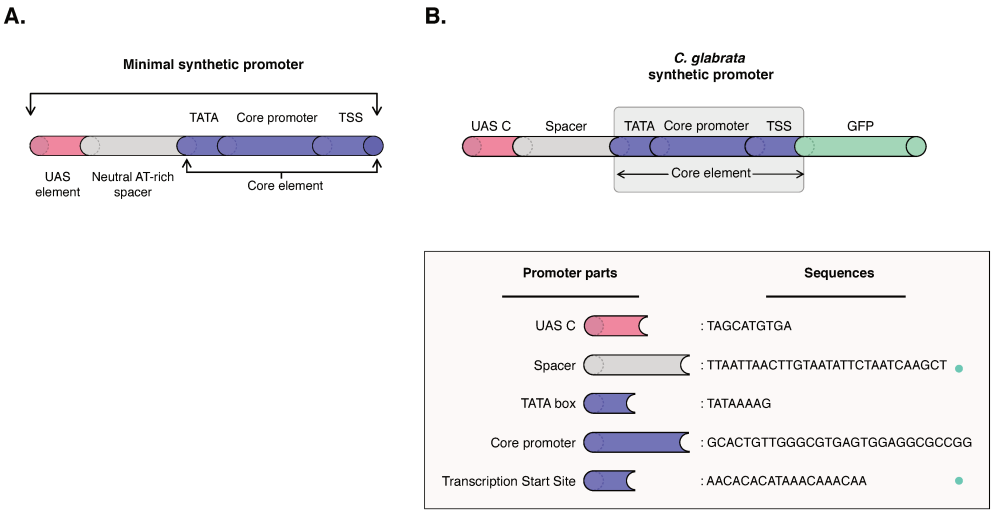

Figure 8. Développement d'un promoteur synthétique pour *C. glabrata*. (A) Les promoteurs synthétiques ont été construits selon les indications de Redden et collègues (Redden et Alper 2015). Leur promoteur était composé d'un élément d'activation synthétique en amont (UAS), d'une séquence neutre riche en A/T de 30 pb, coeur promoteur synthétique de 30 pb et d'un site de démarrage de la transcription (TSS). (B) Le premier promoteur synthétique conçu pour *C. glabrata* présentait une structure similaire. Concernant les modifications apportées pour *C. glabrata* (points verts) : l'élément UAS C a été choisi ; un espaceur riche en A/T de 30 bp et un site de départ de transcription riche en A/T (TSS) de 19 PB ont été générés aléatoirement.

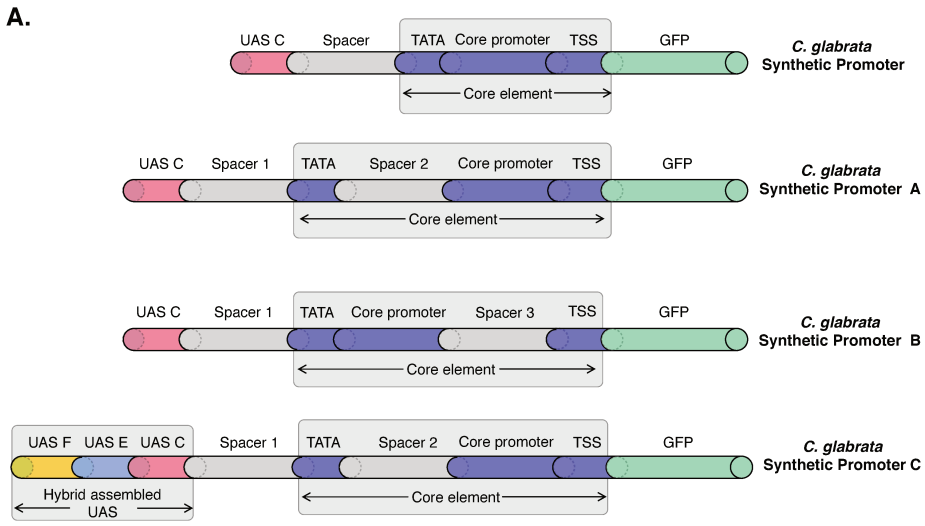

Figure 9. Amélioration du promoteur synthétique de *C. glabrata*. Le promoteur synthétique original de *C. glabrata* a été modifié pour permettre la synthèse d'une protéine fonctionnelle. **Promoteur A** : une séquence riche en A de 30 pb (Spacer 2) a été insérée en amont du promoteur du coeur promoteur pour l'éloigner de la boîte TATA. **Promoteur B** : une petite séquence de 6 pb (espaceur 3) a été insérée en amont du TSS. **Promoteur C** : sur la base du promoteur A, les éléments UAS F et UAS E ont été combinées avec l'élément UAS C.

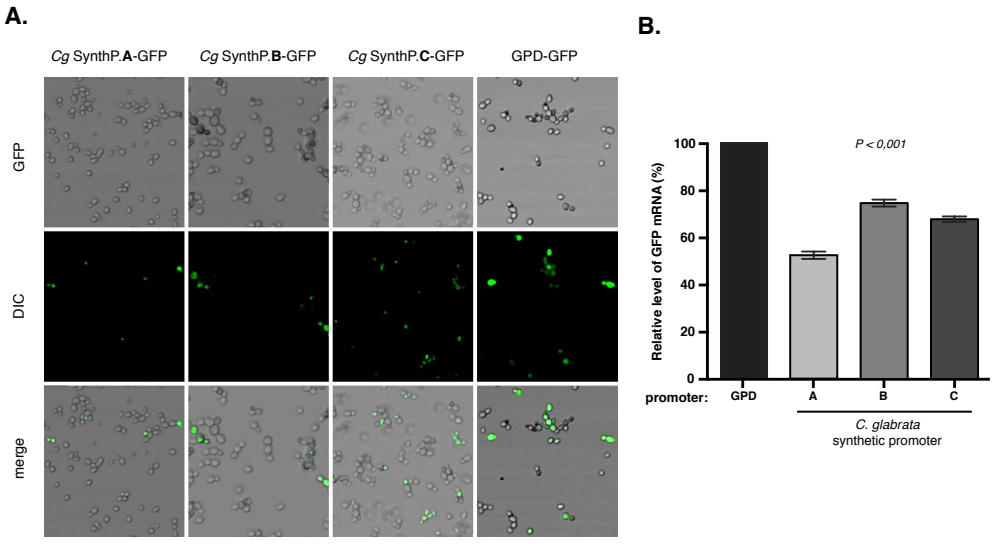

Figure 10. Les trois nouveaux promoteurs synthétiques de *C. glabrata* permettent une expression efficace de la protéine GFP. (A) L'efficacité de traduction de la protéine GFP a été visualisée par microscopie confocale (LSM 700, Zeiss). Les images ont été traitées à l'aide des logiciels ZEN (Zeiss) et ImageJ. (B) L'expression de l'ARNm GFP a été mesurée par RT-qPCR et l'analyse des résultats a été effectuée à l'aide du logiciel Prism (GraphPad).

Conclusion

Mes travaux de thèse ont dans un premier temps permis de caractériser les partenaires protéiques de la composante ARN de la RNase P nucléaire de *Candida glabrata*. Il a été mis pour la première fois en évidence la présence de la protéine Rcl1 au sein d'un complexe de RNase P nucléaire. Néanmoins nous ne savons pas à l'heure actuelle si sa présence est due aux domaines additionnels de RPR1 ou si sa détection est liée l'utilisation de techniques récentes de spectrométrie de masse peu utilisée dans la caractérisation de la RNase P d'autre espèces. De part sa fonction, la présence de cette protéine pourrait alors expliquer les résultats récemment publiés impliquant la RNase P dans des processus de maturation d'ARN ribosomiques chez les eucaryotes. Ce travail de thèse a aussi apporté des réponses préliminaires aux rôle des domaines additionnels de RPR1 qui sembleraient être indispensable à la fonction essentielle de la RNase P à savoir la maturation des ARN de transfet. Dans un second temps, mes travaux de thèse ont contribué a étoffer de façon importante la palette d'outils disponibles pour l'édition du génome de *C. glabrata*. Ces nouvelles méthodes modulables mises en place faciliteront grandement l'analyse de gènes dans des souches isolées cliniquement mais aussi la mise en place de banque de mutations à large échelle.

Bibliography

1. Abiko, Y., Y. Jinbu, T. Noguchi, M. Nishimura, K. Kusano, P. Amaratunga, T. Shibata and T. Kaku (2002). "Upregulation of human beta-defensin 2 peptide expression in oral lichen planus, leukoplakia and candidiasis. an immunohistochemical study." Pathology-Research and Practice **198**(8): 537-542.

2. Al-Yasiri, M. H., A.-C. Normand, C. L'Ollivier, L. Lachaud, N. Bourgeois, S. Rebaudet, R. Piarroux, J.-F. Mauffrey and S. Ranque (2016). "Opportunistic fungal pathogen Candida glabrata circulates between humans and yellow-legged gulls." Scientific reports **6**: 36157.

3. Alam, M. T., A. Zelezniak, M. Mülleder, P. Shliaha, R. Schwarz, F. Capuano, J. Vowinckel, E. Radmaneshfar, A. Krüger and E. Calvani (2016). "The metabolic background is a global player in Saccharomyces gene expression epistasis." Nature microbiology **1**(3): 15030.

4. Alcoba-Florez, J., S. Mendez-Alvarez, J. Cano, J. Guarro, E. Perez-Roth and M. del Pilar Arevalo (2005). "Phenotypic and molecular characterization of Candida nivariensis sp. nov., a possible new opportunistic fungus." J Clin Microbiol **43**(8): 4107-4111.

5. Alderton, A. J., I. Burr, F. A. Mühlschlegel and M. F. Tuite (2006). "Zeocin resistance as a dominant selective marker for transformation and targeted gene deletions in Candida glabrata." Mycoses **49**(6): 445-451.

6. Alexander, B. D., M. D. Johnson, C. D. Pfeiffer, C. Jiménez-Ortigosa, J. Catania, R. Booker, M. Castanheira, S. A. Messer, D. S. Perlin and M. A. Pfaller

(2013). "Increasing echinocandin resistance in Candida glabrata: clinical failure correlates with presence of FKS mutations and elevated minimum inhibitory concentrations." Clinical Infectious Diseases **56**(12): 1724-1732.

7. Alexopoulos, C. J. (1962). "Introductory mycology." Introductory mycology.

8. Alifano, P., F. Rivellini, C. Piscitelli, C. M. Arraiano, C. B. Bruni and M. S. Carlomagno (1994). "Ribonuclease E provides substrates for ribonuclease P-dependent processing of a polycistronic mRNA." Genes & development **8**(24): 3021-3031.

9. Almeida, R. S., D. Wilson and B. Hube (2009). "Candida albicans iron acquisition within the host." FEMS Yeast Res **9**(7): 1000-1012.

10. Altman, S. and L. Kirsebom (1999). "Ribonuclease P." Cold spring harbor monograph series **37**: 351-380.

11. Alves, C. T., X.-Q. Wei, S. Silva, J. Azeredo, M. Henriques and D. W. Williams (2014). "Candida albicans promotes invasion and colonisation of Candida glabrata in a reconstituted human vaginal epithelium." Journal of Infection **69**(4): 396-407.

12. Anderson, H. W. (1917). The Yeast-like Fungi of the Human Intestinal Tract, University of Illinois.

13. Aylward, J., E. T. Steenkamp, L. L. Dreyer, F. Roets, B. D. Wingfield and M. J. Wingfield (2017). "A plant pathology perspective of fungal genome sequencing." IMA Fungus **8**(1): 1-15.

14. Bairwa, G., M. Rasheed, R. Taigwal, R. Sahoo and R. Kaur (2014). "GPI (glycosylphosphatidylinositol)-linked aspartyl proteases regulate vacuole homoeostasis in Candida glabrata." Biochemical Journal **458**(2): 323-334.

15. Beebe, J. A., J. C. Kurz and C. A. Fierke (1996). "Magnesium ions are required by Bacillus subtilis ribonuclease P RNA for both binding and cleaving precursor tRNAAsp." Biochemistry **35**(32): 10493-10505.

16. Bitar, D., O. Lortholary, Y. Le Strat, J. Nicolau, B. Coignard, P. Tattevin, D. Che and F. Dromer (2014). "Population-based analysis of invasive fungal infections, France, 2001–2010." Emerging infectious diseases **20**(7): 1149.

17. Blackwell, M. (2011). "The Fungi: 1, 2, 3… 5.1 million species?" American journal of botany **98**(3): 426-438.

18. Borst, A., M. T. Raimer, D. W. Warnock, C. J. Morrison and B. A. Arthington-Skaggs (2005). "Rapid acquisition of stable azole resistance by Candida glabrata isolates obtained before the clinical introduction of fluconazole." Antimicrobial agents and chemotherapy **49**(2): 783-787.

19. Bothwell, A. L., B. C. Stark and S. Altman (1976). "Ribonuclease P substrate specificity: cleavage of a bacteriophage phi80-induced RNA." Proceedings of the National Academy of Sciences **73**(6): 1912-1916.

20. Botstein, D., S. A. Chervitz and M. Cherry (1997). "Yeast as a model organism." Science **277**(5330): 1259-1260.

21. Bradsher, R. W., S. W. Chapman and P. G. Pappas (2003). "Blastomycosis." Infectious disease clinics of North America **17**(1): 21-40, vii.

22. Brand, A., D. M. MacCallum, A. J. Brown, N. A. Gow and F. C. Odds (2004). "Ectopic expression of URA3 can influence the virulence phenotypes and proteome of Candida albicans but can be overcome by targeted reintegration of URA3 at the RPS10 locus." Eukaryot Cell **3**(4): 900-909.

23. Brown, G. D., D. W. Denning, N. A. Gow, S. M. Levitz, M. G. Netea and T. C. White (2012). "Hidden killers: human fungal infections." Sci Transl Med **4**(165): 165rv113.

24. Brown, J. W. (1998). "The ribonuclease P database." Nucleic acids research **26**(1): 351-352.

25. Brunke, S., J. Quintin, L. Kasper, I. D. Jacobsen, M. E. Richter, E. Hiller, T. Schwarzmüller, C. d'Enfert, K. Kuchler and S. Rupp (2015). "Of mice, flies–and men? Comparing fungal infection models for large-scale screening efforts." Disease models & mechanisms: dmm. 019901.

26. Byrnes Iii, E. J., W. Li, Y. Lewit, H. Ma, K. Voelz, P. Ren, D. A. Carter, V. Chaturvedi, R. J. Bildfell and R. C. May (2010). "Emergence and pathogenicity of highly virulent Cryptococcus gattii genotypes in the northwest United States." PLoS Pathog **6**(4): e1000850.

27. Carreté, L., E. Ksiezopolska, C. Pegueroles, E. Gómez-Molero, E. Saus, S. Iraola-Guzmán, D. Loska, O. Bader, C. Fairhead and T. Gabaldón (2018). "Patterns of genomic variation in the opportunistic pathogen Candida glabrata suggest the existence of mating and a secondary association with humans." Current Biology **28**(1): 15-27. e17.

28. Castaño, I., S. J. Pan, M. Zupancic, C. Hennequin, B. Dujon and B. P. Cormack (2005). "Telomere length control and transcriptional regulation of subtelomeric adhesins in Candida glabrata." Molecular microbiology **55**(4): 1246-1258.

29. Chamberlain, J. R., Y. Lee, W. S. Lane and D. R. Engelke (1998). "Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP." Genes & development **12**(11): 1678-1690.

30. Charles, P. E., J. M. Doise, J. P. Quenot, H. Aube, F. Dalle, P. Chavanet, N. Milesi, L. S. Aho, H. Portier and B. Blettery (2003). "Candidemia in critically ill patients: difference of outcome between medical and surgical patients." Intensive care medicine **29**(12): 2162-2169.

31. Chen, J.-L. and N. R. Pace (1997). "Identification of the universally conserved core of ribonuclease P RNA." Rna **3**(6): 557.

32. Colombo, A. L., M. Garnica, L. F. Aranha Camargo, C. A. Da Cunha, A. C. Bandeira, D. Borghi, T. Campos, A. L. Senna, M. E. Valias Didier and V. C. Dias (2013). "Candida glabrata: an emerging pathogen in Brazilian tertiary care hospitals." Med Mycol **51**(1): 38-44.

33. Cormack, B. P. and S. Falkow (1999). "Efficient homologous and illegitimate recombination in the opportunistic yeast pathogen Candida glabrata." Genetics **151**(3): 979-987.

34. Cormack, B. P., N. Ghori and S. Falkow (1999). "An adhesin of the yeast pathogen Candida glabrata mediating adherence to human epithelial cells." Science **285**(5427): 578-582.

35. Cornely, O. A., M. Bassetti, T. Calandra, J. Garbino, B. J. Kullberg, O. Lortholary, W. Meersseman, M. Akova, M. C. Arendrup and S. Arikan-Akdagli (2012). "ESCMID* guideline for the diagnosis and management of Candida diseases 2012: non-neutropenic adult patients." Clinical Microbiology and Infection **18**: 19-37.

36. Correia, A., P. Sampaio, S. James and C. Pais (2006). "Candida bracarensis sp. nov., a novel anamorphic yeast species phenotypically similar to Candida glabrata." Int J Syst Evol Microbiol **56**(Pt 1): 313-317.

37. Crary, S. M., S. Niranjanakumari and C. A. Fierke (1998). "The Protein Component of Bacillus s ubtilis Ribonuclease P Increases Catalytic Efficiency by Enhancing Interactions with the 5 'Leader Sequence of Pre-tRNAAsp." Biochemistry **37**(26): 9409-9416.

38. Cuéllar-Cruz, M., M. Briones-Martin-del-Campo, I. Cañas-Villamar, J. Montalvo-Arredondo, L. Riego-Ruiz, I. Castaño and A. De Las Penas (2008). "High resistance to oxidative stress in the fungal pathogen Candida glabrata is mediated by a single catalase, Cta1p, and is controlled by the transcription factors Yap1p, Skn7p, Msn2p, and Msn4p." Eukaryot Cell **7**(5): 814-825.

39. de Groot, P. W., E. A. Kraneveld, Q. Y. Yin, H. L. Dekker, U. Gross, W. Crielaard, C. G. de Koster, O. Bader, F. M. Klis and M. Weig (2008). "The cell wall of the human pathogen Candida glabrata: differential incorporation of novel adhesin-like wall proteins." Eukaryot Cell **7**(11): 1951-1964.

40. Delsuc, F., H. Brinkmann and H. Philippe (2005). "Phylogenomics and the reconstruction of the tree of life." Nature Reviews Genetics **6**(5): 361.

41. Dollive, S., Y.-Y. Chen, S. Grunberg, K. Bittinger, C. Hoffmann, L. Vandivier, C. Cuff, J. D. Lewis, G. D. Wu and F. D. Bushman (2013). "Fungi of the murine gut: episodic variation and proliferation during antibiotic treatment." PLoS One **8**(8): e71806.

42. Donlan, R. M. and J. W. Costerton (2002). "Biofilms: survival mechanisms of clinically relevant microorganisms." Clinical microbiology reviews **15**(2): 167-193.

43. Doudna, J. A. and E. Charpentier (2014). "The new frontier of genome engineering with CRISPR-Cas9." Science **346**(6213): 1258096.

44. Douglas, C. M., J. A. D'Ippolito, G. J. Shei, M. Meinz, J. Onishi, J. A. Marrinan, W. Li, G. K. Abruzzo, A. Flattery and K. Bartizal (1997). "Identification of the FKS1 gene of Candida albicans as the essential target of 1, 3-beta-D-glucan synthase inhibitors." Antimicrobial agents and chemotherapy **41**(11): 2471-2479.

45. Drummond, R. A., S. L. Gaffen, A. G. Hise and G. D. Brown (2014). "Innate defense against fungal pathogens." Cold Spring Harb Perspect Med: a019620.

46. Dujon, B., D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. De Montigny, C. Marck, C. Neuvéglise and E. Talla (2004). "Genome evolution in yeasts." Nature **430**(6995): 35.

47. Edskes, H. K. and R. B. Wickner (2013). "The [URE3] prion in Candida." Eukaryot Cell: EC. 00015-00013.

48. Ellis, J. C. and J. W. Brown (2010). The evolution of RNase P and its RNA. Ribonuclease P, Springer**:** 17-40.

49. Enkler, L., D. Richer, A. L. Marchand, D. Ferrandon and F. Jossinet (2016). "Genome engineering in the yeast pathogen Candida glabrata using the CRISPR-Cas9 system." Scientific reports **6**: 35766.

50. Esakova, O. and A. S. Krasilnikov (2010). "Of proteins and RNA: the RNase P/ MRP family." Rna.

51. Evans, D., S. M. Marquez and N. R. Pace (2006). "RNase P: interface of the RNA and protein worlds." Trends in biochemical sciences **31**(6): 333-341.

52. Ferrari, S., F. Ischer, D. Calabrese, B. Posteraro, M. Sanguinetti, G. Fadda, B. Rohde, C. Bauser, O. Bader and D. Sanglard (2009). "Gain of function mutations in CgPDR1 of Candida glabrata not only mediate antifungal resistance but also enhance virulence." PLoS Pathog **5**(1): e1000268.

53. Ferrari, S., M. Sanguinetti, R. Torelli, B. Posteraro and D. Sanglard (2011). "Contribution of CgPDR1-regulated genes in enhanced virulence of azole-resistant Candida glabrata." PLoS One **6**(3): e17589.

54. Filler, S. G. and B. J. Kullberg (2002). "Deep-seated candidal infections." In: Calderone R.A., editor. Candida and Candidiasis. ASM Press; Washington, DC: 341-348.

55. Fishman, J. A. (2017). "Infection in organ transplantation." American Journal of Transplantation **17**(4): 856-879.

56. Flannagan, R. S., G. Cosio and S. Grinstein (2009). "Antimicrobial mechanisms of phagocytes and bacterial evasion strategies." Nat Rev Microbiol **7**(5): 355-366.

57. France, M. M. and J. R. Turner (2017). "The mucosal barrier at a glance." J Cell Sci **130**(2): 307-314.

58. Fu, J., M. Blaylock, C. F. Wickes, W. Welte, A. Mehrtash, N. Wiederhold and B. L. Wickes (2016). "Development of a Candida glabrata dominant nutritional transformation marker utilizing the Aspergillus nidulans acetamidase gene (amdS)." FEMS Yeast Res **16**(3).

59. Gabaldon, T. and L. Carrete (2016). "The birth of a deadly yeast: tracing the evolutionary emergence of virulence traits in Candida glabrata." FEMS Yeast Res **16**(2): fov110.

60. Gabaldón, T., T. Martin, M. Marcet-Houben, P. Durrens, M. Bolotin-Fukuhara, O. Lespinet, S. Arnaise, S. Boisnard, G. Aguileta and R. Atanasova (2013). "Comparative genomics of emerging pathogens in the Candida glabrata clade." BMC genomics **14**(1): 623.

61. Galluzzi, L., A. Buque, O. Kepp, L. Zitvogel and G. Kroemer (2015). "Immunological effects of conventional chemotherapy and targeted anticancer agents." Cancer cell **28**(6): 690-714.

62. Gerwien, F., A. Safyan, S. Wisgott, S. Brunke, L. Kasper and B. Hube (2017). "The Fungal Pathogen Candida glabrata Does Not Depend on Surface Ferric Reductases for Iron Acquisition." Front Microbiol **8**: 1055.

63. Ghannoum, M. A. and L. B. Rice (1999). "Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance." Clinical microbiology reviews **12**(4): 501-517.

64. Giegé, R., C. Florentz and T. W. Dreher (1993). "The TYMV tRNA-like structure." Biochimie **75**(7): 569-582.

65. Gobert, A., B. Gutmann, A. Taschner, M. Gößringer, J. Holzmann, R. K. Hartmann, W. Rossmanith and P. Giegé (2010). "A single Arabidopsis organellar protein has RNase P activity." Nature structural & molecular biology **17**(6): 740.

66. Gopalan, V., N. Jarrous and A. S. Krasilnikov (2018). "Chance and necessity in the evolution of RNase P." Rna **24**(1): 1-5.

67. Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace and S. Altman (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." Cell **35**(3): 849-857.

68. Guinea, J. (2014). "Global trends in the distribution of Candida species causing candidemia." Clin Microbiol Infect **20 Suppl 6**: 5-10.

69. Guo, Z. and F. Sherman (1996). "Signals sufficient for 3'-end formation of yeast mRNA." Molecular and cellular biology **16**(6): 2772-2776.

70. Gupta, A., A. Gupta and A. Varma (2015). "Candida glabrata candidemia: An emerging threat in critically ill patients." Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine **19**(3): 151.

71. Gupta, A. K., R. Batra, R. Bluhm, T. Boekhout and T. L. Dawson Jr (2004). "Skin diseases associated with Malassezia species." Journal of the American Academy of Dermatology **51**(5): 785-798.

72. Gutmann, B., A. Gobert and P. Giegé (2012). "PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis." Genes & development.

73. Haas, E. S. and J. W. Brown (1998). "Evolutionary variation in bacterial RNase P RNAs." Nucleic acids research **26**(18): 4093-4099.

74. Harju, S., H. Fedosyuk and K. R. Peterson (2004). "Rapid isolation of yeast genomic DNA: Bust n'Grab." BMC biotechnology **4**(1): 8.

75. Hartmann, R. K., J. Heinrich, J. Schlegl and H. Schuster (1995). "Precursor of C4 antisense RNA of bacteriophages P1 and P7 is a substrate for RNase P of Escherichia coli." Proceedings of the National Academy of Sciences **92**(13): 5822-5826.

76. Hazen, K. C. (1995). "New and emerging yeast pathogens." Clinical microbiology reviews **8**(4): 462-478.

77. Helgason, T., I. J. Watson and J. P. W. Young (2003). "Phylogeny of the Glomerales and Diversisporales (Fungi: Glomeromycota) from actin and elongation factor 1-alpha sequences." FEMS Microbiology Letters **229**(1): 127-132.

78. Hipp, K., K. Galani, C. Batisse, S. Prinz and B. Böttcher (2011). "Modular architecture of eukaryotic RNase P and RNase MRP revealed by electron microscopy." Nucleic acids research **40**(7): 3275-3288.

79. Hittinger, C. T., J. L. Steele and D. S. Ryder (2018). "Diverse yeasts for diverse fermented beverages and foods." Current opinion in biotechnology **49**: 199-206.

80. Holzmann, J., P. Frank, E. Löffler, K. L. Bennett, C. Gerner and W. Rossmanith (2008). "RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme." Cell **135**(3): 462-474.

81. Hood, M. I. and E. P. Skaar (2012). "Nutritional immunity: transition metals at the pathogen–host interface." Nature Reviews Microbiology **10**(8): 525.

82. Horn, D. M., S. L. Mason and K. Karbstein (2011). "RCL1, a novel nuclease for 18S rRNA production." Journal of Biological Chemistry: jbc. M111. 268649.

83. Houser-Scott, F., S. Xiao, C. E. Millikin, J. M. Zengel, L. Lindahl and D. R. Engelke (2002). "Interactions among the protein and RNA subunits of Saccharomyces cerevisiae nuclear RNase P." Proceedings of the National Academy of Sciences **99**(5): 2684-2689.

84. Howard, M. J., X. Liu, W. H. Lim, B. P. Klemm, C. A. Fierke, M. Koutmos and D. R. Engelke (2013). "RNase P enzymes: Divergent scaffolds for a conserved biological reaction." RNA biology **10**(6): 909-914.

85. Isa-Isa, R., R. Arenas, R. F. Fernández and M. Isa (2012). "Rhinofacial conidiobolomycosis (entomophthoramycosis)." Clinics in dermatology **30**(4): 409-412.

86. Istel, F., T. Schwarzmüller, M. Tscherner and K. Kuchler (2015). "Genetic transformation of Candida glabrata by electroporation." Bio-protocol **5**(14).

87. Izumida, F. E., E. B. Moffa, C. E. Vergani, A. L. Machado, J. H. Jorge and E. T. Giampaolo (2014). "In vitro evaluation of adherence of Candida albicans, Candida

glabrata, and Streptococcus mutans to an acrylic resin modified by experimental coatings." Biofouling **30**(5): 525-533.

88. Jacobsen, I., S. Brunke, K. Seider, T. Schwarzmüller, A. Firon, C. d'Enfert, K. Kuchler and B. Hube (2010). "Candida glabrata persistence in mice does not depend on host immunosuppression and is unaffected by fungal amino acid auxotrophy." Infection and immunity **78**(3): 1066-1077.

89. Jacobsen, I. D., K. Grosse, A. Berndt and B. Hube (2011). "Pathogenesis of Candida albicans infections in the alternative chorio-allantoic membrane chicken embryo model resembles systemic murine infections." PLoS One **6**(5): e19741.

90. Jarrous, N. (2017). "Roles of RNase P and its subunits." Trends in Genetics **33**(9): 594-603.

91. Jarrous, N. and V. Gopalan (2010). "Archaeal/eukaryal RNase P: subunits, functions and RNA diversification." Nucleic acids research **38**(22): 7885-7894.

92. Jayatilake, J., Y. H. Samaranayake, L. K. Cheung and L. P. Samaranayake (2006). "Quantitative evaluation of tissue invasion by wild type, hyphal and SAP mutants of Candida albicans, and non−albicans Candida species in reconstituted human oral epithelium." Journal of oral pathology & medicine **35**(8): 484-491.

93. Johnson, M. D. and J. R. Perfect (2003). "Caspofungin: first approved agent in a new class of antifungals." Expert opinion on pharmacotherapy **4**(5): 807-823.

94. Jung, Y. H. and Y. Lee (1995). "RNases in ColE1 DNA metabolism." Molecular biology reports **22**(2-3): 195-200.

95. Just, G., D. Steinheimer, M. Schnellbach, C. Bottinger, E. B. Helm and W. Stille (1989). "Change of causative organisms under antifungal treatment in immunosuppressed patients with HIV-infections." Mycoses **32 Suppl 2**: 47-51.

96. Kachouri, R., V. Stribinskis, Y. Zhu, K. S. Ramos, E. Westhof and Y. Li (2005). "A surprisingly large RNase P RNA in Candida glabrata." Rna **11**(7): 1064-1072.

97. Katiyar, S. K., A. Alastruey-Izquierdo, K. R. Healey, M. E. Johnson, D. S. Perlin and T. D. Edlind (2012). "Fks1 and Fks2 are functionally redundant but differentially regulated in Candida glabrata: implications for echinocandin resistance." Antimicrobial agents and chemotherapy: AAC. 00813-00812.

98. Kauffman, C. A. and T. T. Yoshikawa (2001). "Fungal infections in older adults." Clinical Infectious Diseases **33**(4): 550-555.

99. Kazakov, S. and S. Altman (1991). "Site-specific cleavage by metal ion cofactors and inhibitors of M1 RNA, the catalytic subunit of RNase P from Escherichia coli." Proceedings of the National Academy of Sciences **88**(20): 9193-9197.

100. Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nature protocols **10**(6): 845.

101. Kennedy, W. A., C. Laurier, D. Gautrin, H. Ghezzo, M. Paré, J.-L. Malo and A.-P. Contandriopoulos (2000). "Occurrence and risk factors of oral candidiasis treated with oral antifungals in seniors using inhaled steroids." Journal of clinical epidemiology **53**(7): 696-701.

102. Kinchen, J. M. and K. S. Ravichandran (2008). "Phagosome maturation: going through the acid test." Nature reviews Molecular cell biology **9**(10): 781.

103. Kirk, P. M., P. F. Cannon, D. W. Minter and J. A. Stalpers (2008). "Dictionary of the Fungi CABI." Wallingford, UK **396**.

104. Kirsebom, L. A. (2007). "RNase P RNA mediated cleavage: substrate recognition and catalysis." Biochimie **89**(10): 1183-1194.

105. Klemm, B. P., N. Wu, Y. Chen, X. Liu, K. J. Kaitany, M. J. Howard and C. A. Fierke (2016). "The diversity of ribonuclease P: Protein and RNA catalysts with analogous biological functions." Biomolecules **6**(2): 27.

106. KoMine, Y., M. Kitabatake, T. Yokogawa, K. Nishikawa and H. Inokuchi (1994). "A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli." Proceedings of the National Academy of Sciences **91**(20): 9223-9227.

107. Kontoyiannis, D. P., K. A. Marr, B. J. Park, B. D. Alexander, E. J. Anaissie, T. J. Walsh, J. Ito, D. R. Andes, J. W. Baddley and J. M. Brown (2010). "Prospective surveillance for invasive fungal infections in hematopoietic stem cell transplant recipients, 2001–2006: overview of the Transplant-Associated Infection Surveillance Network (TRANSNET) Database." Clinical Infectious Diseases **50**(8): 1091-1100.

108. Koutmou, K. S., J. Hsieh and C. A. Fierke (2010). Kinetic mechanism of bacterial RNase P. Ribonuclease P, Springer**:** 93-111.

109. Krcmery, V. and A. J. Barnes (2002). "Non-albicans Candida spp. causing fungaemia: pathogenicity and antifungal resistance." Journal of Hospital Infection **50**(4): 243-260.

110. Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling and T. R. Cech (1982). "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena." Cell **31**(1): 147-157.

111. Kurtzman, C. P. (2003). "Phylogenetic circumscription of Saccharomyces, Kluyveromyces and other members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces, Naumovia, Vanderwaltozyma and Zygotorulaspora." FEMS Yeast Res **4**(3): 233-245.

112. Kurtzman, C. P., J. W. Fell and T. Boekhout (2011). "Definition, Classification and Nomenclature of the Yeasts." The Yeasts, a Taxonomic Study (fifth edition): 3-5.

113. Kurtzman, C. P. and C. J. Robnett (1998). "Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences." Antonie van Leeuwenhoek **73**(4): 331-371.

114. Lemke, A., A. F. Kiderlen and O. Kayser (2005). "amphotericin B." Applied microbiology and biotechnology **68**(2): 151-162.

115. Li, L. and A. Dongari-Bagtzoglou (2007). "Oral epithelium-Candida glabrata interactions in vitro." Oral Microbiol Immunol **22**(3): 182-187.

116. Li, X., D. N. Frank, N. Pace, J. M. Zengel and L. Lindahl (2002). "Phylogenetic analysis of the structure of RNase MRP RNA in yeasts." Rna **8**(6): 740-751.

117. Linde, J., S. Duggan, M. Weber, F. Horn, P. Sieber, D. Hellwig, K. Riege, M. Marz, R. Martin and R. Guthke (2015). "Defining the transcriptomic landscape of Candida glabrata by RNA-Seq." Nucleic acids research **43**(3): 1392-1406.

118. Lodder, J. and N. F. De Vries (1938). "Some notes on Torulopsis glabrata (Anderson) nov. comb." Mycopathologia **1**(2): 98-103.

119. Lorenz, R., S. H. Bernhart, C. H. Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler and I. L. Hofacker (2011). "ViennaRNA Package 2.0." Algorithms for Molecular Biology **6**(1): 26.

120. Luo, G., L. P. Samaranayake, B. P. K. Cheung and G. Tang (2004). "Reverse transcriptase polymerase chain reaction (RT−PCR) detection of HLP gene expression in Candida glabrata and its possible role in in vitro haemolysin production." Apmis **112**(4−5): 283-290.

121. Marques, S. A. (2012). "Paracoccidioidomycosis." Clinics in dermatology **30**(6): 610-615.

122. Marquez, S. M., J. L. Chen, D. Evans and N. R. Pace (2006). "Structure and function of eukaryotic Ribonuclease P RNA." Molecular cell **24**(3): 445-456.

123. Marvin, M. C., S. Clauder-Münster, S. C. Walker, A. Sarkeshik, J. R. Yates, L. M. Steinmetz and D. R. Engelke (2011). "Accumulation of noncoding RNA due to an RNase P defect in Saccharomyces cerevisiae." Rna.

124. McKinnon, P. S., D. A. Goff, J. W. Kern, J. W. Devlin, J. F. Barletta, S. J. Sierawski, A. C. Mosenthal, P. Gore, A. J. Ambegaonkar and T. J. Lubowski (2001).

"Temporal assessment of Candida risk factors in the surgical intensive care unit." Archives of Surgery **136**(12): 1401-1408.

125. Moens, E. and M. Veldhoen (2012). "Epithelial barrier biology: good fences make good neighbours." Immunology **135**(1): 1-8.

126. Muller, H., A. Thierry, J.-Y. Coppée, C. Gouyette, C. Hennequin, O. Sismeiro, E. Talla, B. Dujon and C. Fairhead (2009). "Genomic polymorphism in the population of Candida glabrata: gene copy-number variation and chromosomal translocations." Fungal Genetics and Biology **46**(3): 264-276.

127. Nagi, M., K. Tanabe, H. Nakayama, K. Ueno, S. Yamagoe, T. Umeyama, H. Ohno and Y. Miyazaki (2016). "Iron-depletion promotes mitophagy to maintain mitochondrial integrity in pathogenic yeast Candida glabrata." Autophagy **12**(8): 1259-1271.

128. Netea, M. G., G. D. Brown, B. J. Kullberg and N. A. R. Gow (2008). "An integrated model of the recognition of Candida albicans by the innate immune system." Nature Reviews Microbiology **6**(1): 67.

129. Netea, M. G., J. W. M. Van der Meer and B. J. Kullberg (2007). Recognition of fungal pathogens by Toll-like receptors. Immunology of Fungal Infections, Springer**:** 259-272.

130. Newhart, A., S. L. Powers, P. K. Shastrula, I. Sierra, L. M. Joo, J. E. Hayden, A. R. Cohen and S. M. Janicki (2016). "RNase P protein subunit Rpp29 represses histone H3. 3 nucleosome deposition." Molecular biology of the cell **27**(7): 1154-1169.

131. Notredame, C., D. G. Higgins and J. Heringa (2000). "T-coffee: a novel method for fast and accurate multiple sequence alignment1." Journal of molecular biology **302**(1): 205-217.

132. Orkwis, B. R., D. L. Davies, C. L. Kerwin, D. Sanglard and D. D. Wykoff (2010). "Novel acid phosphatase in Candida glabrata suggests selective pressure and niche specialization in the phosphate signal transduction pathway." Genetics.

133. Otto, V. and D. H. Howard (1976). "Further Studies on the Intracellular Behavior of Torulopsis glabrata." Infection and immunity **14**(2): 433-438.

134. Pappas, P. G., C. A. Kauffman, D. R. Andes, C. J. Clancy, K. A. Marr, L. Ostrosky-Zeichner, A. C. Reboli, M. G. Schuster, J. A. Vazquez and T. J. Walsh (2015). "Clinical practice guideline for the management of candidiasis: 2016 update by the Infectious Diseases Society of America." Clinical Infectious Diseases **62**(4): e1-e50.

135. Pearson, W. R. (2013). "An introduction to sequence similarity ("homology") searching." Current protocols in bioinformatics **42**(1): 3.1. 1-3.1. 8.

136. Peck-Miller, K. A. and S. Altman (1991). "Kinetics of the processing of the precursor to 4· 5 S RNA, a naturally occurring substrate for RNase P from Escherichia coli." Journal of molecular biology **221**(1): 1-5.

137. Perederina, A., I. Berezin and A. S. Krasilnikov (2018). "In vitro reconstitution and analysis of eukaryotic RNase P RNPs." Nucleic acids research.

138. Perederina, A., O. Esakova, H. Koc, M. E. Schmitt and A. S. Krasilnikov (2007). "Specific binding of a Pop6/Pop7 heterodimer to the P3 stem of the yeast RNase MRP and RNase P RNAs." Rna **13**(10): 1648-1655.

139. Perederina, A., O. Esakova, C. Quan, E. Khanova and A. S. Krasilnikov (2010). "Eukaryotic ribonucleases P/MRP: the crystal structure of the P3 domain." The EMBO journal **29**(4): 761-769.

140. Perez-Torrado, R., S. Llopis, L. Jespersen, T. Fernandez-Espinar and A. Querol (2012). "Clinical Saccharomyces cerevisiae isolates cannot cross the epithelial barrier in vitro." Int J Food Microbiol **157**(1): 59-64.

141. Pfaller, M. A., M. Castanheira, S. R. Lockhart, A. M. Ahlquist, S. A. Messer and R. N. Jones (2012). "Frequency of decreased susceptibility and resistance to echinocandins among fluconazole-resistant bloodstream isolates of Candida glabrata: results from the SENTRY Antimicrobial Surveillance Program (2006-2010) and the Centers for Disease Control and Prevention Population-Based Surveillance (2008-2010)." J Clin Microbiol: JCM. 06112-06111.

142. Pfaller, M. A., D. J. Diekema, D. L. Gibbs, V. A. Newell, D. Ellis, V. Tullio, A. Rodloff, W. Fu, T. A. Ling and G. Global Antifungal Surveillance (2010). "Results from the ARTEMIS DISK Global Antifungal Surveillance Study, 1997 to 2007: a 10.5-year analysis of susceptibilities of Candida Species to fluconazole and voriconazole as determined by CLSI standardized disk diffusion." J Clin Microbiol **48**(4): 1366-1377.

143. Phan, Q. T., P. H. Belanger and S. G. Filler (2000). "Role of Hyphal Formation in Interactions ofCandida albicans with Endothelial Cells." Infection and immunity **68**(6): 3485-3490.

144. Poláková, S., C. Blume, J. Á. Zárate, M. Mentel, D. Jørck-Ramberg, J. Stenderup and J. Piškur (2009). "Formation of new chromosomes as a virulence

mechanism in yeast Candida glabrata." Proceedings of the National Academy of Sciences **106**(8): 2688-2693.

145. Pound, M. W., M. L. Townsend, V. Dimondi, D. Wilson and R. H. Drew (2011). "Overview of treatment options for invasive fungal infections." Med Mycol **49**(6): 561-580.

146. Rai, M. N., S. Balusu, N. Gorityala, L. Dandu and R. Kaur (2012). "Functional genomic analysis of Candida glabrata-macrophage interaction: role of chromatin remodeling in virulence." PLoS Pathog **8**(8): e1002863.

147. Ramage, G., S. P. Saville, D. P. Thomas and J. L. Lopez-Ribot (2005). "Candida biofilms: an update." Eukaryot Cell **4**(4): 633-638.

148. Ramirez-Ortiz, Z. G. and T. K. Means (2012). "The role of dendritic cells in the innate recognition of pathogenic fungi (A. fumigatus, C. neoformans and C. albicans)." Virulence **3**(7): 635-646.

149. Redden, H. and H. S. Alper (2015). "The development and characterization of synthetic minimal yeast promoters." Nature communications **6**: 7810.

150. Rehner, S. A. and E. Buckley (2005). "A Beauveria phylogeny inferred from nuclear ITS and EF1-α sequences: evidence for cryptic diversification and links to Cordyceps teleomorphs." Mycologia **97**(1): 84-98.

151. Reiner, R., N. Krasnov-Yoeli, Y. Dehtiar and N. Jarrous (2008). "Function and assembly of a chromatin-associated RNase P that is required for efficient transcription by RNA polymerase I." PLoS One **3**(12): e4072.

152. Rodrigues, C. F., S. Silva and M. Henriques (2014). "Candida glabrata: a review of its features and resistance." Eur J Clin Microbiol Infect Dis **33**(5): 673-688.

153. Roetzer, A., T. Gabaldón and C. Schüller (2010). "From Saccharomyces cerevisiae to Candida glabrata in a few easy steps: important adaptations for an opportunistic pathogen." FEMS Microbiology Letters **314**(1): 1-9.

154. Roetzer, A., N. Gratz, P. Kovarik and C. Schuller (2010). "Autophagy supports Candida glabrata survival during phagocytosis." Cell Microbiol **12**(2): 199-216.

155. Sabourin, M., C. T. Tuzon, T. S. Fisher and V. A. Zakian (2007). "A flexible protein linker improves the function of epitope−tagged proteins in Saccharomyces cerevisiae." Yeast **24**(1): 39-45.

156. Salinas, K., S. Wierzbicki, L. Zhou and M. E. Schmitt (2005). "Characterization and purification of Saccharomyces cerevisiae RNase MRP reveals a new unique protein component." Journal of Biological Chemistry **280**(12): 11352-11360.

157. Sam, Q. H., M. W. Chang and L. Y. A. Chai (2017). "The fungal mycobiome and its interaction with gut bacteria in the host." International journal of molecular sciences **18**(2): 330.

158. Samji, H., A. Cescon, R. S. Hogg, S. P. Modur, K. N. Althoff, K. Buchacz, A. N. Burchell, M. Cohen, K. A. Gebo and M. J. Gill (2013). "Closing the gap: increases in life expectancy among treated HIV-positive individuals in the United States and Canada." PLoS One **8**(12): e81355.

159. Sampaio, P. and C. Pais (2014). "Epidemiology of invasive candidiasis and challenges for the mycology laboratory: specificities of Candida glabrata." Current Clinical Microbiology Reports **1**(1-2): 1-9.

160. Sanglard, D., F. Ischer, D. Calabrese, P. A. Majcherczyk and J. Bille (1999). "The ATP binding cassette transporter GeneCgCDR1 from Candida glabrata is involved in the resistance of clinical isolates to azole antifungal agents." Antimicrobial agents and chemotherapy **43**(11): 2753-2765.

161. Savastano, C., E. d. O. Silva, L. L. Gonçalves, J. M. Nery, N. C. Silva and A. L. T. Dias (2016). "Candida glabrata among Candida spp. from environmental health practitioners of a Brazilian Hospital." brazilian journal of microbiology **47**(2): 367-372.

162. Schelenz, S. (2008). "Management of candidiasis in the intensive care unit." Journal of Antimicrobial Chemotherapy **61**(suppl_1): i31-i34.

163. Schoch, C. L., G.-H. Sung, F. López-Giráldez, J. P. Townsend, J. Miadlikowska, V. Hofstetter, B. Robbertse, P. B. Matheny, F. Kauff and Z. Wang (2009). "The Ascomycota tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits." Systematic biology **58**(2): 224-239.

164. Schwarzmüller, T., B. Ma, E. Hiller, F. Istel, M. Tscherner, S. Brunke, L. Ames, A. Firon, B. Green and V. Cabral (2014). "Systematic phenotyping of a large-scale Candida glabrata deletion collection reveals novel antifungal tolerance genes." PLoS Pathog **10**(6): e1004211.

165. Seider, K., S. Brunke, L. Schild, N. Jablonowski, D. Wilson, O. Majer, D. Barz, A. Haas, K. Kuchler, M. Schaller and B. Hube (2011). "The facultative intracellular

pathogen Candida glabrata subverts macrophage cytokine production and phagolysosome maturation." J Immunol **187**(6): 3072-3086.

166. Serruya, R., N. Orlovetskie, R. Reiner, Y. Dehtiar-Zilber, D. Wesolowski, S. Altman and N. Jarrous (2015). "Human RNase P ribonucleoprotein is required for formation of initiation complexes of RNA polymerase III." Nucleic acids research **43**(11): 5442-5450.

167. Shaffer, R. L. (1975). "The major groups of Basidiomycetes." Mycologia **67**(1): 1-18.

168. Sheehan, D. J., C. A. Hitchcock and C. M. Sibley (1999). "Current and emerging azole antifungal agents." Clinical microbiology reviews **12**(1): 40-79.

169. Shields, R. K., M. H. Nguyen, E. G. Press, A. L. Kwa, S. Cheng, C. Du and C. J. Clancy (2012). "The presence of an FKS mutation rather than MIC is an independent risk factor for failure of echinocandin therapy among patients with invasive candidiasis due to Candida glabrata." Antimicrobial agents and chemotherapy **56**(9): 4862-4869.

170. Shin, J. H., M. J. Chae, J. W. Song, S.-I. Jung, D. Cho, S. J. Kee, S. H. Kim, M. G. Shin, S. P. Suh and D. W. Ryang (2007). "Changes in karyotype and azole susceptibility of sequential bloodstream isolates from patients with Candida glabrata candidemia." J Clin Microbiol **45**(8): 2385-2391.

171. Silva, D. B. d. S., L. M. C. Rodrigues, A. A. d. Almeida and K. M. P. d. Oliveira (2016). "Novel point mutations in the ERG11 gene in clinical isolates of azole resistant Candida species." Memórias do Instituto Oswaldo Cruz **111**(3): 192-199.

172. Silva, S., M. Henriques, A. Martins, R. Oliveira, D. Williams and J. Azeredo (2009). "Biofilms of non-Candida albicans Candida species: quantification, structure and matrix composition." Med Mycol **47**(7): 681-689.

173. Sorrell, T. C. and S. C. A. Chen (2009). Fungal-derived immune modulating molecules. Pathogen-Derived Immunomodulatory Molecules, Springer**:** 108-120.

174. Spellberg, B. (2011). "Vaccines for invasive fungal infections." F1000 Med Rep **3**: 13.

175. Stoltzfus, A. (2012). "Constructive neutral evolution: exploring evolutionary theory's curious disconnect." Biology Direct **7**(1): 35.

176. Sui, X., L. Yan and Y. Y. Jiang (2017). "The vaccines and antibodies associated with Als3p for treatment of Candida albicans infections." Vaccine **35**(43): 5786-5793.

177. Taff, H. T., K. F. Mitchell, J. A. Edward and D. R. Andes (2013). "Mechanisms of Candida biofilm drug resistance." Future microbiology **8**(10): 1325-1337.

178. Taschner, A., C. Weber, A. Buzet, R. K. Hartmann, A. Hartig and W. Rossmanith (2012). "Nuclear RNase P of Trypanosoma brucei: a single protein in place of the multicomponent RNA-protein complex." Cell reports **2**(1): 19-25.

179. Tauszig-Delamasure, S., H. Bilak, M. Capovilla, J. A. Hoffmann and J.-L. Imler (2002). "Drosophila MyD88 is required for the response to fungal and Gram-positive bacterial infections." Nature immunology **3**(1): 91.

180. Tay, L. Y., D. R. Herrera, B. P. Gomes, F. A. d. Santos and J. H. Jorge (2014). "Identification of Candida Spp. In Patients with Denture Stomatitis: Relationship with

Gender, Age, Time of Denture use and Newton's Classification." Journal of Dental Applications: 46-50.

181. Timmermans, B., A. De Las Peñas, I. Castaño and P. Van Dijck (2018). "Adhesins in Candida glabrata." Journal of Fungi **4**(2): 60.

182. Tortorano, A. M., C. Kibbler, J. Peman, H. Bernhardt, L. Klingspor and R. Grillot (2006). "Candidaemia in Europe: epidemiology and resistance." International journal of antimicrobial agents **27**(5): 359-366.

183. Ueno, K., Y. Matsumoto, J. Uno, K. Sasamoto, K. Sekimizu, Y. Kinjo and H. Chibana (2011). "Intestinal resident yeast Candida glabrata requires Cyb2p-mediated lactate assimilation to adapt in mouse intestine." PLoS One **6**(9): e24759.

184. Vale-Silva, L., E. Beaudoing, V. D. T. Tran and D. Sanglard (2017). "Comparative Genomics of Two Sequential Candida glabrata Clinical Isolates." G3 (Bethesda) **7**(8): 2413-2426.

185. Vandeputte, P., G. Tronchin, T. Bergès, C. Hennequin, D. Chabasse and J.-P. Bouchara (2007). "Reduced susceptibility to polyenes associated with a missense mutation in the ERG6 gene in a clinical isolate of Candida glabrata with pseudohyphal growth." Antimicrobial agents and chemotherapy **51**(3): 982-990.

186. Vandeputte, P., G. Tronchin, G. Larcher, E. Ernoult, T. Bergès, D. Chabasse and J.-P. Bouchara (2008). "A nonsense mutation in the ERG6 gene leads to reduced susceptibility to polyenes in a clinical isolate of Candida glabrata." Antimicrobial agents and chemotherapy **52**(10): 3701-3709.

187. Vermitsky, J. P. and T. D. Edlind (2004). "Azole resistance in Candida glabrata: coordinate upregulation of multidrug transporters and evidence for a Pdr1-like transcription factor." Antimicrob Agents Chemother **48**(10): 3773-3781.

188. Wainright, P. O., G. Hinkle, M. L. Sogin and S. K. Stickel (1993). "Monophyletic origins of the metazoa: an evolutionary link with fungi." Science **260**: 340-342.

189. Walker, S. C. and D. R. Engelke (2006). "Ribonuclease P: the evolution of an ancient RNA enzyme." Critical reviews in biochemistry and molecular biology **41**(2): 77-102.

190. Wargo, M. J. and D. A. Hogan (2006). "Fungal—bacterial interactions: a mixed bag of mingling microbes." Current opinion in microbiology **9**(4): 359-364.

191. Weber, C., A. Hartig, R. K. Hartmann and W. Rossmanith (2014). "Playing RNase P evolution: swapping the RNA catalyst for a protein reveals functional uniformity of highly divergent enzyme forms." PLoS genetics **10**(8): e1004506.

192. Wegscheid, B., C. Condon and R. K. Hartmann (2006). "Type A and B RNase P RNAs are interchangeable in vivo despite substantial biophysical differences." EMBO reports **7**(4): 411-417.

193. Weitzman, I. and R. C. Summerbell (1995). "The dermatophytes." Clinical microbiology reviews **8**(2): 240-259.

194. Westhof, E., D. Wesolowski and S. Altman (1996). "Mapping in three dimensions of regions in a catalytic RNA protected from attack by an Fe (II)-EDTA reagent." Journal of molecular biology **258**(4): 600-613.

195. Wilusz, J. E., S. M. Freier and D. L. Spector (2008). "3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA." Cell **135**(5): 919-932.

196. Xu, L., H. Chen, X. Hu, R. Zhang, Z. Zhang and Z. Luo (2006). "Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms." Molecular biology and evolution **23**(6): 1107-1108.

197. Yanagida, M. (2002). "The model unicellular eukaryote, Schizosaccharomyces pombe." Genome biology **3**(3): comment2003. 2001.

198. Yapar, N. (2014). "Epidemiology and risk factors for invasive candidiasis." Therapeutics and clinical risk management **10**: 95.

199. Yarrow, D. and S. A. Meyer (1978). "Proposal for amendment of the diagnosis of the genus Candida Berkhout nom. cons." Int J Syst Evol Microbiol **28**(4): 611-615.

200. Zahler, N. H., E. L. Christian and M. E. Harris (2003). "Recognition of the 5′ leader of pre-tRNA substrates by the active site of ribonuclease P." Rna **9**(6): 734-745.

201. Zilberberg, M. D. and A. F. Shorr (2009). "Fungal infections in the ICU." Infectious disease clinics of North America **23**(3): 625-642.

Yacine DAHMAN

Caractérisation de la RNase P nucléaire de *Candida glabrata* et amélioration des outils d'édition de son génome.

Résumé

Candida glabrata est une levure pathogène opportuniste, apparaissant aujourd'hui comme la deuxième cause de candidémie en Europe et en Amérique du Nord. Cette levure présente de nombreuses particularités génomiques telles que la présence de nouveaux domaines structuraux au sein d'ARN non-codants ubiquitaires. Le premier aspect de cette thèse a consisté en l'étude de la sous-unité ARN atypique de la Ribonucléase P nucléaire de *C. glabrata*. Cet ARN contient trois grand domaines additionnels octroyant au transcrit une taille trois fois plus élevée que la moyenne des sous-unités ARN des RNase P eucaryotiques. Les expériences réalisées ont permis une meilleure compréhension du rôle de ces domaines additionnels et ont démontré la présence inédite de la protéine Rcl1 au sein du complexe de la RNase P. Dans un second temps ce travail de thèse a aussi contribué à l'amélioration des outils d'édition du génome de *C. glabrata* existants. De nouvelles cassettes intégratives de faible taille et positivement sélectionnables ont été mises au point. Ces éléments présentent toutes les caractéristiques permettant leur utilisation dans la modification du génome de souches sauvages et d'isolats cliniques de *C. glabrata.*

Mots-clés:

Candida glabrata, levures pathogènes, ARN non-codants, RNase P, edition du génome

Résumé en anglais

Candida glabrata is an opportunistic pathogenic yeast, and is today the second causative agent of candidemia in Europe and North America. This yeast has many genomic peculiarities such as the presence of new structural domains within ubiquitous non-coding RNAs. The first aspect of this thesis was the study of the atypical RNA subunit of the nuclear Ribonuclease P of *C. glabrata*. This RNA contains three large additional domains giving the transcript an overall size more than three times larger than the average eukaryotic RNase P RNA subunits. The experiments performed led to a better understanding of the role of these additional domains and demonstrated for the first time the presence of the Rcl1 protein within the RNase P complex. Secondly, this thesis work also contributed to the improvement of existing genome editing tools in *C. glabrata*. New small and positively selectable integrative cassettes have been developed. These elements exhibited all the required characteristics for their use in wild-type strains and clinical isolates of *C. glabrata*.

Keywords:

Candida glabrata, pathogenic yeasts, non-coding RNA, RNase P, genome editing