Taxonomy, morphology, physiology and DNA content of a new *Blastobotrys* yeast species with potential applications in Biotechnology

Érica Torrico Vieira

Thesis to obtain the Master of Science Degree in

Biotechnology

Supervisors: Dr. Margarida Isabel Rosa Bento Palma

Prof. Dr. Isabel Maria de Sá Correia Leite de Almeida

Examination Committee

Presidente/Chairperson: Prof. Dr. Miguel Nobre Parreira Cacho Teixeira

Supervisors: Dr. Margarida Isabel Rosa Bento Palma

Vogais/Members of the Committee: Prof. Dr. José Paulo Nunes de Sousa Sampaio

July 2020

Preface

The work presented in this thesis was performed at the Institute for Bioengineering and Biosciences of Instituto superior Técnico, Universidade de Lisboa (Lisbon, Portugal), during the period September 2019 - March 2020, under the supervision of Dr. Margarida Isabel Rosa Bento Palma and Prof. Dr. Isabel Maria de Sá Correia Leite de Almeida.

Declaration

I declare that this document is an original work of my authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgements

I would like to express my gratitude to my supervisors, Dr. Margarida Palma and Professor Isabel Sá-Correia. Without them this work wouldn't have been possible, so I am very thankful for their contributions for the development of this thesis.

To professor Isabel Sá-Correia, my supervisor and leader of the Biological Sciences Research Group of the Institute for Bioengineering and Biosciences (iBB), Instituto Superior Técnico, Universidade de Lisboa, I would like to thank the opportunity to join her group and all the critics and advises that led me to achieve the proposed goals.

To Dr. Margarida Palma, my supervisor that worked with me daily side-by-side, I would like to express my deepest gratitude for all the knowledge shared and support throughout all the work. For all her patience and kindness. Without her advice and guidance this work wouldn't have progressed so much.

To all the research group, Ricardo, Cláudia, Luís, Rui, Marta, Miguel e Nuno, a special thank you for also providing guidance and help whenever needed.

The funding received by iBB (UID/BIO/04565/2019) and Yeastpec (ERA-IB-2/0003/2015) from the Portuguese Foundation for Science and Technology (FCT) is acknowledged.

To my university friends, Luís, Carolina, João, Margarida P., and Margarida B., for all the moments spent together during these years. For all the lunch hours shared, all the coffees taken together and the delicious dinners. A special thanks to Margarida P., my lab bench partner, for all the moments of help, for laughing together at our own mistakes and sharing our victories. To my friends Rafaela, Patrícia and Carolina that helped me to relieve the university stress every now and then. A big thanks to all of you, for being present whenever I needed.

Lastly, I want to whole-heartedly appreciate my family, my parents, Sulma e Rodolfo, and my boyfriend, João, for the unconditional support, comprehension, strength and for always encouraging me to accomplish my goals. Without them, I wouldn't have come so far, and I am gratefull to them.

Abstract

Lignocellulosic material residues, rich in xylose, are interesting feedstocks for biotechnology, due to their abundance and low-cost. *Saccharomyces cerevisiae* – the conventional yeast cell factory, cannot use xylose and other pentoses as substrate. So, there is an increased interest in the use of non-conventional yeasts, more suitable for specific applications. In this study, the isolation and identification of different non-conventional yeast strains was performed. Fifty isolates were obtained and molecularly identified, of which thirteen were grown in xylose media. Nine strains previously isolated by the group were also molecularly identified. During this step, strain IST508 belonging to a new yeast species of the genus *Blastobotrys* was isolated from olive tree soil in Alentejo. This new species, preliminarily designated *Blastobotrys solioliva* IST508 due to its ecological niche, presents low ITS identity (95%) to its closest match, *Blastobotrys proliferans*. Moreover, it exhibits a different morphology compared with the other described species of the genus and can assimilate and ferment a large range of carbon sources. This strain is also able to produce organic acids, like formic acid, in small concentrations, and the vitamin riboflavin. The complete genome of the haploid *Blastobotrys solioliva* IST508 was sequenced and analyzed. Sixty-five genes were chosen to be studied in detail, since they were previously associated with overproducing strains of riboflavin. All the genes involved in the main biosynthesis pathway of riboflavin were identified and gene duplications of *SEF1* and *MET2*, involved in the positive regulation of riboflavin main pathway and amino acid biosynthesis, respectively, were found.

Keywords: Non-conventional yeasts, *Blastobotrys sp.*, lignocellulosic residues bioconversion, Added-value bioproducts, Riboflavin

Resumo

Os resíduos de material lenhocelulósico, ricos em xilose, são matérias-primas interessantes para a biotecnologia devido à sua abundância e baixo custo. *Saccharomyces cerevisiae* - a fábrica celular de levedura convencional, não usa xilose e outras pentoses como substrato. Por estas razões, há um interesse crescente no uso de leveduras não-convencionais, mais adequadas para aplicações específicas. Neste estudo, realizou-se o isolamento e identificação de diferentes estirpes de leveduras não-convencionais. Cinquenta isolados foram obtidos e identificados molecularmente. Entre eles, treze foram selecionados por crescimento diferencial em xilose. Nove outras estirpes, previamente isoladas pelo grupo, foram também identificadas molecularmente. A estirpe IST508 pertencente a uma nova espécie de levedura do género *Blastobotrys* foi isolada do solo junto a oliveiras. Esta nova espécie, designada preliminarmente por *Blastobotrys solioliva* IST508, devido ao nicho ecológico onde foi isolada, apresenta baixa identidade na região ITS (95%) relativamente à espécie mais próxima, *Blastobotrys proliferans*. Apresenta uma morfologia diferente das restantes espécies do género, pode assimilar e fermentar uma grande variedade de fontes de carbono e é capaz de produzir ácidos orgânicos, tal como o ácido fórmico em pequenas concentrações, e a vitamina riboflavina. O genoma completo da levedura haplóide *Blastobotrys solioliva* IST508 foi sequenciado e analisado. Sessenta e cinco genes, previamente associados a estirpes sobre-produtoras de riboflavina foram escolhidos para serem estudados em detalhe. Todos os genes associados à via principal de produção foram identificados e encontraram-se duplicações dos genes *SEF1* e *MET2*, envolvidos na regulação positiva da via principal da riboflavina e na biossíntese de aminoácidos, respetivamente.

Palavras-chave: Leveduras não-convencionais, *Blastobotrys sp*., bioconversão de resíduos lenhocelulósicos , Bioprodutos de valor acrescentado , Riboflavina

Contents

Content of Figures

Content of Tables

Abbreviations List

3PG	3-phosphoglycerate
A. gossypii	*Ashbya gossypii*
Ac-CoA	Acetyl Coenzyme A
AMP	Adenosine Monophosphate
ArP	5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione
ArPP	5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5'-phosphate
ARPP	5-amino-6-ribosylamino-2,4 (1H,3H)-pyrimidinedione-5'-phosphate
ATP	Adenosine Triphosphate
B. adeninivorans	*Blastobotrys adeninivorans*
B. meliponea	*Blastobotrys meliponea*
B. nivea	*Blastobotrys nivea*
B. proliferans	*Blastobotrys proliferans*
B.raffinosifermentans	*Blastobotrys raffinosifermentans*
B. subtilis	*Bacillus subtilis*
BLAST	Basic Local Alignment Search Tool
BSRG	Biological Sciences Research Group
C. ammoniagenes	*Corynebacterium ammoniagenes*
C. famata	*Candida famata*
C/N	Carbon/Nitrogen
D1/D2	Domain 1 / Domain 2
DAG	Diacylglycerol
DArPP	2,5-diamino-6-ribitylamino-4 (3H)-pyrimidinone-5'-phosphate
DARPP	5-diamino-6-ribosylamino-4 (3H)-pyrimidinone-5'-phosphate
DHBP	5-phosphate to 3,4-dihydroxy-2-butanone 4-phosphate
DLR	6,7-Dimethyl-8-Ribityl-Lumazine
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
DRL	6,7-dimethyl-8-ribityllumazine

E. coli	*Escherichia Coli*
EBI	European Bioinformatics Institute
EDTA	Ethylenediaminetetraacetic Acid
EMBOSS	European Molecular Biology Open Software Suite
FA	Fatty Acids
FAD	Flavin Adenine Dinucleotide
FMN	Flavin Mononucleotide
gDNA	Genomic Dna
Glc-6P	Glucose 6-Phosphate
Gly	Glycine
GTP	Guanosine Triphosphate
HPLC	High-Performance Liquid Chromatography
iBB	Institute For Bioengineering And Biosciences
IMP	**I**nosine Monophosphate
IST	Instituto Superior Técnico
ITS	Internal Transcribed Spacer
LSU	Large Subunit
ME	Malt Extract
mRNA	Messenger Ribonucleic Acid
MUSCLE	Multiple Sequence Alignment
NCBI	National Center For Biotechnology Information
NCBI	National Center For Biotechnology Information
NEB	New England Biolabs
NGS	Next-Generation Sequencing
OAA	Oxaloacetate
OD	Optical Density
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEP	Phosphoenolpyruvate
PP	Pentose *Phosphate Pathway*

PRA	5-Phosphpribosylamine
PRPP	Phosphoribosylpyrophosphate
Pyr	Pyruvate
QS	*Quorum Sensing*
rDNA	Ribossomal Deoxyribonucleic Acid
RF	Riboflavin
rpm	Rotations Per Minute
Ru-5P/Ribu-5P	Ribulose-5-Phosphate
S. cerevisiae	*Saccharomyces cerevisiae*
SCO	Single Cell Oil
Ser	Serine
Sp.	Species
TAG	Triacylglycerol
TBE	Tris/Borate/EDTA
TCA	Tricarboxylic Acid Cycle
Thr	Threonine
UMP	Uridine Monophosphate
WGD	Whole Genome Duplication
WLN	Wallerstein Nutrient Agar
XMP	Xanthosine Monophosphate
YGAP	Yeast Genome Annotation Pipeline
YM	Yeast extract-malt extract-peptone-glucose
YNB	Yeast Nitrogen Base
YPD	Yeast Extract Peptone Dextrose
YPDX	Yeast Extract Peptone Dextrose Xylose
YPX	Yeast Extract Peptone Xylose

Thesis Outline

The yeast domain contains a wide variety of species of which several present interesting traits for applications in biotechnology [1]. The study of non-conventional yeasts may uncover new strains with different traits that can be useful for bioprocesses improvement: they can be able to efficiently catabolize several sugars and other carbon sources and resist to different stresses, for example, improving the production of added-value products, like pharmaceuticals, chemical intermediates, biofuels, enzymes, among others [2–5].

One of the lines of biotechnology research nowadays is to find microbial strains with the capacity to utilize renewable feedstocks and transform them into biofuels or value-added products through the biorefinery concept, due to the increasing concerns on sustainability issues worldwide [2–4]. The largest source of renewable organic material includes lignocellulosic residues, generated through agricultural and forestry practices. They have a great biotechnological potential due to its abundance and low-cost. It is composed of cellulose, hemicellulose, and lignin [5]. Cellulose and hemicellulose can be processed to release sugar monomers, like glucose and xylose that can be used as substrates in bioprocesses. Xylose is the most abundant sugar of the biosphere after glucose [5]. The most widely used yeast in fermentation processes, *Saccharomyces cerevisiae,* cannot metabolize xylose and other pentoses. In contrast, several non-*Saccharomyces* (non-conventional yeasts) can use this and other pentoses as substrate [5]. Among the non-conventional yeasts with interest are those comprised in the *Blastobotrys* genus, which includes several yeast strains with potential applications in biotechnology. *Blastobotrys nivea,* described in 1967 by von Klopotek, was the first species described and originated this genus [6,7]. Later, Kurtzman *et al.* (2011) described the genus in more detail. Based on the D1/D2 gene sequences, the genus was phylogenetically placed and the assigned species were characterized as anamorphic members of the *Saccharomycetales* order [6,8].

Yeasts from the *Blastobotrys* genus are promising yeasts to be used in bioprocesses, due to their unusual characteristics: most of these strains are able to utilize a wide range of carbon sources, through assimilation or fermentation, including uncommon sources of carbon, like adenine, glycine, uric acid, n-hexadecane, putrescine and branched-chain aliphatic compounds, such as isobutanol, leucine and isoleucine [6,9–11]. Strains of this genus can utilize different sources and mixtures of raw materials, including pentoses, as feedstocks in bioprocesses, and consume most of the sugars present, increasing the process yield [12]. Also, some yeast strains of *Blastobotrys* present unusual stress tolerance to high temperature (some strains can grow at up to 48ºC), dry environments, high salinity conditions and high osmotic pressures [6,13,14].

Currently, there are 24 species reported as belonging to this genus isolated from a wide variety of ecological (soil, water, air, plants and the gut of animals) and geographical sources [6,10]. Some, like *Blastobotrys raffinosifermentans*, are already being used in bioprocesses, for example, for the synthesis of enzymes naturally encoded in its genome, like tannases [13].

In this thesis, the isolation and characterization of fifty yeast strains of non-conventional yeasts was performed. Their ITS and D1/D2 sequences from ribosomal DNA were obtained for taxonomic identification. The objective of this study was the finding of new yeast strains able to catabolize xylose, due to its abundance in lignocellulosic feedstocks. Nine other yeast strains, previously isolated by the group, were also identified and examined.

During the isolation and characterization process, a yeast species showing an identity percentage of approximately 95% with the *Blastobotrys* genus at the level of the ITS sequences was found. This species was isolated from the soil surrounding an olive tree, in *Ferreira do Alentejo, Portugal.* Due to the low percentage of identity found with all the other *Blastobotrys* species described, this species was proposed as a new anamorphic member of *Blastobotrys* and preliminarily designated as *Blastobotrys solioliva* IST508.

Following the identification of this novel yeast species, the second part of the work involved its thorough characterization. The phylogenetic placement of this yeast was assessed, through comparisons of the D1/D2 sequences with other known yeast species, morphological observations were done and physiological tests regarding its ability to ferment or assimilate different compounds were performed, in order to understand its potential for biotechnology. Evidence was found for the production of a yellow color compound since the liquid growth medium turned bright yellow after a few hours of incubation. Based on the UV scan, the colored compound is believed to be the vitamin riboflavin. The complete genome of the yeast *Blastobotrys solioliva* IST508 was sequenced, through Illumina, and the annotated genes were examined to search for interesting functions, with focus on genes involved with riboflavin production.

1. Introduction

1.1 Non-conventional yeasts in Biotechnology

The increasing global population, together with concerns related to increasing energetic needs and global warming renders the need to exploit more sustainable options regarding the production of clean energy and other added-value products as solution for the future [15,16]. In order to recycle and reuse resources towards waste minimization industrial by-products and wastes from manufacturing and consumers are proposed to be used as feedstock, instead of expensive refined sugars, like glucose, to produce added-value compounds such as acetic acid, single cell oils, citric acid, glycerol or ethanol, using yeast cell factories [15–18].

In biorefineries, a wide range of technologies are used to separate the biomass resources (wood, corn, grasses…) into their building blocks (carbohydrates, proteins, triglycerides…), which can be converted by a cell factory into added-value products, like chemicals and biofuels [2,3]. First generation biofuels (bioethanol and biodiesel) are produced from sugar, starch, vegetable oil or animal fats using conventional technologies [2]. Bioethanol is obtained from sugarcane, sugar beet and starch crops, whereas biodiesel is produced from oil-based crops such as rapeseed, sunflower, soybean but also from palm oil and waste edible oils, and biogas is produced by the anaerobic digestion of mixtures of corn, manure and grasses [2,3]. Despite the success of sugar- and starch-based bioprocesses, first generation biofuels have the main disadvantage of competing with food and feed industries. If the prices of fuels increase, a large proportion of land and feedstuff will be shifted towards producing biofuel instead of using it to produce food, raising ethical problems. Besides this, these raw-materials will not be able to meet the increasing demand for bio-based products [2,3]. Second generation biofuels overcome these limitations and are produced by a circular economy, since they are produced from non-food crops and wastes that are re-utilized and its value is regenerated, for example biodiesel can be produced from edible oil waste streams of food industry and bioethanol from lignocellulosic feedstock [2].

The use of renewabe sources and organic residues in Biorefineries requires the sustainable and regular supply of feedstock. The main biomass raw-materials are provided from forestry and agriculture, process residues and leftovers from industries (agro-industrial wastes) and municipal solid waste and wastewaters, and aquaculture [2]. These residues mainly consist of carbohydrates (starch, cellulose and hemicellulose) and lignin, triglycerides (oils and fats) or organic residues (manure, residues from fresh fruit and vegetable industries) [18,19].

Lignocellulosic biomass has special interest since it is an abundant and low-cost feedstock with high organic content, mainly complex carbohydrates, to produce biofuels and added-value products [15,16,19]. However, this feedstock cannot be directly converted to the desired products, due to the fact that the plant carbohydrates are in a complex structure of lignocellulose, which requires a pretreatment method to release simple sugars (glucose, xylose, arabinose, mannose, glucose and galactose) that can then be assimilated and transformed into added-value products [5,15]. Besides, inhibitors of microbial growth are also formed during the hydrolytic pretreatment process, like furfural, acetic acid and formic acid

[3,15,16]. The yeast cell factory used must have a broad substrate range, must contain enzymes able to hydrolyze celluloses and hemicelluloses, must have a high physiological flexibility and be resistant to potential inhibitors formed during the pre-treatment step in order to tolerate the conditions imposed by the process and fulfill the requirements for yield, productivity, and titer [3,15,16].

The yeast domain contains more than 2000 yeast species of which several present potentially interesting traits for biotechnology [1,4]. Nowadays, yeasts with the capacity to utilize renewable feedstocks and transform them into biofuels or value-added products through the biorefinery concept are highly desired, due to the increasing concerns on sustainability issues worldwide [2–4]. Despite being extremely diverse and widespread, *Saccharomyces cerevisiae* strains are the workforces of the biotechnology industry, especially regarding fermentation processes, due to their high tolerance to the presence of ethanol and organic acids, low pH and scarce oxygen availability. This unicellular eukaryote is one of the most popular cell factories since its genetic and physiological background is well known. It was the first sequenced eukaryote and it has been used to develop many molecular and genomic tools [1]. Moreover, it is compatible with high density and large-scale fermentation processes. *S. cerevisiae* has been successfully used, not only for the traditional applications in alcoholic fermentations, but also in modern biotechnological industry to produce a wide variety of products, such as organic acids, amino acids and enzymes [4]. On the contrary, non-conventional yeasts, species other than *S. cerevisiae*, are in general poorly known and studied, with relevant exceptions like the yeasts *Yarrowia lipolytica* and *Pichia pastoris* [1,4,20,21].

The study of different non-conventional yeasts may uncover new strains able to efficiently catabolize several sugars and other carbon sources, resistant to different stresses present during the bioprocesses, and producers of improved added-value products [1]. For these reasons the yeasts chosen for a given bioprocess or to be genetically engineered for a specific bioprocess have a crucial role. *S. cerevisiae* is not able to efficiently use the pentoses xylose, arabinose or galacturonic acid, which are requirements to efficiently produce added-value compounds from a wide-range of agro-industrial wastes [21]. Differently, non-conventional yeasts, like *Pichia stipitis (Scheffersomyces stipitis), Scheffersomyces (Candida) shehatae, Candida tropicalis, Pachysolen tannophilus* and *Cyberlindnera jadinii,* can use a wide range of substrates, including xylose [5,22–25]. For example, in the production of second-generation bioethanol, to increase the fermentation yield, the microorganism used must be able to ferment all the monosaccharides present and, in addition, withstand potential inhibitors in the hydrolysates. *S. cerevisiae* is the most commonly used ethanol producer, however and as already said, it cannot ferment pentoses, which may constitute up to 45% of the raw material, leaving residual sugars after the process [26] and the biosynthetic ability is limited. This gap can be fulfilled by selected non-conventional yeasts.

Non-conventional yeast strains may possess different traits of interest, including a high metabolic diversity, as they have evolved in their diverse habitats [14,21]. In fact, some non-conventional yeasts are naturally equipped to hydrolyze and catabolize some of the substrates cited above, and transform them, for example, into organic acids (citric acid), drug precursors (α-ketoglutaric acid), aroma compounds (γ-decalactone), single cell oils and enzymes (lipases) [16]. For these reasons, the wide variety of non-

conventional yeasts has gained attention for applications in the areas of pharmaceuticals, chemical intermediates, biofuels, enhancement of desirable flavors in beverages, biocontrol and bioremediation and heterologous protein production [5].

1.2. The *Blastobotrys* Genus

The study and use of non-conventional yeast strains for biotechnological purposes is increasing [2–4]. In the next sections, non-conventional yeasts of the *Blastobotrys* genus on the focus of the present work, which also include strains with an interesting potential for biotechnology, will be discussed in detail.

The *Blastobotrys* genus was established by von Klopotek in 1967. Klopotek isolated a strain from compost in Germany, in 1963, that was first described as hyphomycete, having a soft-walled mycelium, hyaline, sparsely but distinctly septate [7]. Conidiophores were upright, simple or more rarely branched. They also described forming spore mother cells [7]. This strain was named *Blastobotrys nivea* and is the first species described that originated this genus, is the type species [6,7]. Table 1 shows the taxonomic rank of the *Blastobotrys* genus [27].

Table 1 - Taxonomic rank of the Blastobotrys genus [27].

Kingdom	*Fungi*
Phylum	*Ascomycota*
Subphylum	*Saccharomycotina*
Class	*Saccharomycetes*
Order	*Saccharomycetales*
Family	*Trichomonascaceae*
Genera	*Blastobotrys*

Later, Kurtzman *et al.* (2011) described the genus as having well-defined colonies, dry, with a snow-white color and presenting sometimes hyphal growth or buds. Members of *Blastobotrys* were found to contain mannose as the predominant cell carbohydrate, whereas rhamnose and xylose were absent. Additionally, septa were reported to have micropores and the coenzyme system observed was Q-9 [6]. Their research demonstrated that in fact, the assigned species to the *Blastobotrys* genus are anamorphic members of the *Saccharomycetales* [8]. So, in opposition to what was previously described, the genus is comprised of the asexual members of this order [6]. Teleomorph and anamorph are the phases of the fungal life cycle. Teleomorphosis is the sexual phase of the life cycle and involves the production of meiospores. Anamorph is the asexual phase of the fungus and involves the development of mitospores (conidia), which is the case of the species from the genus *Blastobotrys* [28].

Taxonomic and phylogenetic analyses of the yeasts from the genus *Blastobotrys* also corroborated the findings cited above. These studies are based on the study of sequences from ribosomal DNA, which has a fundamental role in protein synthesis, having its organization conserved. The most used sequences for the identification of yeast species are those from the internal transcribed spacer (ITS) region and from the D1/D2 region, which are present between the ITS1-ITS2 and at the 26S, respectively [29]. So, based on the D1/D2 gene sequences, the genus *Blastobotrys* was confirmed to belong to the *Saccharomycetales* order, and the suggestion was made that *Blastobotrys, Arxula* and

Sympodiomyces could represent the same anamorphic genus, resulting in the transfer of species assigned to the latter two genera to *Blastobotrys* [8,29]. Later, this was confirmed by multigene phylogenetic analysis. The *Trichomonascus / Blastobotrys* clade includes anamorphic and teleomorphic members and comprises species from the genus *Candida* and *Blastobotrys* as anamorphic members, whereas *Trichomonascus* is the teleomorphic member [6,9,30]. Currently, twenty-four species were described in the *Blastobotrys* genus (Table 2) [6,9–11].

Table 2 - Blastobotrys species described in the present [6,9–11].

Blastobotrys adeninivorans	*Blastobotrys meliponeae*
Blastobotrys americana	*Blastobotrys mokoenaii*
Blastobotrys arbuscula	*Blastobotrys muscicola*
Blastobotrys aristata	*Blastobotrys nivea*
Blastobotrys attinorum	*Blastobotrys parvus*
Blastobotrys bombyces	*Blastobotrys peoriensis*
Blastobotrys capitulata	*Blastobotrys persicus*
Blastobotrys chiropterorum	*Blastobotrys proliferans*
Blastobotrys elegans	*Blastobotrys raffinosifermentans*
Blastobotrys illinoisensis	*Blastobotrys robertii*
Blastobotrys indianensis	*Blastobotrys serpentis*
Blastobotrys malaysiensis	*Blastobotrys terrestris*

The phylogenetic placement of most of the *Blastobotrys* species and its teleomorphic genus *Trichomonascus* are shown in Figure 1 [6].

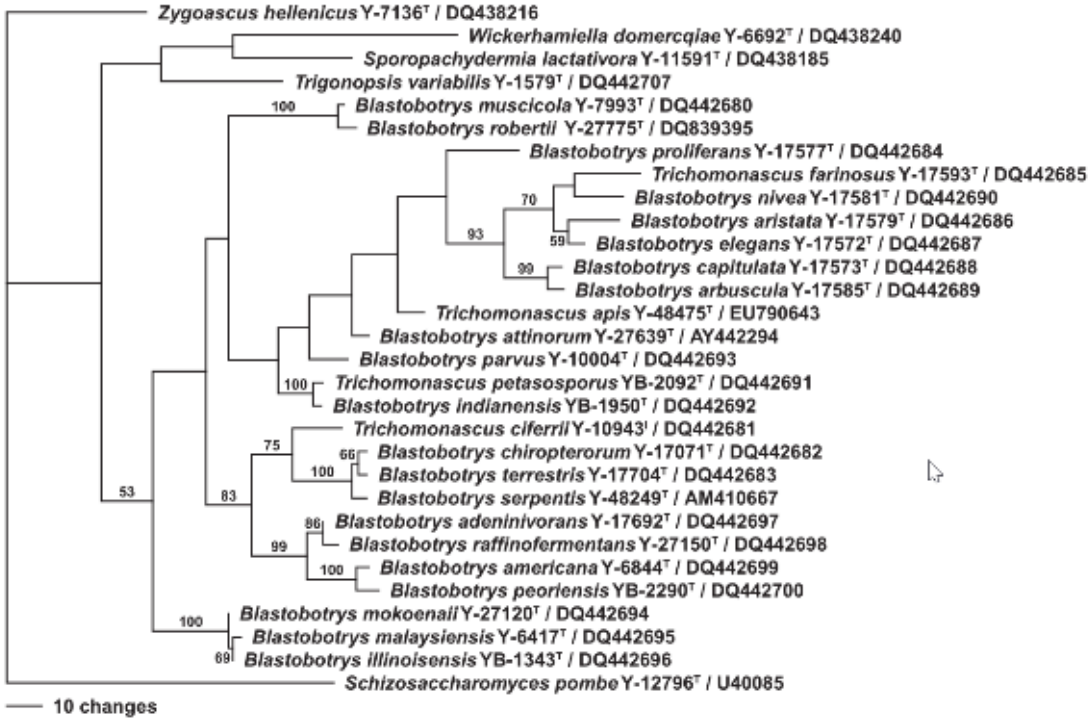

Figure 1 - Phylogenetic tree of Blastobotrys species accepted until 2011 by Kurtzman et al. (2011) and its teleomorphic genus Trichomonascus based on maximum parsimony analysis of the D1/D2 rRNA gene sequences. Each strain name is followed by its accession number in GenBank. Bootstrap values are from 1000 replicates [6].

1.2.1 Geographical Origins and Ecology

The species of the genus *Blastobotrys* have been isolated from different sources like soil, water, air, plants and the gut of animals [6,10]. The great diversity of their habitats is reflected in their different abilities to ferment or to assimilate several compounds.

Recently, 3 novel species of this genus have been reported: *Blastobotrys meliponae*, isolated from the honey of the bee *Melipona scutellaris, Blastobotrys bombycis*, isolated from the gut of *Bombyx mori* (silkworm) and *Blastobotrys persicus* isolated from cave soil [9–11].

Table 3 summarizes the location and isolation substrate of each one of the type strains of the species described so far. In the case of *B. adeninivorans, B. capitulata* and *B. parvus*, information about the several strains found is given [6,9–11].

Table 3 – The origins of Blastobotrys species [6,9–11].

Species	Isolation substrate	Location	Ref
B. adeninivorans	Maize	Netherlands	6
	Soil	Netherlands/South Africa	6
	Liver and intestines Gila monster	Sweden	6
B. americana	Unknown	Unknown	6
B. arbuscula	House air	Finland	6
B. aristata	Moldy plaster	Czecholovakia	6
B. attinorum	Sponge material of a lab nest of a leaf-cutting ant	Brazil	6
B. bombyces	Worm gut	India	10
B. capitulata	Straw	USA	6
	Rotting candelabra tree	South Africa	6
B. chiropterorum	Bat liver	Colombia	6
B. elegans	House air	Finland	6
B. illinoisensis	Dead tree	USA	6
B. indianensis	Fallen pine tree	USA	6
B. malaysiensis	Cave soil	Malaysia	6
B. meliponae	Honey	Brazil	11
B. mokoenaii	Soil	South Africa	6
B. muscicola	Moss on a fallen log	USA	6
B. nivea	Compost	Germany	6
B.parvus	Ocean	Antarctic/Indian/Pacific	6
B. peoriensis	Rotted log	USA	6
B. persicus	Cave soil	Iran	9
B. proliferans	Mite infested nut	Brazil	6
B. raffinosifermentans	Unknown	Unknown	6
B. robertii	Rotted Scoth pine	Netherlands	6
B. serpentis	Snake intestine	India	6
B. terrestris	Grassland soil	South Africa	6

1.2.2 *Blastobotrys* Physiology

Fermentation capacity has great interest in biotechnology, especially if it is coupled with the ability to metabolize common and inexpensive sources of carbon and energy. For example, yeasts capable of metabolizing D-xylose, the second most abundant sugar in lignocellulosic biomass, are of great interest, since they may be useful in producing second-generation bio-ethanol [10]. This desirable trait has encouraged the search for novel yeast strains, with high metabolic diversity, able to ferment several of the sugars present in agro-industrials feedstocks, like xylose, arabinose, galacturonic acid and rhamnose [5,15].

Table 4 shows the ability to ferment different sugars of all the species of the genus *Blastobotrys* known. So far, none of the species described in this genus has been specifically tested for fermentation of some of the sugars present in agro-industrial residues. However, there are some data on assimilation of these carbohydrates (Table 5) [10].

Table 4 – Fermentation ability of the species assigned to the genus Blastobotrys. Symbols: +/w, positive or weak; -, negative; v, variable; n, no data available [6,9–11].

Species	Glucose	Galactose	Sucrose	Maltose	Lactose	Raffinose	Trehalose
B. adeninivorans	+	+	+	+	-	+	+
B. americana	+	-	-	-	-	-	+
B. arbuscula	+	+	-	+	-	-	+
B. aristata	+	v	-	+	-	-	+
B. attinorum	+	-	-	v	-	-	-
B. bombycis	+	+	-	-	-	n	+
B. capitulata	+	+	-	v	-	-	v
B. chiropterorum	-	-	-	-	-	-	-
B. elegans	-	-	-	-	-	-	-
B. illinoisensis	+	+	+	+	-	-	+
B. indianensis	+	+	-	+	-	-	+
B. malaysiensis	+	+	+	+	-	-	+
B. meliponae	+	+	+	-	+	-	n
B. mokoenaii	+	+	-	-	-	-	+
B. muscicola	-	-	-	-	-	-	-
B. nivea	+	+	-	+	-	-	+
B. parvus	-	-	-	-	-	-	-
B. peoriensis	+	+	-	-	-	-	+
B. persicus	-	-	-	-	-	-	-
B. proliferans	+	+	+	+	-	+	v
B. raffinosifermentans	+	+	+	+	-	+	+
B. robertii	-	-	-	-	-	-	-
B. serpentis	-	-	-	-	-	-	-
B. terrestris	-	-	-	-	-	-	-

Besides fermentation, the ability to assimilate different compounds is important in biotechnology. The assimilations' capacity of different chemicals, of all known *Blastobotrys sp.*, is shown in Table 5. Assimilation of a given compound means that the yeast can grow aerobically on it, when it is supplied as the sole source of energy. In general, if a yeast is able to ferment a carbohydrate, is able to grow on it, but many yeasts are able to grow aerobically on sugars that they cannot ferment [6].

Blastobotrys species are able to assimilate a wide range of carbohydrates, as shown in Table 5. The order used for the tables was based on the order suggested by Kurtzman *et al*. (2011) in the summary of species characteristics [6].

As mentioned above, most of the substrates used in biotechnology are rich in mixtures of sugars, being of extreme importance that the yeasts responsible for the process can metabolize all types of carbohydrates present. Otherwise, slow and incomplete reactions can occur, leading to unscheduled loss of tank capacity due to extended processing times and the potential for microbial instability due to residual sugars [12]. All indicates that if in need, the species of this genus can utilize different sources and mixes of raw materials, and consume most of the sugars present, increasing the yield of the process where they are being used. Besides this and as explained before, ability to utilize pentoses is extremely required in the biotechnological industry. Table 5 shows that most of the *Blastobotrys* species are able to assimilate arabinose and xylose, confirming the potential for biotechnological applications of this genus. In terms of temperature, most species grow well between 19 and 30ºC. For many species and compounds there is no data available on assimilation ability, so more experimental tests are required to gain more knowledge on this genus. Tables 4 and 5 were prepared by the same authors and the compounds tested for fermentation and assimilation were not all verified for both cases, so the absence of a compound in one of the tables indicates that there is no information about its assimilation or fermentation in the *Blastobotrys* species.

Table 5-Assimilation ability of the species assigned to the genus Blastobotrys. Symbols: +/w, positive or weak; -, negative; v, variable; n, no data available [6,9–11].

	Glucose	Inulin	Sucrose	Raffinose	Melibiose	Galactose	Lactose	Trehalose	Maltose	Melezitose	Methyl-α-D-glucoside	Soluble starch	Cellobiose	Salicin	L-Sorbose	L-Rhamnose	D-Xylose	L-Arabinose	D-Arabinose	D-Ribose	Methanol	Ethanol	Glycerol	Erythritol	Ribitol	Galactitol	D-Mannitol	D-Glucitol
B. adeninivorans	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+
B. americana	+	-	-	+	+	+	-	+	-	-	+	-	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+
B. arbuscula	+	v	-	-	-	+	-	+	+	-	v	+	+	-	+	-	+	+	-	+	n	-	+	+	+	+	+	+
B. aristata	+	v	+	v	v	+	+	+	+	-	v	+	+	+	+	v	+	+	+	+	n	-	+	+	+	+	+	+
B. attinorum	+	-	-	-	-	+	-	-	+	-	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+
B. bombycis	+	-	-	-	+	+	-	n	+	-	n	-	+	-	+	-	+	+	-	+	n	-	+	+	-	+	+	-
B. capitulata	+	v	v	v	-	+	v	+	v	-	-	+	+	+	+	+	+	+	v	+	n	+	+	+	+	+	+	+
B. chiropterorum	+	-	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+
B. elegans	+	v	-	-	-	+	+	+	+	-	-	v	+	v	v	-	+	v	-	+	n	+	+	+	+	v	+	+
B. illinoisensis	+	-	+	-	-	+	-	-	+	-	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	+
B. indianensis	+	-	-	+	+	+	-	+	+	-	+	+	+	-	+	-	+	+	+	-	-	+	+	-	+	-	+	+
B. malaysiensis	+	-	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+
B. meliponae	+	n	n	n	-	+	n	n	+	n	n	n	n	n	+	+	+	+	n	n	n	n	-	+	n	n	n	n
B. mokoenaii	+	+	+	-	-	+	-	+	+	-	+	+	+	-	+	+	+	+	-	-	-	+	+	+	+	+	+	+
B. muscicola	+	-	+	v	v	+	+	v	+	+	v	+	+	v	-	-	-	-	-	-	-	-	+	-	-	-	+	-
B. nivea	+	v	v	-	+	+	+	+	+	-	v	+	+	+	+	-	+	+	+	+	n	+	+	+	+	+	+	+
B. parvus	+	-	+	-	-	+	-	+	+	-	-	-	+	-	-	+	+	+	v	v	-	-	+	+	+	-	+	+
B. peoriensis	+	-	+	+	+	+	+	-	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+
B. persicus	+	-	+	+	w	+	-	n	+	w	n	-	+	w	n	w	+	n	-	-	-	w	+	n	+	n	-	n
B. proliferans	+	v	+	+	+	+	+	+	+	v	v	v	+	+	+	v	+	+	+	v	n	+	+	+	+	+	+	+
B. raffinosifermentans	+	+	+	+	+	+	-	-	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+
B. robertii	+	-	+	+	+	+	+	+	+	-	v	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-
B. serpentis	+	-	+	+	+	+	-	+	+	-	n	-	+	+	+	+	+	+	+	+	-	-	+	+	+	+	n	+
B. terrestris	+	-	+	+	-	+	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+

Table 5 – Key characters, regarding assimilation ability, of the species assigned to Blastobotrys. Symbols: +/w, positive or weak; -, negative; v, variable; n, no data available [6,9–11]. Cont.

	myo-Inositol	DL-Lactate	Succinate	Citrate	D-Gluconate	D-Glucosamine	N-acetyl-D-glucosamine	Hexadecane	Nitrate	Vitamin-free	2-keto-D-gluconate	5-keto-D-gluconate	Saccharate	Xylitol	L-Arabinitol	L-Lysine	Ethylamine	50%Glucose	10%NaCl/5% glucose	Starch formation	Cycloheximide 0.01%	Growth at 19ºC	Growth at 25ºC	Growth at 30ºC	Growth at 37ºC	Growth at 40ºC	Growth at 45ºC
B. adeninivorans	+	-	+	+	+	+	n	n	+	-	+	n	-	n	n	n	n	+	n	-	+	+	+	+	+	+	n
B. americana	+	+	+	-	+	+	+	+	-	-	-	+	-	n	n	n	n	n	+	-	+	+	+	+	-	-	-
B. arbuscula	-	v	+	-	n	+	n	n	-	-	n	n	n	n	n	n	n	n	n	n	n	+	+	+	-	-	-
B. aristata	v	-	+	v	n	+	n	n	-	-	n	n	n	n	n	n	n	n	n	n	n	+	+	+	-	-	-
B. attinorum	+	+	-	-	+	n	+	-	-	-	-	n	n	n	n	n	n	+	n	-	+	+	+	+	+	n	n
B. bombycis	+	-	+	+	+	-	+	w	-	-	n	n	n	+	n	n	n	n	+	n	n	n	n	+	-	n	n
B. capitulata	+	-	+	v	n	+	n	n	-	-	n	n	n	n	n	n	n	n	n	n	n	+	+	+	+	n	n
B. chiropterorum	+	+	-	+	+	+	+	-	-	-	+	+	-	-	-	+	+	+	+	n	+	+	+	+	+	-	-
B. elegans	-	-	+	v	n	v	n	n	-	-	n	n	n	n	n	n	n	n	n	n	n	+	+	+	-	-	-
B. illinoisensis	+	-	+	+	+	+	+	+	-	+	+	+	-	n	n	n	n	n	+	-	+	+	+	+	+	n	n
B. indianensis	+	-	+	+	-	+	+	+	-	-	+	+	-	n	n	n	n	n	+	-	+	+	+	+	-	-	-
B. malaysiensis	+	+	+	+	-	+	+	-	-	+	+	+	-	n	n	n	n	n	+	-	+	+	+	+	+	n	n
B. meliponae	n	n	n	-	+	-	n	-	-	n	n	n	n	n	n	n	n	n	n	n	n	n	+	+	+	n	n
B. mokoenaii	+	-	+	+	+	+	n	n	-	+	+	n	n	+	+	+	+	-	+	n	+	+	+	+	+	+	+
B. muscicola	-	-	+	+	-	v	+	-	-	v	-	-	-	n	n	n	n	n	-	-	+	+	+	+	+	n	n
B. nivea	+	-	+	+	n	+	n	n	-	-	n	n	n	n	n	n	n	n	n	n	n	+	+	+	+	n	n
B. parvus	+	-	-	-	-	+	+	-	-	-	n	n	-	n	n	n	n	-	-	-	n	+	+	-	-	-	-
B. peoriensis	+	-	+	-	+	+	+	+	+	-	+	+	-	n	n	n	n	n	+	-	+	+	+	+	-	-	-
B. persicus	+	+	-	-	n	w	n	n	-	n	n	n	n	n	+	+	+	+	+	n	-	w	+	+	+	+	-
B. proliferans	+	v	+	+	n	+	n	n	-	-	n	n	n	n	n	n	n	n	n	n	n	+	+	+	+	n	n
B.raffinosifermentans	+	+	+	+	+	+	+	+	+	-	+	+	-	n	n	n	n	n	+	-	+	+	+	+	+	n	n
B. robertii	-	-	+	+	v	-	+	+	-	-	n	n	n	n	n	n	n	n	n	n	n	+	+	+	-	-	-
B. serpentis	+	+	+	+	+	+	n	n	-	-	-	n	-	n	n	n	n	n	n	-	+	+	+	+	-	-	-
B. terrestris	+	+	+	+	+	-	n	n	+	-	-	n	-	+	n	n	n	+	n	-	+	+	+	+	+	+	n

Five species of the *Blastobotrys* genus were chosen to be discussed in more detail in this introduction: *B. nivea, B. proliferans, B. meliponae, B. adeninivorans and B. raffinosifermentans. Blastobotrys nivea* was selected because it is the type species, *B. proliferans* and *B. meliponae* were selected because they are taxonomically close to the new species found during this work, and *B. adeninivorans* and *B. raffinosifermentans* because they are the *Blastobotrys* species more represented in biotechnological studies.

1.2.3 *Blastobotrys nivea*

Blastobotrys nivea was described for the first time by von Klopotek, in 1967 and isolated from compost, in Germany [7]. It was observed that the colonies grew slowly on malt agar. Its aspect was dense, often with a raised surface, dry, crusty and no pigments were observed in the nutrient medium [7]. This species gave rise to the genus *Blastobotrys.* Conidia and conidiophores of this species are illustrated in Figure 2.

Figure 2 - Illustration of Conidia and conidiophores of Blastobotrys nivea [7].

B. nivea is closely related to *B. proliferans*, but they differ in the ability to ferment sucrose and grow on raffinose (*B. nivea* is unable to ferment sucrose and to grow on raffinose). *B. nivea* CBS 163.67 is the only strain described for this species and to the extent of our knowledge, there is no information about its biotechnological potential and clinical importance [6].

1.2.4 *Blastobotrys proliferans*

This species was initially isolated from a mite-infested nut, in Brazil by Marvanová, in 1976 [6,31]. *B. proliferans* is characterized by conidiophores that proliferate along a main axis.

The colonies are white, grow slowly at 24°C and are surrounded by a margin of submerged hyphae. Conidiophores are hyaline, erect or ascending, septate, producing 1-5 crowded conidiogenous cells (Figure 3A). When released, they leave short blunt denticles on the conidiophore apex arranged so that they seem to be the result of sympodial branching (Figure 3B). Conidiogenous cells are discrete, shortly pedicellate, ampulliform, with a highly refractive granular body in the middle. They bear up to 8 conidia, situated mostly laterally (Figure 3A, C). Its hyphae are thin, septate, branched, forming hyaline thick-

walled chlamydospores which may be spherical or irregular, terminal or intercalary, often in chains (Figure 3D) [31].

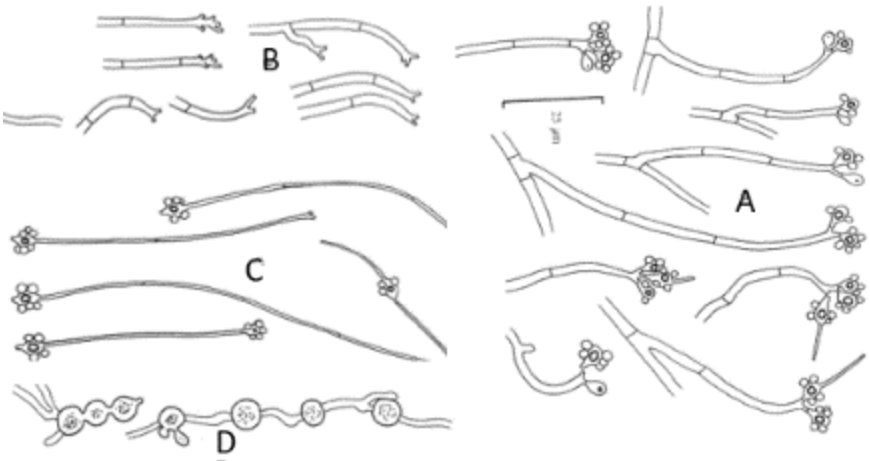

Figure 3 - Blastobotrys proliferans. A, Fertile conidiophores with conidiogenous cells at various stages of development. B, conidiophore tips after the liberation of conidiogenous cells; C, free conidiogenous cells bearing often fertile segments of conidiophore; D, chlamydospores [31].

This species, in terms of physiology is similar to *B. nivea* but it does ferment sucrose and is able to grow on raffinose. Despite that several strains of this species have been studied [6], there is no information about its biotechnological potential. In terms of clinical aspects, this species was associated to a peritonitis case.

1.2.5 *Blastobotrys adeninivorans*

B. adeninivorans was first described by Middelhoven *et al*. (1984), as *Trichosporon adeninivorans* CBS 8244 [6,32]. It was first isolated from soil, using an enrichment method, where uric acid or adenine were used as the sole source of carbon, nitrogen and energy [32]. This yeast caught the scientists' attention, since it showed an unusual biochemical activity, due to its ability to assimilate adenine and other purine compounds as a source of carbon and energy, showing potential to be used as a host for the production of recombinant proteins or other added-value compounds [33].

Later, in 1990, this species was assigned to the genus *Arxula*, proposed by Van der Walt, and composed by xerotolerant, ascomycetous, anamorphic, arthroconidial, nitrate positive yeasts [33]. The phylogenetic studies performed by Kurtzman and Robnett in 1998 showed that the species is a member of the clade that includes *Trichomonascus* and its anamorph *Blastobotrys* [6,29]. Consequently, these authors transferred *A. adeninivorans* to the genus *Blastobotrys* [6].

Nowadays, *B. adeninivorans* has gained importance in industry due to its unusual stress tolerance: it is thermotolerant (compatible with growth at high temperatures), xerotolerant (able to grow on dry environments), halotolerant (tolerates high salinity conditions) and osmotolerant (compatible with growth in high osmotic pressure environments) [13,14]. Also, due to its metabolic flexibility, this species is able to assimilate a wide spectrum of carbon and nitrogen sources, which includes not only conventional substrates such as glucose, xylose, and starch, but also rarely metabolized substances such as n-

butanol, tannic acid and protocatechuate [13,14]. These characteristics are very desirable for different biotechnological processes, like the production of added-value compounds from agro-industrial wastes, since these feedstocks are rich in mixtures of sugars not efficiently metabolized by the conventional yeast strains [5,14,15,21].

Besides, this yeast is used as a host to produce recombinant proteins and for the detection of estrogenic activity in various aqueous media because it tolerates a number of environmental parameters (thermotolerant up to 48°C, halotolerant up to 20% NaCl), and it can use a broad range of carbon and nitrogen sources which is both unique and beneficial in whole cell biosensors [13,34].

1.2.6 *Blastobotrys raffinosifermentans*

Gienow *et al*. (1990) found the strain LS3 (PAR-4) from wood hydrolysates in Siberia, with characteristics similar to *B. adeninivorans* CBS 8244 [13,35]. After studying its genetic and physiological properties, they confirmed it as a second strain of the species *Blastobotrys adeninivorans* [14,35]. Recently, and with the evolution of genetic analysis techniques, Thomas *et a*l. (2019) re-classified this strain and assigned it to the *Blastobotrys raffinosifermentans* species [36]. *B. raffinosifermentans* owns its name to its ability to ferment raffinose, which is a rare characteristic among species of the *Trichomonascus/Blastobotrys* clade [6].

LS3 is the most characterized *Blastobotrys raffinosifermentans* strain [13,14]. In 2014, the complete genome of this strain was sequenced [13] and phylogenetic studies were done, where the genome of 17 different species were compared to the one of the strain LS3. From the group of species chosen, *Yarrowia lipolytica* was the closest relative to *B.raffinosifermentans* LS3 (Figure 4) [13].

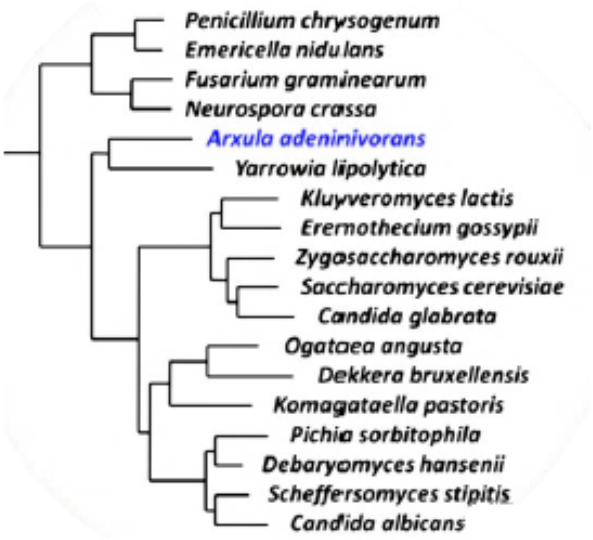

Figure 4 - Phylogenetic relationship of B. raffinosifermentans (previously known as Arxula or Blastobotrys adeninivorans) with other known yeasts [13].

Similar to its relative, *B. adeninivorans, Blastobotrys raffinosifermentans* also has interesting characteristics for biotechnology: it has an unusual metabolic flexibility that allows it to use a wide range

of carbon and nitrogen sources, it is also thermotolerant, xerotolerant and osmotolerant [13]. This yeast is currently explored as a biocatalyst for the synthesis of various biotechnological products such as tannases, 1-(S)-phenylethanol or β-D-galactopyranoside[13]. *B. raffinosifermentans* also contains 4 purine-degrading enzymes: adenine deaminase, guanine deaminase, xanthine oxidoreductase and urate oxidase, which simultaneously break down purines to a water-soluble 5-hydroxyisourate, so it is used to produce food with low purine content, important for patients with gout caused by accumulation of the purine uric acid [37].

Depending on the cultivation temperature, *B. raffinosifermentans* exhibits different morphological forms (<42 °C – budding cells; 42 °C – pseudomycelia; >42 °C – mycelial – Figure 5) and various post-translational modifications and protein expression profiles that are strongly correlated with the morphology. The dimorphism of *B. raffinosifermentans* is reversible, with temperature being the key factor [14].

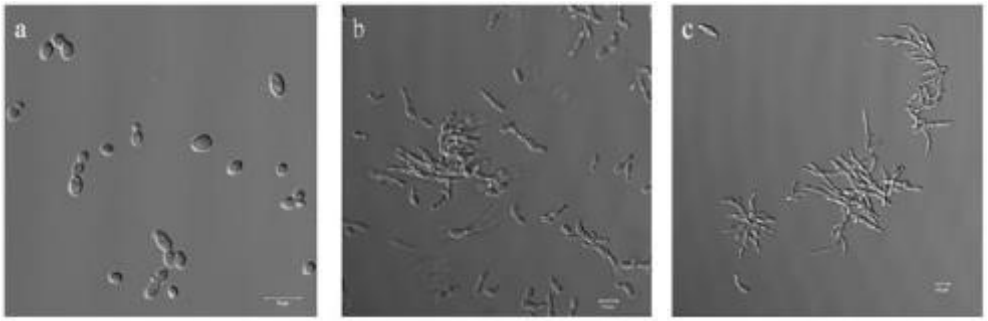

Figure 5 - Cell morphology of B. raffinosifermentans LS3, dependent on temperature: budding cells (a), pseudomycelia (b) and mycelia (c). The yeast cells were cultured in YPD medium for 18 h at 30 °C, 42 °C and 45 °C, respectively. [14].

1.2.7 *Blastobotrys meliponae*

This species was isolated from the honey of the bee *Melipona scutellaris* by Barbosa *et al.* (2016) and was collected in the Atlantic Forest in Brazil [11]. The isolate formed a clade with *B. proliferans,* based on phylogenetic analyses using only sequences of the D1/D2 domains. However, the sequences showed only 91% identity with the large subunit (LSU) sequences of that species in a BLASTn analysis, indicating that these isolates represented a new yeast species of the genus *Blastobotrys* [11].

Members of *Blastobotrys* with high similarity of the LSU rDNA D1/D2 domains to this species were: *B. attinorum* (GenBank GU373758; 92 %), *B. proliferans* (GenBank EF584541; 91 %) and *B. nivea* (GenBank DQ442690; 90 %) [11].

Blastobotrys meliponae differs from *B. nivea* in having sympodial branched conidiophores, lacking budding cells and chlamydospores, but with lateral conidia forming directly on the hyphae [11]. The species can be distinguished from *B. aristata* by the size of the conidiogenous cells (3–8 × 4.5–9 µm), conidiophore branching, number and size (100 µm) of setae, absence of lateral conidia formed directly on the hyphae, absence of chlamydospores and growth at 37ºC. *Blastobotrys proliferans* has a different branching of the conidiophores, conidiogenous cell size (3–4.5 × 4.5–7 µm), setae with a spathulate apex in older cultures and presence of distinct refraction bodies in the conidiogenous cells [11].

1.2.8 *Blastobotrys sp.* pathogenicity

Infections with opportunistic fungal pathogens are an increasing medical concern. These infections can have high mortality rates, which is highly dependent on early, accurate, and precise information from diagnostic testing [38]. In general, *Blastobotrys* species are not considered as pathogenic [6]. However, in recent studies, some species were isolated from samples of different patients, indicating that they might be contributing to the illness [38]. In the cases described, the strains seem to act as opportunistic pathogens, since the hosts had suffered from previous perturbations, like other disease, wound, medication, prior infection, immunodeficiency, and ageing [38–40].

In the study developed by Jean *et al*, 2019, the species *B. raffinosifermentans*, was isolated from a sputum specimen from a patient with a bronchopulmonary disorder [38]. In the case of *B. proliferans* and *B. serpentis*, their ability to infect mammals, and possible reptiles, is known. These two species cause invasive mycosis in humans [27]. *B. proliferans* was related to a case of peritonitis in a 46-year-old man undergoing peritoneal dialysis [6]. Despite this, the exact mechanisms behind the infection process are not known. Also, Scwharz *et al.* (2018) reported a case where *B. adeninivorans* was identified as causing invasive pulmonary mycosis and fungemia in a patient already affected by cystic fibrosis. This was the first time this species was described as a human pathogen [40].

1.2.9 Relevance of *Blastobotrys sp.* in the biotechnological industry

1.2.9.1 Biosensor

B. adeninivorans is resistant to high concentrations of sodium chloride in water which, together with its high tolerance to elevated temperatures, makes it possible to use some strains as biosensors for use in salty wastewater and other contaminated waters [14,34]. Strains of this species were used as hosts to express endocrine receptors together with a reporter gene, in order to detect specific molecules, like hormones in wastewaters, since due to the growth of pharmaceutical industry, the contamination of wastewaters with difficult molecules to remediate is also increasing. In 2016, there were 10 diagnosis kits based on the yeast *B. adeninivorans* [14,34].

1.2.9.2 Bioremediation

Blastobotrys sp. has a wide variety of enzyme activities that can be used for bioremediation processes. Huyben *et al*. (2018) showed that *B. adeninivorans* contains proteolytic enzymes that can be used to degrade contaminants from the environment or feedstock, like the prions present in organic wastes. This species can be used in the conversion of organic waste into high-quality protein and lipid sources for animal feeds and even human food, recapturing nutrients in a circular food production system [41].

Another study reported the ability of *B. raffinosifermentans* to use tannins and gallic acid as carbon sources. The concentrations of these two compounds are increasing in the environment, since they are widely used due to their antifungal and anticancer properties [42]. However, in certain concentrations, these compounds are toxic and hard to degrade, leading to its accumulation in water and soil, increasing

environmental pollution. So, since *B. raffinosifermentans* can grow on these compounds because it has fundamental enzymes, like the gallic acid decarboxylase [42], it can be used for bioremediation of this problem.

1.2.9.3 Cell Factories and Added-Value Products

Industry is constantly requiring new molecules and now there is the need to produce them in a sustainable way, to reduce costs and maximize the process efficiency. So, for many purposes, it is still important to search for alternative organisms, able to tolerate extreme environmental conditions such as high temperature fluctuations, media with high osmolarity, capable to grow on a wide variety of substrates and product accumulation to high levels [14]. As already discussed, *Blastobotrys* species, especially *B. raffinosifermentans,* fulfills these criteria.

Kunze *et al*. (2014) sequenced the strain *B. raffinosifermentans* LS3. They observed 914 introns within 6116 genes, turning this strain as the one of the richest in intron sequences in hemiascomycetes, to date. *Yarrowia lipolytica* was the closest species of *B. raffinosifermentans*, as a single outgroup, as observed in Figure 5 [13].

The genome and transcriptome analyses revealed enzymes with biotechnological potential such as glucoamylase, cutinase and two extracellular tannases involved in the tannic acid catabolic route and a novel pathway for n-butanol assimilation, which have remained unexplored in yeast, offering clues for further biotechnological developments. Due to the presence of new metabolic routes and other attractive characteristics already addressed, *B. raffinosifermentans* is a promising cell factory for many more applications and a useful gene donor [13,14]. In fact, it is already being used to produce some of these enzymes and the first protein industrially produced by this yeast was a recombinant tannase [14].

1.2.9.3.1 Antibacterial Activity

Many natural products are rich in preservative properties that can reduce microorganism's growth. The study published by Mat *et al.* (2016) showed that when using starter cultures of *Candida sp*. and *Blastobotrys sp*. in cocoa fermentation, the content of bioactive compounds, like γ-tocopherol, stigmasterol, and β-sitosterol, in cocoa beans was increased, hence increasing the antibacterial activity of the fermented bean extracts against *Bacillus cereus, Bacillus subtilis, Escherichia coli, Klebsiella pneumoniae, Proteus mirabilis, Salmonella enterica,* and *Staphylococcus aureus* [43]*.* The use of these strains might help in decreasing the problem of food spoilage and food poisoning [43]*.*

1.2.9.3.2 Toxins Degradation

Blastobotrys proliferans and *Blastobotrys muscicola* can biodegrade a trichothecene T-2 toxin, by the action of its acetyltransferases and glucosyltransferases, respectively. This toxin can cause cartilage lesions and chondrocyte necrosis in humans, and can be found in food and feedstuffs [44]. Therefore, these yeasts represent a good source of enzymes with an interesting potential for toxins removal from maize, wheat, barley, oat, and winter rye grain crops, avoiding human contamination.

1.2.9.3.3 Anti-Signaling Molecules Activity

Blastobotrys parvus showed a promising anti-Quorom Sensing (QS) activity against *P. aeruginosa*. *Pseudomonas* infection in cystic fibrosis patients is one of the major causes of morbidity [45]. The crude extract of *B. parvus* was rich in bioactive compounds that were able to reduce different virulence traits of *P. aeruginosa* PAO1 such as the production of pyocyanin, elastase, protease, chitinase, swimming and swarming motility, biofilm formation, exopolysaccharide production and alginate production at different sub-MIC concentrations [45].

1.2.9.3.4 Lipid Producer

Yeasts are attractive microorganisms to use in the production of sustainable oil from diverse renewable feedstocks. Despite of their great potential, there are still some drawbacks to their use in bulk production, being the major problem associated with the costs of the process [36]. Microbial lipids, also called single cell oils (SCO), have been intensively studied, since they have great potential for different biotechnology applications and for their attractive properties such as renewability and sustainability [46]. They have applications as food additives, pharmaceuticals, feed ingredients for aquaculture, biosurfactants, lubricants and oleochemicals, including fuels [46–48]. One of their characteristics is the fact that their fatty acid composition is like vegetable oils: are mainly composed of diacylglycerols (DAG), triacylglycerols (TAG) and sterol esters, that consist of long-chain fatty acids which have an important added value [46,49,50].

SCOs are produced by oleaginous microorganisms, which are those that accumulate over than 20% oil as a percentage of cell dry weight, since they can convert substrates such as carbon dioxide, sugars, and organic acids to oils [49]. Oleaginous yeasts generally synthesize and abundantly store lipids in lipid droplets under specific environmental conditions, one of which is nitrogen limitation [36]. However, proper identification of oleaginous strains is difficult, since SCOs production depends on the growth conditions, like the aeration rate, stage of growth, temperature, C/N ratio, carbon and nitrogen source and level, pH…[49].

Regarding the *Blastobotrys* genus, Malak *et al*. (2016) reported that C16:0 is the main fatty acid present in wild-type strains of *B. adeninivorans* (19.5–36.8% total fatty acids). C17:1 fatty acids were synthesized at up to 30.6% of total fatty acids and monounsaturated fatty acids accumulated to between 38.3% and 52.3% [14]. C18:1 shows different profiles in different strains and temperature conditions, and it ranges from 12,6% to being completely absent [14]. In another interesting study done by Thomas *et al*. (2019), the ability of *B. adeninivorans* and *B. raffinosifermentans* species to naturally produce and accumulate lipids from a set of different substrates was assessed. These species were the starting point since they exhibit a versatile metabolism, indicative of the presence of various degradative enzymes and pathways linked to central metabolism, exhibit also extremophilic traits, such as halotolerance, osmotolerance, and thermotolerance. Besides, in the case of *B. raffinosifermentans*, its genome has been already sequenced and engineered, and is closely related phylogenetically to *Yarrowia lipolytica*, a well-known lipid producer [36]. Thomas *et al*. (2019) found that several genes involved in lipid metabolism where

encoded in the *Blastobotrys raffinosifermentans* genome, like the cytoplasmic ATP citric acid lyase (ACL), which can efficiently convert respiratory citric acid to acetyl-CoA for fatty acid biosynthesis, is a signature of oleaginous yeasts. Effectively, strains of the genus *Blastobotrys* were able to produce and store lipids up to 45ºC. The strain LS3 was the best lipid-producer in comparison to the other strains tested [36]. Although it showed a temperature dependence for growth, glucose consumption, and lipid storage, its native performance in the tested range of 28 to 45 °C is a good starting point, ranging from 24 to 30% dry-cell weight, being 30ºC the optimal temperature for lipid production [36].

1.3. Riboflavin

Biotechnological processes are gaining prominence and in many cases are starting to replace the chemical ones. In order to allow the continuous growth of biotechnological industry, cellular factories able to convert renewable feedstocks into other products are extremely demanded and are under development. Also, with the growing concerns of consumers with the environment and their own personal health, industries face challenges in order to meet the increasing demand on sustainable products [51].

Riboflavin (RF) or Vitamin B_2 is an essential vitamin that belongs to flavins, which are yellow water-soluble-pteridine-based organic compounds derived from the isoalloxazine ring. This vitamin was discovered in 1879, as a yellow pigment from milk, but its chemical structure was only deciphered in 1930 [52,53]. It belongs to B-group vitamins, which act in synergy to maintain the body's homeostasis by playing major roles in metabolic processes [51]. RF is an obligatory component of cellular metabolism and is responsible for normal development, growth, reproduction, lactation, physical performance of well-being. Metabolically, RF is the precursor of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), both of which are indispensable as active groups in the majority of flavoproteins/flavo-coenzymes (Figure 6). These flavo-coenzymes play key roles in multiple crucial physiological functions, including redox homeostasis, protein folding, DNA repair, fatty acid β-oxidation, amino acid oxidation, and choline metabolism [51,52,54]. Moreover, riboflavin is part of a flavoprotein called cryptochrome, a photoreceptor in charge of the upkeep of the circadian clock [55].

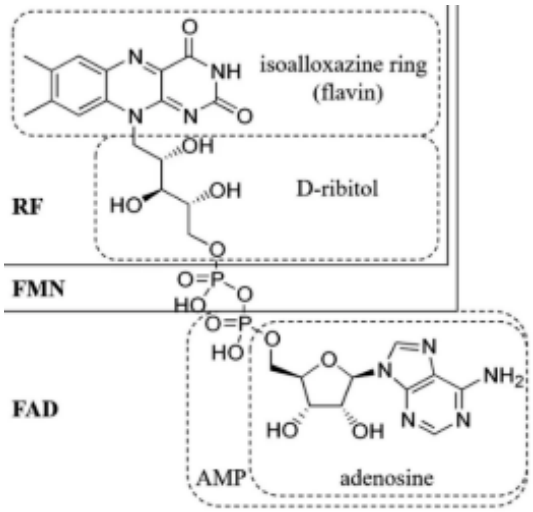

Figure 6 - Chemical structure of riboflavin (RF) and its derivates flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) [52].

Besides being present in milk and dairy products, RF is also found in green vegetables, eggs and meat and is required by bacteria, plants and animals [54]. Many microorganisms and plants possess the biosynthetic ability to produce RF. However, vertebrates, including humans, lack this ability and must obtain it from their diet [54].

In Western countries, RF daily intake comes from meat, fish, green vegetables, milk and dairy products, but is also commercially synthesized for nutritional use in the fortification of various food products like bread and breakfast cereals [51,54,56]. Another important role of this vitamin in industry is that is used as colorant (E-101), due to its bright yellow color [53,54]. Besides this and in contrast to the highly purified RF needed for multivitamin juices, colorant or pharmaceutical applications, more than 80% of the industrial production is used in a less pure form as animal feed additives [53,54].

Despite riboflavin's presence in a wide variety of foods, in many cases food processing causes the loss of this vitamin, causing the persistency of human RF deficiency, both in developing and industrialized countries. This deficiency can have many consequences: impaired vision, reduced growth rate, increased levels of homocysteine with consequent cardiac risk, pre-eclampsia, oxidative stress and anemia, liver and skin damage, sore throat, hyperemia, edema of oral and mucous membranes [51,52,54]. Therefore, it is important to produce this vitamin in order to be used as supplement and in the treatment of many diseases: for migraine and headache management, to combat progression of neurodegenerative diseases, can act as an adjuvant in cancer treatment [51,54]… Besides, RF is the precursor of the redox coenzymes, FMN and FAD, which are important biochemical reagents with significant application potential in the enzyme industry [52]. To avoid deficiency symptoms, the recommended daily intake for RF is 1,3 mg/day for men, 1,1 mg/day for women and between 1 and 4 mg/kg diet for animals [51,53]. The excess of RF supply, to humans, is harmless, because there is an efficient excretion via urine [51,53].

1.3.1 Riboflavin, FMN and FAD production in Bacteria and Fungi

Nowadays commercial vitamin B_2, is mainly used in the food, feed and pharmaceutical industries and is exclusively obtained by microbial fermentation [52]. RF is naturally synthesized by plants and many microorganisms and its biosynthesis pathways have been extensively studied both in the bacteria *B. subtilis* and in the filamentous fungi *Ashbya gossypii* [51,57], however for nearly 50 years, it was almost exclusively produced by chemical synthesis processes, which essentially consisted of six to eight chemical steps starting from glucose or ribose [52,56]. Despite achieving considerable productivity and purity, chemical processes involved the use of toxic agents and produced many waste products that required stringent environmental control and special forms of effluent treatment [56].

The emerging of competitive biotechnological processes with the use of microbial cell factories was a good solution in order to use more sustainable and low-cost processes. The paradigm of vitamins production, which have been traditionally produced by chemical synthesis, started to shift to the use of

bioprocesses [51]. Besides, this change offers several advantages: there is a cost reduction of the process, the waste generated and energy requirements are also reduced and renewable sources can be used as feedstock [51].

The hemiascomycetous filamentous fungus *Ashbya gossypii* and yeast *Candida famata* (teleomorph *Debaryomyces hansenii*) are naturally occurring overproducers of this vitamin. Despite being natural producers, several attempts to increase production have been made using mutagenesis and selection, achieving titers of 24,28 mg/g of biomass and 16,4 g/L, respectively [52,56]. The overproduction of vitamin B_2 by *A. gossypii* has been related to sporulation and oxidative stress conditions and is thought to represent an ecological advantage, by conferring the fungi protection against plant defenses [58]. To ensure that the Gram-positive bacterium *Bacillus subtilis* produces high titers of RF, steps of genetic engineering are required, like promoting the deregulation of purine synthesis and a mutation in a flavokinase/FAD-synthetase, achieving 15,7 g/L of titer [52,53].

The chemical reactions involved in the terminal biosynthesis process of RF by *B. subtilis* and *A. gossypii* are shown on Figure 7. In the bacteria, flavinogenesis is encoded by the RF or rib biosynthesis operon, encompassing five non-overlapping genes (*ribGBAHT*) [52]. This operon contains all the proteins necessary for the synthesis of RF, except for the ArPP phosphatase that was unknown. Recently, evidence was found that this step could be catalyzed by multiple enzymes from the haloacid dehalogenase (HAD) superfamily [52]. In the fungi, the process is encoded by *RIB1*, *RIB2*, *RIB7*, *RIB3*, *RIB4* AND *RIB5* [55]. The process in fungi and bacteria is similar, with exception of the reaction of deamination of the diamino pyrimidinone that occurs prior to the reduction of the ribosyl residue in bacteria, and the sequential order of these reactions is inverted in fungi [53,54].

This biosynthesis pathway contains a total of seven enzymatic steps starting from two different branches: the purine biosynthesis and the Pentose Phosphate (PP) pathway (Figure 7). The GTP cyclohydrolase, encoded by *RIB1* in *A. gossypii* and by ribA in *B. subtilis,* catalyzes the cleavage of GTP with release of formate. Then, a reduction reaction follows in the fungus, carried out by the gene product of *RIB7* (DARPP reductase), and a subsequent deamination (*RIB2*, DArPP deaminase). In *B. subtilis*, a bifunctional enzyme, encoded by ribG, catalyzes these latter two steps in reverse order. The phosphatase that cleaves 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5'-phosphate (ArPP) into 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione (ArP) is the only, less known, enzyme of the riboflavin biosynthetic pathway. DHBP synthase (*RIB3* and ribA in *A. gossypii and B.subtilis*) converts ribulose 5-phosphate to 3,4-dihydroxy-2-butanone 4-phosphate (DHBP). At this point, the two different branches of the riboflavin pathway merge into one. The condensation of DHBP and ArP is catalyzed by the lumazine synthase (*RIB4* for *A. gossypii*, ribH for *B. subtilis*) and yields one molecule of 6,7-dimethyl-8-ribityllumazine (DRL). In the final step of the pathway, riboflavin synthase (*RIB5* for *A.gossypii* and ribB for *B.subtilis*) converts two mole of DRL into one mole of riboflavin and one mole of ArP, which is recycled in the previous step [55]. RF can be further on transformed into FMN by a riboflavin kinase (*FMN1* for *A.gossypii* and ribR for *B.subtilis)* and FMN into FAD by FAD synthetase (*FAD1* for *A.gossypii* and ribC for *B.subtilis)* [52,59].

Despite of the large knowledge gathered about RF biosynthetic pathway, its regulation is poorly known, both in fungi and bacteria. In *B.subtilis*, it is believed that the pathway is regulated by a regulatory region in the rib operon (*ribO),* containing the main promoter P1 and two additional internal promoters, P2 and P3), but also by riboswitches: sequences that folds into a secondary structure and directly binds to the rib operon, repressing it [55]. In the case of *A.gossypii*, there is evidence that the pathway is upregulated as a result of growth or oxidative stress [55], however there is much to uncover still.This vitamin is one of the fewest B-vitamins biotechnologically produced [57]. Currently, the industrial strains used to produce RF are *Bacillus subtilis* and *Ashbya gossypii* strains genetically modified. In the past, the yeast *Candida famata* was also used, however due to the low stability of the strain the process was unprofitable and its use was abandoned [51,52]. In terms of small scale production, probiotic lactic acid bacteria, like *Lactobacillus plantarum* have been used for the development of dairy- and cereal-based functional foods for the in situ delivery of this vitamin to consumers [52,53].

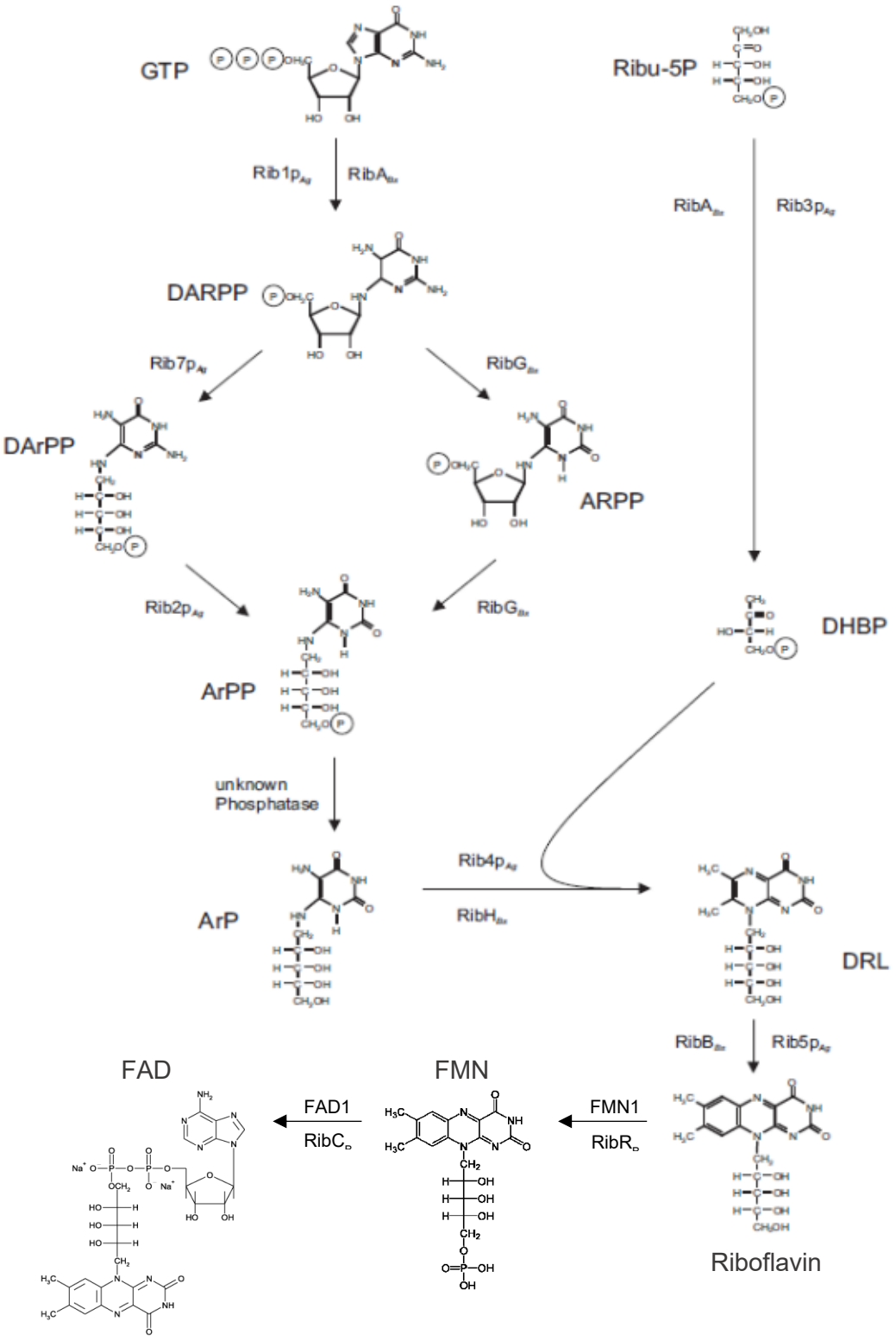

*Figure 7 - Riboflavin biosynthesis pathway in Bacillus subtilis (genes ribGBAHT, ribR and ribC) and in A.gossypii (genes rib1, rib7, rib2, rib3, rib4, rib5, FMN1 and FAD). In B. subtilis, RibA and RibG are bifunctional enzymes. In contrast to A. gossypii, deamination takes place before reduction in B. subtilis. In both organisms reduction of the ribosyl residue is NADPH-dependent. Enzymes catalyzing dephosphorylation of 5-amino-6-ribitylamino-2,4 (1H,3H)-pyrimidinedione-5'-phosphate (ArPP) were only recently characterized. GTP: Guanosine triphosphate, Ribu-5P: ribulose-5'-phosphate, DARPP 2: 5-diamino-6-ribosylamino-4 (3H)-pyrimidinone-5'-phosphate, DArPP: 2,5-diamino-6-ribitylamino-4 (3H)-pyrimidinone-5'-phosphate, ARPP: 5-amino-6-ribosylamino-2,4 (1H,3H)-pyrimidinedione-5'-phosphate, ArP: 5-amino-6-ribitylamino-2,4 (1H,3H)-pyrimidinedione, DHBP: 3,4-dihydroxy-2-butanone 4-phosphate, DRL: 6,7-dimethyl-8-*ribityl-lumazine, *FMN: Flavin Mononucleotide, FAD: Flavin Adenine Dinucleotide* [53,59].

As already said, plants, fungi and most bacteria, are able to produce RF. Its production starts with guanosine triphosphate (GTP), which is formed in the purine biosynthesis, and ribulose 5-phosphate (Ru5P) originating from the pentose phosphate (PP) pathway (Figure 7,8) [52,60]. According to the microorganism, different substrates can be used. Fungi like *A. gossypii* prefer oil as substrate, whereas bacteria such as *B. subtilis* prefer to use carbohydrates, like glucose [52,53,60].

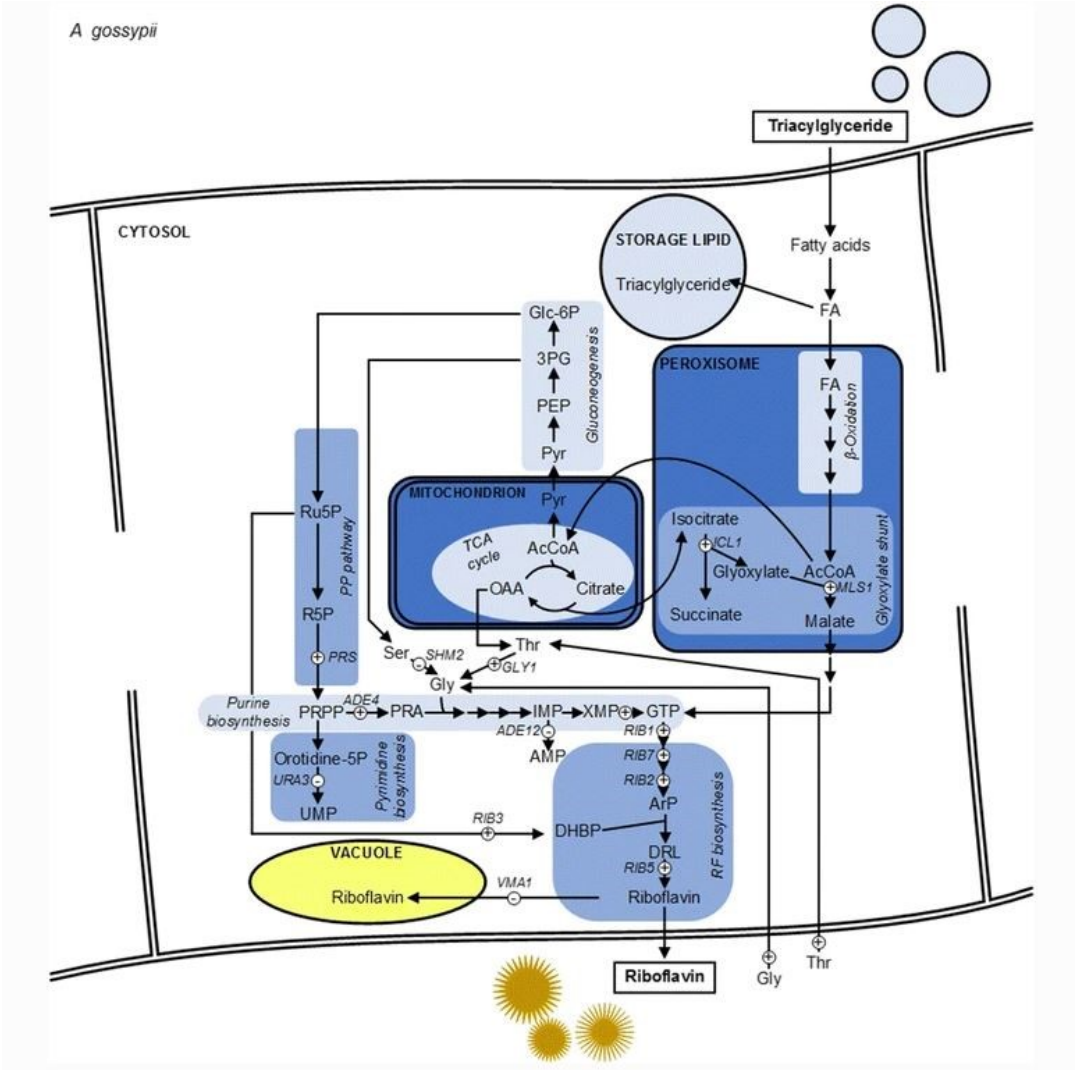

Figure 8 – Riboflavin biosynthesis in Ashbya gossypii. The symbols "+" and "–" indicate increased or decreased fluxes or activities both with positive effect on riboflavin biosynthesis. 3PG 3-phosphoglycerate, AcCoA acetyl-coenzyme A, AMP adenosine monophosphate, ArP 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione, DHBP 3,4-dihydroxy-2-butanone 4-phosphate, DLR 6,7-dimethyl-8-ribityl-lumazine, FA fatty acids, Glc-6P glucose 6-phosphate, Gly glycine, GTP guanosine triphosphate, IMP inosine monophosphate, OAA oxaloacetate, PEP phosphoenolpyruvate, PP pathway pentose phosphate pathway, PRA 5-phosphpribosylamine, PRPP phosphoribosylpyrophosphate, Pyr pyruvate, R5P ribose 5-phosphate, RF riboflavin, Ru5P ribulose 5-phosphate, Ser serine, TCA cycle tricarboxylic acid cycle, Thr threonine, UMP uridine monophosphate, XMP xanthosine monophosphate, ADE4 PRPP amidotransferase, ADE12 adenylosuccinate synthase, GLY1 threonine aldolase, ICL1 isocitrate lyase, MLS1 malate synthase, PRS PRPP synthetase, RIB1 GTP cyclohydrolase II, RIB2 DArPP deaminase, RIB3 DHBP synthase, RIB5 riboflavin synthase, RIB7 DARPP reductase, SHM2 serine hydroxymethyltransferase, URA3 orotidine 5-phosphate decarboxylase, VMA1 vacuolar ATPase subunit A [55].

Especially in *A. gossypii*, the production of this vitamin is a complex and multi-compartmental process. The oil substrate is cleaved into fatty acids and glycerol, by an extracellular lipase. Are then transported into the cell and oxidized into acetyl-CoA via the β-oxidation pathway, located in the peroxisome. Acetyl-

CoA is then metabolized via the glyoxylate cycle, the tricarboxylic acid (TCA) cycle in the mitochondria, gluconeogenesis, the PP pathway, the purine pathway and the terminal RF biosynthesis in the cytosol (Figure 8) [52,53,55,60].

Metabolic engineering of yeast strains has been done in order to improve RF biosynthesis by obtaining overproducing strains. This is accomplished by overexpression of genes in the relevant pathways, suppression of competing pathways, and disruption of regulatory genes responsible for feedback inhibition [52].

1.3.2 Microbial Cell Factories Used to Produce Flavins

As discussed previously, RF production is obtained by microbial fermentation and the industrial strains used are genetically modified strains of *B.subtilis* and *A.gossypii* [52], achieving titers of 26-30 g/L in normal industrial processes [52]. Still, engineered techniques to produce RF have also been applied to other species: *C. famata*, *C. ammoniagenes*, *E. coli*, *Pichia guilliermondii* and *Eremothecium gossypii* (*E. ashbyi*) [52].

Overproducing strains of RF, FMN and FAD were developed by combinations of mutagenesis, to achieve resistance against different antimetabolites (structural analogues of RF, inhibitors, antagonists…) and several rounds of genetic engineering to improve cells' metabolism [52]. Metabolic engineering allows rewiring the metabolism of cells to enhance production of a metabolite, endogenous or not. Different strategies can be used according to the metabolic pathway chosen to modify. In RF synthesis the central carbon metabolism, the synthesis of glycine, cell-scale optimization, and so on, are all candidates to be tested, since all can affect the main pathway [52].

In general, RF overproducing strains were obtained by overexpressing the synthetic pathways of RF or its precursors, like the pentose phosphate pathway or the purine biosynthetic pathway, which was accomplished by the combination of direct gene duplication, replacement of the native promoter with a strong one, disruption of competing pathways, or modification of regulatory genes [55]. Other strategies focused on the improvement of the host traits, for example by reducing the maintenance metabolism and overexpression of Heat Shock Proteins, that lead to an increase of over 23% of the titers [52].

In addition, in order to enhance the biosynthesis of this vitamin, supplements can be added to the medium. Addition of precursors of GTP, the first metabolite of the pathway, increases productivity. The addition of glycine or its overproduction by the cell was also found to increase productivity by ninefold in *A. gossypii* [52,53]. Glycine can act as simply a limiting precursor or can also act as an inducer of gene transcription. Besides, Stahman *et al.*, (2000) found that this biosynthesis in *C. famata* is negatively affected by iron [53].

These metabolic systems used are at early stages of development and thus, the capacity of these microorganisms to produce RF is expected to be boosted in the next years, together with the improvement of techniques and genome-scale models important for metabolic engineering processes [56]. Besides focusing only on the pathway of the vitamin production, different targets can be pursued by

genetic engineering. For example, it is also important the use of strains with a broad substrate range, enabling it to produce the vitamin from cheaper carbon sources such as lignocellulosic material or agro-industrial residues, strains with increased robustness to reduce the need of substrate sterilization or even to permit open-air fermentations, and strains with increased temperature resistance to diminish the costs of refrigeration in the bioreactor and save time for the cooling of the feeding streams [56].

1.4 Unveiling strains potential through genomics

The model yeast *S. cerevisiae* has established much of our understanding of eukaryotic biology, however, a single yeast is not representative of all the others [61]. Thus, there are many important biotechnological applications and highly divergent physiological capabilities of lesser-known yeast species that are yet to be discovered and explored. The use of non-conventional yeasts that contain already a natural capacity to produce a given product of interest can facilitate the process of strain optimization, since most of them need to suffer only straightforward metabolic engineering techniques, to obtain higher yields, titers and production rate [62].

The access to a yeast genome is paramount to increase and optimize our knowledge of key genes and pathways of industrial relevance, fungal infection as well as host-pathogen interactions [63]. The evolution of DNA sequencing techniques and genomics is facilitating the prediction of these biochemical traits with industrial interest [61]. The main advances on these techniques, comparing to the initial methods used, is their higher throughput and lower costs per base [64]. These new techniques, called Next-Generation Sequencing (NGS), use massively parallel sequencing to generate up to several gigabases of sequence information per day, making possible new applications that were previously difficult to perform. One example of NGS technique is Illumina. This technique initiates with an amplification step by bridge PCR, where templates of the DNA are made, then sequencing occurs by reversible termination, where nucleotides labelled with fluorophores bind to the template, identifying the correct nucleotide per position [65]. Illumina can be used in different sequences devices [64]. The HiSeq X series, with very high throughput and small cost ($0.001 per 1000 bases), is used for the study of whole-genomes and has an output range of 125-1500 Gb, the run time is 1-3.5 days, it has 2.5 billion reads per flow cell, and the maximum read length is 2 x 150 bp. The data that results of this process requires great power of analysis from the bioinformatic systems. A problem with this technique is the fact that the reads are relatively short, hampering the assembling stage [64].

NGS has applications in most fields of science related to DNA studies. It can be used for food control, for the detection of viral components in vaccines and other medications, to study the evolution of organisms, analysis of fetus DNA from the mother's blood, in the forensic field, among others. Also, the diagnostics field has evolved with these techniques, resulting in a better understanding of diseases and drug efficiency [64].

The complementation of biotechnology, that establishes the tools necessary to modify genes, with genomics, that provides a new platform for the high-throughput genetic analysis, has potential to increase the information known on gene expression, metabolic pathways, protein levels, subcellular

localization yielding a ‘genomic understanding’ of how yeast metabolizes, grows, and reproduces, but also to improve industrially relevant strains [63]. Genomics is based on genetic mapping, which provides the location of genes and association to their function and DNA sequencing [66]. The era of eukaryotic genomics started in 1996, with the complete sequencing of *S. cerevisiae*’s genome, and further evolved with the collaborative work of different labs to comprehend and study eukaryotic gene expression and function [63]. Completion of genomes sequences and the resultant data has allowed to gain insights into the general functions required for eukaryotic life, explore the overall genetic composition and do genomic comparison of the medically, industrially, and environmentally important fungi and yeasts. In the past few years, there has been an impressive rise in fungal genomics, expanding the understanding of the genetic, physiological, and ecological diversity of the organisms studied [67]. Besides genomics, other “omics” technologies are evolving to study different molecules: mRNA molecules (transcriptomics), proteins (proteomics), and metabolites (metabolomics), which allow to characterize entirely the activity of a biological system through mathematical modeling [66].

Genome-mining, which involves the analysis and identification of previously uncharacterized natural product biosynthetic gene clusters within the genomes of sequenced organisms and the enzymes encoded by it [68], is indicating that the capability of fungi to produce secondary metabolites has been substantially underestimated because their gene clusters are silent under standard cultivation conditions. Bioinformatic algorithms such as SMURF, antiSMASH, and FungiFun allow identification of secondary metabolism gene clusters [67]. By combining genomic and biochemical information it is possible to predict the production of a compound and also target genes of importance for heterologous pathway engineering in order to produce natural and novel compounds with potential new bioactivities [67].

The understanding of the unique metabolisms and physiologies of specific non-conventional yeast has become possible thanks to the increasing availability of next generation sequencing, genome editing tools, and the development of system wide genomics studies. With these new tools, it will be possible to extend the knowledge obtained by the study of the model yeast, *S. cerevisiae* to other eukaryotic organisms, maximize the desired phenotypes by genome engineering and metabolic re-wiring of genes, and increase productivity to reach industrially relevant production yields of new products [62,63].

2. Materials and Methods

2.1 Yeast Isolation and Culture Conditions

The yeasts used in this work were isolated from different soil samples (soil underneath an olive tree from Ferreira do Alentejo (38°02'43.6"N 8°06'34.3"W, olive tree soil 1), soil underneath an olive tree (olive tree soil 2) and an oak tree from Póvoa-Cadaval, soil from a vegetable garden from Póvoa-Cadaval, and soil from Berlengas and Arrábida), from fruits and vegetables (plum, walnut green husk, physalis, juniper berries, cabbage, olives, olives curing water), and from surface seawater (coast of Arrábida and Berlengas). The work of isolation and identification of non-conventional yeasts was performed in collaboration with my colleague of the M.Sc. in Biotechnology Margarida Pataco.

2.1.1 Differential Isolation

To isolate microorganisms from marine samples the method used was the one described by Zaky *et al.* (2016), with minor modifications [69]. 500mL, 1L, 2L and 3L of water were filtered using a metallic filtration system (SS Filter Holder 100ml 47mm, Merck), and all the media involved was prepared using this filtered marine water, to assure the osmotic needs of yeasts isolated from this environment. The resulting filter was inoculated in 500 mL of YPDX medium (30 g/L glucose (Scharlau), 30 g/L xylose (Sigma-Aldrich), 3 g/L malt extract (Sigma-Aldrich), 3 g/L yeast extract (ThermoFisher), 5 g/L peptone (ThermoFisher), 1 g/L $(NH_4)_2SO_4$ (Panreac) and 0.25 g/L KH_2PO_4 (Panreac), pH 5.0), supplemented with chloramphenicol (100 µg/mL) and incubated at 30ºC, 150 rpm, 48h (Primary Enrichment). Then, 20mL from Primary Enrichment were added to 180 mL of YPDX medium, and incubated again at 30ºC, 150 rpm, 48h (Scale-Up Enrichment). To differentiate yeasts with the ability to grow in different carbon sources a differential enrichment step was performed, where 10 mL from Scale-Up Enrichment was added to 90 mL of YPD (60 g/L glucose, 3 g/L yeast extract, 5 g/L peptone, 1 g/L $(NH_4)_2SO_4$ and 0.25 g/L KH_2PO_4, pH 5.0) or to YPX medium (60 g/L xylose, 3 g/L yeast extract, 5 g/L peptone, 1 g/L $(NH_4)_2SO_4$ and 0.25 g/L KH_2PO_4, pH 5.0), and incubated at the same conditions as before. After 48h, the samples were diluted in 0,085% NaCl solution and poured to YPX Agar (60 g/L xylose, 3 g/L yeast extract, 5 g/L peptone, 1 g/L $(NH_4)_2SO_4$ and 0.25 g/L KH_2PO_4, 20 g/L agar (NzyTech)) or YPD Agar (60 g/L glucose, 3 g/L yeast extract, 5 g/L peptone, 1 g/L $(NH_4)_2SO_4$ and 0.25 g/L KH_2PO_4, 20 g/L agar) and incubated at 30ºC. After 48h, it was possible to observe isolated yeast colonies in several plates. Isolated yeast colonies were streaked into new agar plates to assure the purity of the isolate.

Regarding soil samples, one table spoon of the soil (~1g, contains 10^3 yeast cells [70]) was used as inoculum. In the case of the fruits and vegetables, they were covered with sterile demineralized water, vortexed for 1 min and 1 mL of this water was used as inoculum. In the case of plum (5 mL), physalis and juniper berries (10 mL) a higher volume was used as inoculum, since no growth was observed with 1 mL. In the sample olives' curing water, 0.5 mL of the sample was directly used as inoculum. Afterwards, the procedure used was similar to the one used for marine samples [69], with the difference that the water used to prepare the media was demineralized and the volumes used were different:

primary enrichment was done in 50 mL YPDX, 5 mL of this enrichment were added to 45 mL YPDX (Scale-up) and differential enrichment was done by adding 5 mL from Scale-up to 45 mL of YPD or YPX. All the colonies isolated were streaked into new agar plates to guarantee the purity of isolate. Pure yeast cultures were frozen at -80ºC in appropriate media supplemented with 15% glycerol (v/v) (Sigma) and an IST number ID was attributed, and they were added to the yeast collection of IST.

2.1.2 Direct Isolation

Cabbage and walnut samples were directly isolated in Wallerstein Nutrient Agar (WLN) medium. Samples were covered in sterile demineralized water, vortexed for 2 min and 100 µL of the resulting water was inoculated into WLN agar medium (Yeast extract 4 g/L, Tryptone 5 g/L, Glucose 50 g/L, Potassium dihydrogen phosphate 0.55 g/L, Potassium chloride 0.425 g/L, Calcium chloride 0.125 g/L (Merck), Magnesium sulphate 0.125 g/L (Sigma), Ferric chloride 0.0025 g/L (Chem-Lab), Manganese sulphate 0.0025 g/L (Sigma-Aldrich), Bromocresol green 0.022 g/L (Sigma-Aldrich), Agar 15 g/L), at pH 5.5, supplemented with chloramphenicol (100 µg/mL) and incubated at 30ºC for 48 hours. All of the colonies isolated were streaked into new agar plates to guarantee the purity of isolate. Pure yeast cultures were frozen at -80ºC and an IST number ID was attributed, and they were added to the yeast collection of Instituto Superior Técnico.

In the samples where no yeast growth was observed after the differential isolation (Arrábida sea, juniper berries, physalis, garden soil and olive tree soil 1), 100 µL from Primary Enrichment were also directly inoculated in WLN agar and incubated at 30ºC for 48 hours. Isolated yeast colonies were streaked into new WLN agar plates to assure the purity of the isolate.

2.2 Identification of yeast strains

2.2.1 DNA extraction

To identify the isolated strains, genomic DNA was extracted from the samples and sequenced. Isolates of beer must, IST400b, IST453, IST454, IST455 and IST456, previously isolated at the Biological Sciences Research Group (BSRG) laboratory by Dr. Margarida Palma and were kept at −80°C in appropriate media supplemented with 15% glycerol (v/v) (Sigma), were also studied. The cells were previously grown in solid media and colonies were picked and resuspended in 200 µl of sterile demineralized water in microcentrifuge tubes, approximately 3 spoons of 0.5 mm glass sterile spheres and 300 µl of phenol (Amresco) were added and vortexed for 2 minutes, and then centrifuged 5 minutes at 14000 rpm, 4ºC (ScanSpeed 1730R, LaboGene). The supernatant was transferred to a new microcentrifuge tube and 300 µl of phenol were added and vortexed for 10 minutes, followed by centrifugation at 4ºC, 14000 rpm. After discarding the lower phase, 300 µl of ether were added and vortexed for 20 seconds, then centrifuged for 5 minutes (14000 rpm, 4ºC). One mililiter of absolute ethanol (Carlo Erba) at -20ºC was added only to the lower phase and incubated for over 20 minutes at -20ºC. Afterwards, samples were centrifuged for 15 minutes (14000 rpm, 4ºC) and the resulting pellet was washed with ethanol 70% (v/v), centrifuged for 5 minutes (14000 rpm, 4ºC) and after discarding the

supernatant, the pellet was vacuum dried for 15 minutes. Finally, the yeast total DNA was resuspended in 100 µl of sterile water and quantified using the NanoDrop ND-1000 spectrophotometer (Alfagene) before being stored at -20ºC.

2.2.2 Polymerase Chain Reaction (PCR) and gel electrophoresis

The DNA obtained was amplified by Polymerase Chain Reaction (PCR) using two possible sets of primer pairs. To amplify the region D1/D2 of ribosomal DNA the primers NL-1 (5'- GCATATCAATAAGCGGAGGAAAAG-3') and NL-4 (5'- GGTCCGTGTTTCAAGACGG- 3') [71] were used and for the Internal Transcribed Spacer (ITS) region of ribosomal DNA was amplified using ITS-1 (5'- TCCGTAGGTGAACCTGCGG- 3') and ITS-4 (5'- TCCTCCGCTTATTGATATGC- 3') [72]. PCR was performed in 50 µl reaction volume, composed of 29.5 µl of water, 10 µl of Enzyme Buffer (New England BioLabs), 1 µl dNTPS (NZYtech), 2.5 µl of each primer 10µM (StabVida), 1.5 µl of DMSO (Dimethyl Sulfoxide, New England BioLabs), 2.5 µl of DNA and 0.5 µl of Phusion enzyme (New England BioLabs). The first denaturation was performed at 98ºC, for 30 seconds, followed by 35 cycles consisting of 10 seconds of denaturation at 98ºC, 20 seconds of annealing at 52ºC and 30 seconds of extension at 72ºC. A final extension step was performed for 8 minutes at 72ºC.

A total of 8 µl of loading buffer 6X (Takara) were added to the amplified products and the mix was loaded on a 0.8% agarose gel (1X TBE buffer) (NZYTech) with green safe and separated for about 1 hour by electrophoresis at 110V, 400mA, 100W. NZYladder III (NZYTech) was used for comparison of the DNA fragment sizes. The resulting bands were extracted and purified using the NZYGelpure purification kit (NZYTech) and afterwards, quantified in the NanoDrop spectrophotometer.

2.2.3 DNA sequencing

After purification and quantification of the DNA samples, they were sequenced using Sanger sequencing, at STAB VIDA. The resulting sequences were compared to homologous sequences deposited in the NCBI database, using the Standard Nucleotide BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) in order to taxonomically identify the isolated yeast species. The software SnapGene v.4.3.11 was used to access the quality of the sequences obtained.

2.3 Phenotypic and Physiological Characterization of *Blastobotrys solioliva* IST508

2.3.1 Phylogenetic analysis

The sequences of the ITS region and D1/D2 domain of the yeast isolate IST508, previously amplified and sequenced, were submitted for BLAST search in the NCBI database. Sequences of related species from *Blastobotrys, Candida, Trichomonascus, Sympodiomyces, Sugiyamaella, Zygoascus, Wickerhamiella, Trigonopsis, Sporopachyadermia and Schizosaccharomyces* genera were retrieved from the GenBank and aligned iteratively by using the multiple alignment tool CLUSTAL X, from the

software MEGA-X v.10.0.5. This software was also used to phylogenetically place the new species sequence, using the evolutionary distance data using the two-parameter model and the maximum likelihood method. The robustness of the clades was assessed using bootstrap analysis with 1000 replicates [9].

2.3.2 Morphological observations

Morphological characteristics were observed by the stereomicroscope and light microscopy after culture on malt extract agar ME (15g/L malt extract, 20 g/L agar) and YPD agar, after 3 and 7 days of incubation at 30ºC, as described by Kurtzman *et. al* (2011) [6]. Temperature effect on growth was assessed by growing the cells on ME and YPD agar at 25 ºC, 37 ºC and 40ºC, and the results were observed after 3 days of incubation. Cells were also plated on 50% glucose agar (500 g/L glucose, 10 g/L yeast extract, 20 g/L agar) to assess the osmotic tolerance of the species, and the results were observed after 7 days.

To analyse the sporulation ability, yeast cells were streaked, at first, into a pre-sporulation medium (0.8% yeast extract, 0.3% peptone, 10% glucose, 2% agar) and incubated at 30 ºC for 1 day. Second, yeast cells were transferred to the sporulation medium (0.5% potassium acetate and 2% agar, pH 6.5-7.0) [6]. The evolution of the morphology and sporulation of cells was observed after 3 and 7 days of incubation.

2.3.3 Calcofluor Staining

In order to confirm the reproduction method of *Blastobotrys solioliva* IST508, cells were stained for 5 min with 5 µl of a stock solution of 1% Calcofluor White, then centrifuged at 8000 rpm, 3 min, and washed with PBS 1X (Phosphate Buffered Saline, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) [73]. Cells were grown on Pre-Sporulation Media, Sporulation Media and YPD, and observed after 2, 4 and 7 days of incubation at 30ºC. All the pictures were taken with KL2500 LCD (ZEISS) Stereomicroscope or with ZEN – Digital Imaging for Light Microscopy (ZEISS) connected to an optical microscope Hitech Zeiss Axioplan with one camera Axiocam 503 colour, using a filter number 5 (Calcofluor White: excitation at 440nm and emission at 500-520 nm) and power supply 12V 100W.

2.3.4 Assimilation Tests

The assimilation tests were performed as described by Kurtzman *et al*. (2011) with minor modifications [6]. A pre-culture of *Blastobotrys solioliva* IST508 was prepared in 50 mL YNB (6.7 g/L Yeast Nitrogen Base, Difco), with 20 g/L of glucose and incubated for 24h. Afterwards, sterile glass tubes with 3 mL of YNB with 20 g/L of each compound (stock concentration of 100g/L) were inoculated with the pre-culture at an initial OD_{600nm} = 0.1. Compounds tested were fructose (Sigma-Aldrich), glucose (Merck), galactose (Sigma-Aldrich), ribose (Sigma-Aldrich), melezitose (Sigma-Aldrich), trehalose (Merck), sucrose (Sigma-Aldrich), xylose (Sigma-Aldrich), arabinose (Sigma-Aldrich), maltose (Sigma-Aldrich), galacturonic acid (Sigma-Aldrich), raffinose (Alfa-aesar), rhamnose (Merck), mannose (Sigma-Aldrich), myo-inositol (Merck), sorbitol (Sigma-aldrich), succinic acid (Fisher-scientific), mannitol (CalbioChem), xylitol (Sigma-aldrich), soluble starch (BDH chemicals), melibiose (Sigma-Aldrich), lactose (Sigma-Aldrich), inulin (Sigma-Aldrich), galactitol (Sigma-Aldrich, sorbose (Sigma-Aldrich, rybitol (Sigma-

Aldrich), ethanol 5% (v/v) (Carlo Erba), methanol 5% (v/v) (VWR chemicals), glycerol 5% (v/v) (Sigma-Aldrich), ethanol 1% (v/v), methanol 1% (v/v), glycerol 1% (v/v). Vitamin requirements of the yeast were tested by inoculating a *Blastobotrys solioliva* IST508 pre-culture prepared as described before into a sterile glass flask with 3 mL of Vitamin-Free Yeast Base (5g/l ammonium sulfate (Panreac), 10g/L glucose, 10 mg/L L-Histidine monohydrochloride (Acros), 20 mg/L DL-Methionine (Sigma), 20mg/L Tryptophan (Merck), 0.85g/L Potassium phosphate monobasic, 0.15g/L potassium phosphate dibasic, 0.5g/L magnesium sulfate (Riedel-de-Haen), 0.1g/L sodium chloride, 0.1g/L calcium chloride, 500 μg/L boric acid (Merck), 40 μg/L copper sulfate (Merck), 100 μg/L potassium iodide (Sigma-Aldrich), 200 μg/L ferric chloride, 400 μg/L manganese sulfate (Sigma-Aldrich), 200 μg/L sodium molybdate (Sigma-Aldrich), 400 μg/L zinc sulfate (Panreac)) [6]. Osmotic tolerance was assessed by growing a *Blastobotrys solioliva* IST508 pre-culture prepared as described before in 3mL of 10% Sodium Chloride + 5% glucose (10% sodium chloride + 5% glucose media (100 g/L NaCl, 6.7 g/L Yeast Nitrogen Base, 50 g/L glucose) [6]. After 7 days of incubation at 30ºC, 250 rpm, optical densities at 600 nm were measured. A total of 3 replicates were performed for each assay, with exception to inulin, since the solution formed crystals after the first replicate, and methanol 5%, ethanol 5% and glycerol 5% assimilation, where only one replicate was done. The mean and standard deviations were calculated by GraphPad Prism 6, for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

2.3.5 Fermentation Tests

Fermentation sterile flasks with 16 mL of Fermentation Basal Media (4.5g/L yeast extract, 7.5g/L peptone, 0.0267 g/L bromothymol blue) and 4 mL of sugar (glucose, fructose, trehalose, galactose, maltose, xylose, raffinose, arabinose and sucrose) stock solution (100 g/L) were inoculated at an initial OD_{600nm} = 0.1 with a pre-culture prepared with approximately 7 hours of incubation at 30 ºC and 250 rpm, in 50 mL of YM (yeast extract-malt extract-peptone-glucose) liquid medium (3 g/L of malt extract, 3 g/L yeast extract, 5 g/L peptone and 10 g/L glucose (ThermoFisher)). All flasks were capped with a perforated rubber cork and a needle was inserted in the middle of the cap to allow evaporation of gases, before incubation at 30 ºC without agitation for ~17 days. Fermentations were monitored daily by weighting the fermentation flasks and plotting weight loss over time. Three replicates were performed for each assay and the mean and standard deviations were calculated by GraphPad Prism 6, for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

2.3.6 Growth Analysis, Sugar Consumption and Determination of Extracellular Metabolites Produced by HPLC

Blastobotrys adeninivorans PYCC 4638*, Blastobotrys proliferans* PYCC 5116 and *Blastobotrys solioliva* IST508 were initially grown in 50 mL of YNB with 100 g/L of xylose or glucose. These cultures were grown overnight, at 30ºC and 250 rpm, and used to inoculate at an initial OD_{600nm}=0.1. Erlenmeyer flasks with 50 mL YNB containing 20g/L of carbon source (glucose or xylose). These flasks were incubated at 30ºC, 250 rpm, for 145 h. The growth of the yeast cultures was assessed by taking daily measures of the OD_{600nm} of the culture media. In addition, 500 μL of culture media at each point was taken, in order to analyze the metabolites produced and sugar consumption by High-Performance Liquid

Chromatography (HPLC). These samples were centrifuged (8000 rpm, 3 min) and the supernatant was collected and frozen at -80ºC.

These supernatants were diluted 10 times in 0.005 M H_2SO_4 in HPLC vials. Samples were analyzed by the HPLC system (Merck Hitachi, Darmstadt, Germany) with a refractive index detector (Hitachi LaChrom Elite L-7490) and a UV-Vis detector (Hitachi LaChrom Elite L-2420), and an Aminex HPX-87H column (300 mm × 7.8 mm, Bio-Rad) eluted at 65°C with 0.005M H_2SO_4 at a flow rate of 0.6 mL/min. The sugar concentrations (glucose and xylose) were determined, as well as the concentrations of glycerol and ethanol. Also, the concentrations of oxalic acid, citric acid, malic acid, succinic acid, formic acid and acetic acid were determined for each sample. Concentrations were estimated based on calibration curves previously prepared and three replicates were performed for each strain tested. The data was analyzed using GraphPad Prism 6, for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

2.3.7 DNA Content Analysis by Flow Cytometry

The procedure taken to determine the DNA content of *Blastobotrys solioliva* IST508*, Blastobotrys adeninivorans* PYCC4638*, Blastobotrys proliferans* PYCC5116 was similar to the one described by Fortuna *et al*. (2000) [74]. Pre-cultures of cells of *Blastobotrys solioliva* IST508*, Blastobotrys adeninivorans* PYCC4638*, Blastobotrys proliferans* PYCC5116*, Saccharomyces cerevisiae* BY4741 (haploid strain) and *Saccharomyces cerevisiae* BY4743 (diploid strain) [75], were grown overnight in YPD growth medium, at 30ºC, 250 rpm. It was assumed that an OD_{600nm}=1, was equivalent to 10^7 cells, so they were harvested and centrifuged at 8000 rpm, for 3 min. Then, cells were washed twice with ice-cold deionized, distilled water, and resuspended in ethanol 70% and left overnight at 4ºC. Afterwards, cells were washed with 50 mM sodium citric acid buffer, pH 7.5 and resuspended in 850 µL of the same buffer and 250 µL of RNAse A (NZYTech) (1mg/mL) were added and incubated for 1 h at 50 ºC. After, 50 µL of proteinase K (NZYTech) (20mg/mL) were added and the mix was incubated again for 1 h at 50ºC. The cell suspension was stained with SYBR GREEN I (Thermofisher) at a final concentration equivalent to 500-fold dilution of the commercial solution and incubated overnight at 4ºC. Triton X – 100 was added at a final concentration of 0.25% (v/v), followed by vortex and sonication of the samples in Branson Sonifier 250 (3-4 pulses with an output power of 3 and 30% duty cycle) and analyzed in BD AccuriTM C6 flow cytometer with a limit of 40.000 events for each strain. For each sample, three replicates were done and the fluorescence background before staining was also measured. *Saccharomyces cerevisiae* BY4741 (haploid strain) and *Saccharomyces cerevisiae* BY4743 (diploid strain) [75] were used as references, through which a linear relation was created between the mean fluoresce emitted and the number of megabases already known (y=11505x+38892) that allowed to identify the DNA content of *Blastobotrys adeninivoras* PYCC4638*, Blastobotrys proliferans* PYCC5116 and *Blastobotrys solioliva* IST508. The software FlowJo, (RRID:SCR_008520) was used to analyze the data.

2.4. Riboflavin Detection Tests

Blastobotrys solioliva IST508 cells were grown on Erlenmeyer flasks with 50 mL of two different media: YPD and YNB, and incubated for 3 days at 30ºC, 250 rpm. Afterwards, samples of both the culture media and also from the sterilized non-inoculated media were taken to be analyzed by HPLC. 500 µL of media from each flask was taken, in order to analyze the presence of riboflavin. These samples were centrifuged (8000 rpm, 3min) and the supernatant was collected. Also, a positive control with 1 g/L of riboflavin was prepared. These supernatants were diluted 10 times in water in HPLC vials. Samples were analyzed by the HPLC system (Merck Hitachi, Darmstadt, Germany) with a UV-Vis detector (Hitachi LaChrom Elite L-2420, at 271 nm) and a Chromalith Purospher® RP-18 endcapped column (Merck) eluted at room temperature with an isocratic mobile phase (48% H2O+ 52% acetonitrile), at a flow rate of 1mL/mim. Additionally, the samples were scanned in a UV-VIS spectrophotometer (Shimadzu) to obtain the absorbance spectrum at a wavelength ranging from 210 to 800 nm.

2.5. Genome sequencing and annotation of *Blastobotrys solioliva* IST508

Blastobotrys *solioliva* IST508 was grown overnight in liquid yeast extract-peptone-dextrose (YPD) medium at 30°C with orbital shaking (250 rpm) and reinoculated in fresh YPD medium until early log phase. For genomic DNA (gDNA) extraction, 15 mL of this culture was centrifuged at 5000 rpm, for 5 minutes, at 4ºC, to remove the growth medium. The cells were then resuspended in 1mL sorbitol (1M) and EDTA 0.1M, at pH 7.5. A total of 10 mg/mL zymolyase were added and this mixture was incubated at 37ºC until protoplast formation. The protoplasts were centrifuged at 5000 rpm for 5 minutes and resuspeded in 1 mL Tris-HCl (pH 7.4) 50 mM + EDTA 20 mM. A total of 30 µL of SDS 10% were added and mixed well. This mixture was incubated at 65 °C for 30 min. After this period, 250 µL of Potassium Acetate (5 M) were added, incubated in ice for 1 hour and centifuged at 10000 rpm for 10 min. The supernatant was transferred to a new microcentrifuge tube and 1 volume of cold Isopropanol was added. This mixture was then centrifuged at 5000 rpm for 15 min. The supernatant was discarded, and the pellet was washed twice with 1 mL Ethanol 70%. The pellet was dried and ressuspended in 200 µL TE (pH 7.4). A total of 0.5 µL of RNase (stock 10 mg/ml) was added this mixture was incubated for 1 hour at 37 °C, followed by centrifugation for 15 min at 10000 rpm. The supernatant containing the gDNA was then transferred to new tubes.

Genomic DNA was sequenced in an Illumina Novaseq 6000 platform, producing 2 × 150-bp paired-end reads. Library preparation (NEBNext® DNA Library Prep Kit from NewEngland Biolabs) and sequencing were carried out by Novogene Corporation. Illumina sequencing produced 21753322 raw paired-end reads. Low-quality bases and adapters were removed using BBDuk from BBMap package (https://sourceforge.net/projects/bbmap/). Read duplicates were removed using PRINSEQ (v0.20.4) [76]. Ultimately, 17196026high-quality reads were used for subsequent analysis. The draft genomes were assembled into scaffolds using SPAdes (v3.11.1) [77]. Scaffolds smaller than 1000 bp were filtered out, and the remaining sets of scaffolds were used as draft assemblies. Assembly quality was analyzed using Quality Assessment Tool for Genome Assemblies (QUAST; v4.6.3) [78]. The assembly consists of 32.53% GC content, 77 scaffolds, a total length of 11535679 bp, an N50 value of 515248 bp, a longest

scaffold of 1053043 bp, and an average coverage of 174x. Genome annotation was performed by submitting the assemblies to the Yeast Genome Annotation Pipeline (YGAP), based on the Yeast Gene Order Browser (v7) [79]. Annotation was performed specifying a pre-whole-genome duplication (pre-WGD) species and predicted 5442 genes. Manual validation was performed for 652 annotated genes. This work was performed by Dr. Margarida Palma.

The Genome Analysis was carried out using the CLC Genomics Workbench 20.0 (https://digitalinsights.qiagen.com). Protein sequences of the genes involved in the RF pathway (*RIB1, RIB2, RIB3, RIB4, RIB5, RIB7*), as well as others that have been previously associated to overproducing strains (*GLY1, VMA1, ADE4,BAS1, SHM2, ADE1, ADE12, IMD3, ICL, SEF1, OCT1, ATP1, SEN2, SHM2, ARO2, ILV2, ILV6, LYS5, CDD1, FUR1, SOK2, OAF1, SDH1, SDH2, SDH3, PDX1, NDI1, DLD1, CBR1, GLR1, MTO1, MET5, PUT1, FAS1, AgHEM14, ERV2, ERO1, FMN1, TRP2,TRP5, MET6, STR3, MET17 ,MET2, SAM2, MET10, RKI1, PRS1, ADE5,7, ADE8, ADE2, MSH2, MSH3, MSH6, MLH1, MLH2, MLH3, PMS1*) [55,80,81] were retrieved from KEGG [82] and analyzed by BLAST, against the *Blastobotrys solioliva* IST508 genome. Afterwards, for the genes that matched with more than one gene from the novel species, Pairwise Sequences Alignments were done, to identify regions of similarity that could indicate functional and structural relationships between the two biological sequences, using the EBI tool EMBOSS Water for pairwise alignments [83,84]. Comparisons were also made using the online tools ScanProsite, Pfam and InterProScan. ScanProsite is a tool that allows to scan proteins for matches against the PROSITE collection of motifs [85,86] Pfam is a widely used database of protein families and domains [87]. InterProScan is a software where sequences of proteins are scanned against the InterPro database that provides functional analysis of protein sequences by classifying them into families and predicting the presence of domains and important sites [88].

3.Results

3.1 Identification of Yeast Isolates

In this study, samples from different origins (soil, sea, fruits, vegetables and beer must) were analyzed. For some, cycles in xylose enrichment were performed, so that only strains able to grow on xylose were isolated, while others were directly isolated in glucose-rich media. In order to identify the different strains isolated, sequences of either ITS or D1/D2 (or both) regions were amplified and sequenced and the results compared through BLAST (Basic Local Alignment Search Tool) with the sequences deposited in the NCBI database (Table 6). Samples highlighted in green were selected in glucose rich media, in orange were selected in xylose rich media and in blue were isolated in WLN media. In cases were the percentage of identity was equal or superior to 98% for one of the ribosomal DNA regions, it was considered that the isolate belonged to the corresponding species.

Isolation and molecular identification of fifty yeast strains was performed (Table 6), of which thirteen were selected by differential growth in xylose media, plus nine strains that were previously isolated by the Biological Sciences Research Group, were also molecularly identified (Table 7). In total, fifty-nine strains from twenty-six different species were identified. In cases where the identification was ambiguous, a second ribosomal DNA region was amplified and sequenced to confirm the results, as it was the case of the isolates IST508, IST513, IST475, IST471, IST497 and IST505.

The sample walnut green husk was the one that originated more isolates (Table 6). From the sample olive tree soil 2 and garden soil we were not able to isolate yeasts due to the confluent growth of filamentous fungi and in the case of physalis and juniper berries, we were not able to isolate any yeast. For the strains IST471, IST475, IST540, IST505, IST563, IST510 the identification was not conclusive.

Table 6 – Identity, in %, of the sequences of the regions D1/D2 or ITS of the different strains isolated in this work. Primers used were NL-1 (5'- GCATATCAATAAGCGGAGGAAAAG-3') and NL-4 (5'- GGTCCGTGTTTCAAGACGG- 3') to amplify the ribosomal DNA region D1/D2 [71] and ITS-1 (5'- TCCGTAGGTGAACCTGCGG- 3') and ITS-4 (5'- TCCTCCGCTTATTGATATGC- 3') for the Internal Transcribed Spacer (ITS) region of ribosomal DNA [72]. Lines highlighted in green are from samples that were selected in glucose rich media, in orange were selected in xylose rich media and in blue were isolated in WLN media. IST ID refers to the number of the yeast strain in the yeast culture collection of Instituto Superior Técnico. Symbol – means that region was not amplified and sequenced. For IST471, IST475, IST540, IST505, IST563, IST510 strains the identification was not conclusive.

Sample	IST ID	D1/D2 Identity		ITS Identity		Species
Oak Soil	IST464	99 %	*Kluyveromyces lactis*	-		*Kluyveromyces lactis*
	IST465	99 %	*Kluyveromyces lactis*	-		*Kluyveromyces lactis*
	IST494	97 %	*Pichia sporocuriosa*	99 %	*Meyerozyma caribbica*	*Meyerozyma caribbica*
	IST503	99 %	*Hanseniaspora osmophila*	-		*Hanseniaspora osmophila*
	IST504	99 %	*Kluyveromyces marxianus*	-		*Kluyveromyces marxianus*
	IST471		*Pichia membranifaciens*			

		97 %		79 %	*Pichia membranifaciens*	*Pichia membranifaciens (sequence with ambiguities)*
Olives Water	IST472	99 %	*Pichia kluyveri*		-	*Pichia Kluyveri*
	IST473	99 %	*Candida boidinii*		-	*Candida boidinii*
	IST493	98 %	*Pichia membranifaciens*		-	*Pichia membranifaciens*
	IST496	99 %	*Pichia kluyveri*		-	*Pichia Kluyveri*
	IST509		-	99 %	*Candida boidinni*	*Candida boidinii*
Olives	IST474	99 %	*Candida silvanorum*		-	*Candida silvanorum*
	IST475	99 %	*Schwanniomyces vanrijae /Debaryomyces*	99 %	*Schwanniomyces vanrijae/Schwanniomyces polymorphus*	*Schwanniomyces vanrijiae*
	IST492	99 %	*Candida silvanorum*		-	*Candida Silvanorum*
Berlengas Sea	IST476	99 %	*Candida palmioleophila*		-	*Candida palmioleophila*
	IST477	99 %	*Candida palmioleophila*		-	*Candida palmioleophila*
Berlengas Soil	IST495	98 %	*Candida membranifaciens*		-	*Candida membranifaciens*
	IST466	99 %	*Moliniella spathulata*		-	*Moliniella spathulata*
Arrabida Sea	IST540		-	99 %	*Candida carpophila/Meyerozyma guilliermondii/caribbica*	*Candida carpophila/Meyerozyma guilliermondii/caribbica*
	IST541		-	99 %	*Meyerozyma caribbica*	*Meyerozyma caribbica*
	IST542		-	99 %	*Meyerozyma caribbica*	*Meyerozyma caribbica*
Arrabida Soil	IST497	98 %	*Kazachstania viticola*	87 %	*Kazachstania viticola/Saccharomyces sp.*	*Kazachstania viticola*
	IST506		-	99 %	*Candida membranifaciens*	*Candida membranifaciens*
Olive Tree Soil 1	IST498	99 %	*Candida membranifaciens*		-	*Candida membranifaciens*
	IST499	99 %	*Hanseniaspora guilliermondii*		-	*Hanseniaspora guilliermondii*
	IST500	99 %	*Pichia manshurica*		-	*Pichia manshurica*
	IST507			99 %	*Candida membranifaciens*	*Candida membranifaciens*
	IST508	98 %	*Blastobotrys proliferans*	95 %	*Blastobotrys proliferans*	*Blastobotrys solioliva*
	IST513	97 %	*Pichia sp.*	99 %	*Pichia manshurica*	*Pichia manshurica*
Plums	IST502	99 %	*Meyerozyma guillermondii*		-	*Meyerozyma guilliermondii*
	IST505	99 %	*Meyerozyma guillermondii/caribbica*	99 %	*Meyerozyma carpófila/ Meyerozyma guillermondii/caribbica*	*Meyerozyma guilliermondii*
Cabbage	IST510	99 %	*Ustilaginomycetes sp.*		-	*Ustilaginomycetes sp.*
	IST511	99 %	*Pseudozyma aphidis*		-	*Pseudozyma aphidis*
	IST512	99 %	*Rhodotorula mucilaginosa*		-	*Rhodotorula mucilaginosa*

	IST563		-	99 %	*Ustilaginomycetes sp.*	*Ustilagomycetes sp.*
	IST539	99 %	*Pseudozyma aphidis*		-	*Pseudozyma aphidis*
Walnut Green Husk	IST543		-	99 %	*Sporobolomyces ruberrimus*	*Sporobolomyces ruberrimus*
	IST544		-	99 %	*Cystobasidium slooffiae*	*Cystobasidium slooffiae*
	IST545		-	99 %	*Cystobasidium slooffiae*	*Cystobasidium slooffiae*
	IST546		-	99 %	*Cystobasidium slooffiae*	*Cystobasidium slooffiae*
	IST547		-	97 %	*Cystobasidium slooffiae*	*Cystobasidium slooffiae*
	IST548		-	99 %	*Cystobasidium slooffiae*	*Cystobasidium slooffiae*
	IST549		-	99 %	*Rhodotorula(Rhodosporidium) babjevae*	*Rhodotorula(Rhodosporidium) babjevae*
	IST550		-	99 %	*Rhodotorula Babjevae*	*Rhodotorula babjevae*
	IST551		-	99 %	*Sporobolomyces roseus*	*Sporobolomyces roseus*
	IST552		-	99 %	*Aureobasidium pullulans*	*Aureobasidium pullulans*
	IST553		-	99 %	*Aureobasidium pullulans*	*Aureobasidium pullulans*
	IST554		-	99 %	*Aureobasidium pullulans*	*Aureobasidium pullulans*
	IST555		-	100%	*Aureobasidium pullulans*	*Aureobasidium pullulans*
	IST556		-	99 %	*Aureobasidium pullulans*	*Aureobasidium pullulans*
	IST557		-	99 %	*Aureobasidium pullulans*	*Aureobasidium pullulans*
	IST558		-	99 %	*Filobasidium magnum*	*Filobasidium magnum*
	IST559		-	98 %	*Cryptococcus flavescens*	*Cryptococcus flavescens*
	IST560		-	99 %	*Cryptococcus carnescens*	*Cryptococcus carnescens*
	IST561		-	99 %	*Cryptococcus carnecens*	*Cryptococcus carnecens*
	IST562		-	99 %	*Cryptococcus flavescens*	*Cryptococcus flavescens*
Olive Tree Soil 2					-	
Garden Soil					-	
Juniper Berries					-	
Physalis					-	

Table 7 - Identity, in %, of the sequences of the regions D1/D2 or ITS of the different strains previously isolated by the Biological Sciences Research Group by Dr. Margarida Palma. Primers used were NL-1 (5'- GCATATCAATAAGCGGAGGAAAAG-3') and NL-4 (5'- GGTCCGTGTTTCAAGACGG- 3') to amplify the ribosomal DNA region D1/D2 [71] and ITS-1 (5'- TCCGTAGGTGAACCTGCGG-3') and ITS-4 (5'- TCCTCCGCTTATTGATATGC- 3') for the Internal Transcribed Spacer (ITS) region of ribosomal DNA [72]. Symbol – means that the identity for that region was not tested.

IST ID	D1/D2 identity (%)		ITS identity (%)	Species
IST457	-	99%	*Kluyveromyces marxianus*	*Kluyveromyces marxianus*
IST461	-	100%	*Rhodotorula mucilaginosa*	*Rhodotorula mucilaginosa*
IST458	-	99%	*Pichia kudriavzevii*	*Pichia kudriavzevii*

IST462		-	100%	*Saccharomyces cerevisiae*	*Saccharomyces cerevisiae*
IST448		-	99%	*Pichia sporocuriosa*	*Pichia sporocuriosa*
IST453	99%	*Saccharomyces cerevisiae*		-	*Saccharomyces cerevisiae*
IST454	100%	*Saccharomyces cerevisiae*		-	*Saccharomyces cerevisiae*
IST455		-	99%	*Saccharomyces cerevisiae*	*Saccharomyces cerevisiae*
IST456		-	99%	*Saccharomyces cerevisiae*	*Saccharomyces cerevisiae*

3.2 Characterization of *Blastobotrys solioliva* IST508

The strain IST508 isolated from an olive tree soil from *Alentejo, Portugal* (38°02'43.6"N 8°06'34.3"W) presented an ITS identity of 95% and D1/D2 identity of 98% with *Blastobotrys proliferans*, also no ambiguities were found in the sequence obtained, indicating that it could be a new species. For this reason, from this point on, the new species will be referred as *Blastobotrys solioliva* IST508.

3.2.1 Phylogenetic Placement

The sequences of related species retrieved from GenBank were aligned iteratively by using the multiple alignment tool CLUSTAL X available from the software MEGA v.10.0.5. This software was also used to study the phylogenetic placement of the species under study, using the two-parameter model and the maximum likelihood method (Figure 9). The robustness of the clades was assessed using bootstrap analysis with 1000 replicates. The closest known species to the new one is *Blastobotrys proliferans*.

Figure 9 - Phylogenetic placement of Blastobotrys solioliva IST508 and its teleomorphic genus Trichomonascus among members of the Saccharomycetales, based on maximum parsimony analysis of D1/D2 rRNA gene sequences created on an alignment of sequences of the D1/D2 domains of the 26S rDNA region, computed using the maximum likelihood method and the Tamura–Nei model, performed by the software Mega-X. The numbers provided on branches are frequencies with which a given branch appeared in 1000 bootstrap replications. Location of the novel species found is highlighted by the red box.

3.2.1 Cellular Morphology

3.2.1.1 Agar-grown colonies

In order to morphologically characterize the proposed new species, colonies were grown on 5% malt extract agar. The morphology of the colonies was observed, under a stereomicroscope, after 3 days of incubation, at 30ºC (Figure 10A), and under the microscope after 7 days of incubation (Figure 10B).

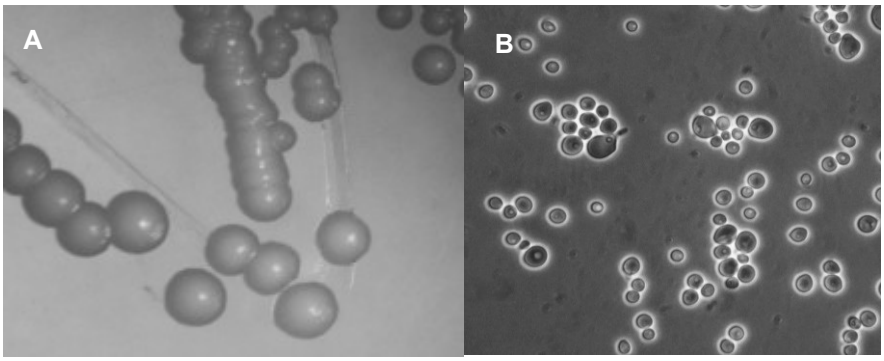

Figure 10 – A) Blastobotrys solioliva IST508. Stereomicroscope observation of colonies after 3 incubation days, 5% malt extract agar, 30ºC. B) Microscopic observation, 7 incubation days, 5% malt extract agar, 30ºC (1000x).

After 3 days of incubation, at 30ºC, colonies were dry, raised, butyrous, soft, cream, circular with a smooth surface, margins are entire, and no hyphae or pseudohyphae are observed. After 7 days of incubation, and according to microscopic observations, cells range from spherical to ovoid, and occur as single cells, in pairs or in small clusters, pseudohyphae and hyphae growth are absent. No growth was observed in the 50% glucose agar medium.

3.2.1.2 Growth in pre-sporulation and sporulation media

In order to assess the new species ability to form spores, or to reproduce sexually, colonies were plated on pre-sporulation media, followed by sporulation media. This media is poor in nutrients, so cells were under starvation and were forced to produce spores. Observations under the microscope were performed after 3 and 7 days of incubation at 30ºC (Figure 11).

After 7 days of incubation, cells maintained the morphology in pre-sporulation media (Figure 11A), however when transferred to sporulation media, it changed (Figure 11B). Cells were more elongated and seemed to have small buds surrounding a bigger cell.

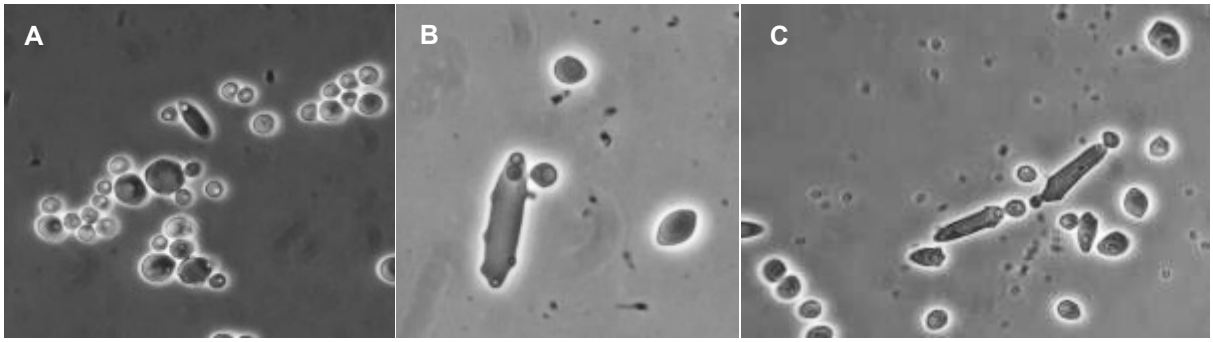

Figure 11- A). Blastobotrys solioliva IST508, microscopic observation after 7 days of incubation in pre-sporulation media, at 30ºC, (1000x). B) Microscopic observation after 7 incubation days in sporulation media at 30ºC, 1000x. C) microscopic observation, after 3 incubation days in sporulation media, 30ºC, (1000x).

Cells' morphology after 3 (Figure 11C) or 7 days of incubation is similar. Some cells also present a more elongated shape, with buds surrounding a bigger cell.

3.2.1.3 Temperature Effect

It is known that temperature can affect the morphology of the cells, so this condition was also evaluated. Cells were grown on YPD and ME, at 25ºC, 37ºC and 40ºC in order to evaluate differences in growth and morphology. Cells were observed under the microscope after 3 days of incubation. Results in YPD and ME were similar, and for this reason, only growth in ME medium is shown. Growth was observed in all agar plates, except for the plate incubated at 40ºC (Figure A1 - Annexes).

Cells grown in ME Agar, at 25ºC are shown in Figure 12A. It is observed that most cells maintain their spherical to ovoid form, however there are a few cells that are much larger and have small spheres around it, suggesting that cells are budding. At 37ºC (Figure 12B) cells are similar to cells grown at 25 and 30ºC.

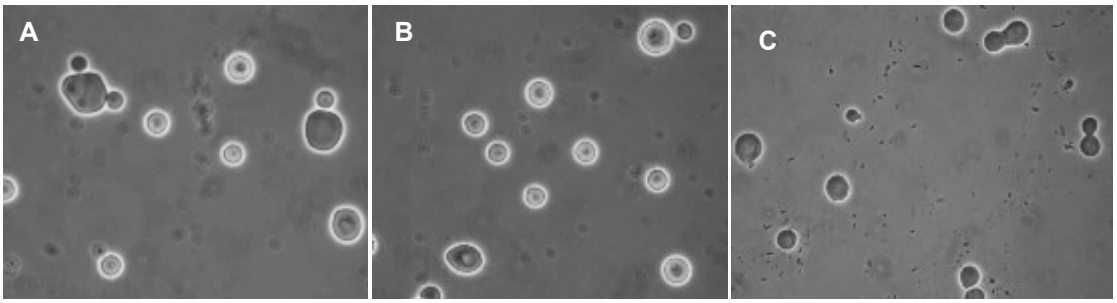

Figure 12 - Blastobotrys solioliva IST508, microscopic observation. A) 3 incubation days, ME Agar, 25ºC (1000x). B) 3 incubation days, ME Agar, 37ºC (1000x). C) 3 incubation days, ME Agar, 40ºC (1000x).

Although a weak growth was observed at 40ºC in solid medium, we were able to take some cells from the agar plate and observe in the microscope (Figure 12C). At this temperature, cells are still round, but several seem to have stopped at the final stages of budding. The intracellular content is not distinguishable.

3.2.1.4 Calcofluor Staining

Blastobotrys solioliva IST508 cells were stained with Calcofluor White to confirm if they were reproducing asexually by checking the presence of bud scars. Figures 13, 14 and 15 show the cells stained with Calcofluor White for 2 and 7 days, in Pre-Sporulation, Sporulation and YPD media, respectively. In general, cells of *Blastobotrys solioliva* IST508 kept their circular form, and it is possible to clearly distinguish between mother cells, where the fluorescence is seen stronger in the cell wall, and daughter cells, where the signal is weaker. In Figures 13, 14 and 15, cells caught in the process of budding show brighter zones in the cell wall rich in chitin (highlighted by the white arrow) and also mother cells with bud scars can be seen (red arrow), confirming that this novel species is an anamorphic

member of the order. Also, a smaller number of cells are bigger than the rest and some presented a longer shape (Figures 14 – 15) as observed before (Figure 11).

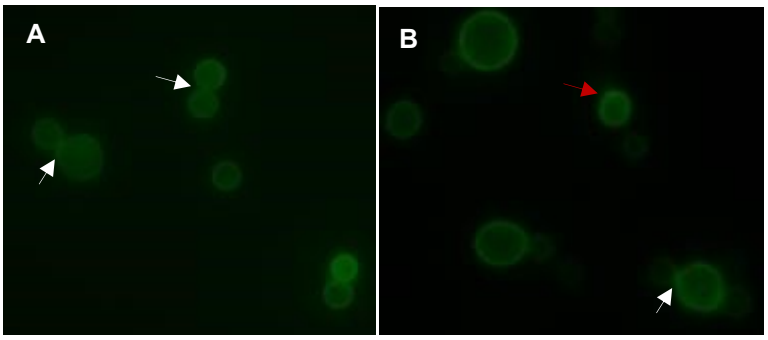

Figure 13 – Budding patterns of Blastobotrys solioliva IST508 stained with Calcofluor White and grown in solid Pre-Sporulation Media, at 30ºC for 2 (A) and 7 (B) days. White arrows indicate brighter areas (rich in chitin) of the membrane of budding cells and red arrows indicate bud scars in mother cells.

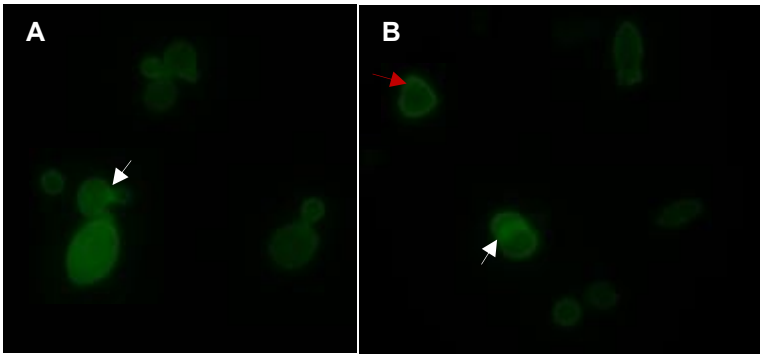

Figure 14 - Budding patterns of Blastobotrys solioliva IST508, stained with Calcofluor White and grown in solid Sporulation Media, at 30ºC for 2 (A) and 7 (B) days. White arrows indicate brighter areas (rich in chitin) of the membrane of budding cells and red arrows indicate bud scars in mother cells.

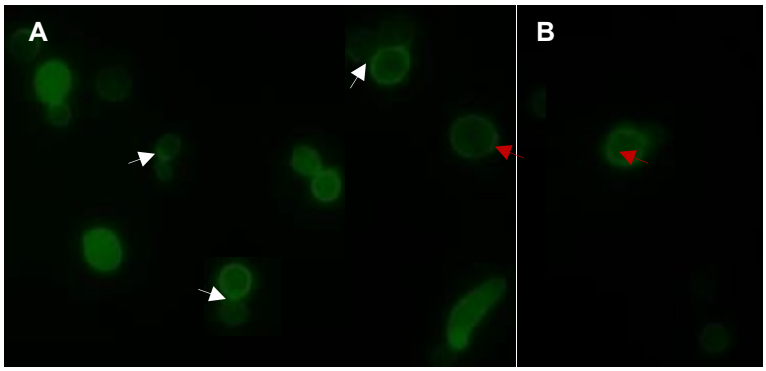

Figure 15 - Budding patterns of Blastobotrys solioliva IST508 , stained with Calcofluor White and grown in YPD Agar, at 30ºC for 2 (A) and 7 (B) days. White arrows indicate brighter areas (rich in chitin) of the membrane of budding cells and red arrows indicate bud scars in mother cells.

3.2.2 Physiological Characterization

To characterize *Blastobotrys solioliva* IST508 physiology, a set of tests was performed. In terms of its ability to ferment and assimilate, different compounds were tested. Vitamins requirements, growth at different temperatures, and osmotolerance are other physiological tests that were performed [6].

3.2.2.1 Assimilation Assays

To assess the assimilation ability of *Blastobotrys solioliva* IST508, yeast cells were cultured in YNB containing 20 g/L of each of the tested compounds. After 7 days of incubation, at 30ºC, the optical density at 600 nm was measured.

The optical densities measured for each culture of *Blastobotrys solioliva* IST508 grown in the different compounds tested are shown in Figure 16. The error bars represent the mean and standard deviation of the three replicates performed and were calculated by the software GraphPad Prism. It is observed that the strain of the novel species grew well, in general, in all the compounds tested, with exception of galacturonic acid, myo-inositol, lactose, soluble starch, methanol 1 and 5%, ethanol 5%, where growth was absent (OD smaller than 0.5). In rhamnose and 10% NaCl+5% Glucose growth was weak, reaching optical densities of 3. On the other hand, this strain grew considerably well when cultured with the sugars fructose, galactose, glucose, arabinose, xylose, sucrose, maltose, mannose, melezitose, raffinose, trehalose, melibiose, sorbose and other carbon sources, like glycerol 1% and 5%, sorbitol, mannitol, ethanol 1%, inulin, and in vitamin free-media (Vitamin requirements) reaching optical densities higher than 5. In rhamnose, ribose, succinic acid, xylitol, galactitol growth reached optical densities between 3 and 5.

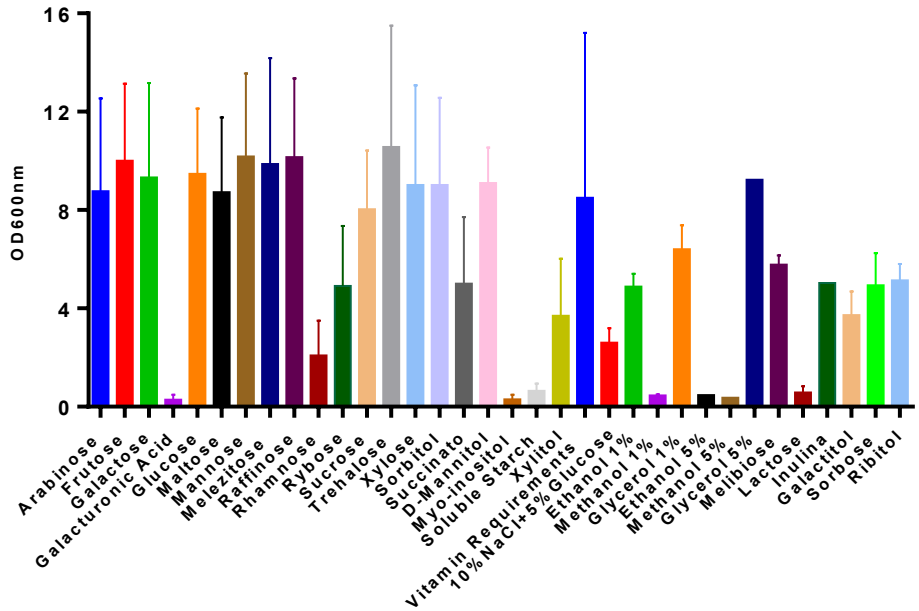

Figure 16 – Optical density at 600 nm, after 7 days of incubation of Blastobotrys solioliva IST508, in YNB, at 30ºC and 250 rpm, with different compounds as carbon sources, at an initial concentration of 20g/L.

3.2.2.2 Fermentation Assays

Fermentation is important in biotechnological applications and when describing a new species, the ability to ferment several carbon sources is usually tested. Generally, fermentation tests are carried out in Durham tubes that allow to observe the formation of gas throughout the fermentation [6]. However, in this study, fermentation tests were carried out in 25 mL fermentation flasks capped with a rubber cork perforated with a needle to allow carbon dioxide release during fermentation. Fermentation basal media was supplemented with different carbohydrates (final concentration of 20 g/L) and inoculated with

Blastobotrys solioliva IST508 yeast cells at an initial optical density (600nm) of 0.1. The fermentation flasks' weight was monitored daily, in order to follow carbon dioxide losses.

Figures 17-18 represent the weight loss, in grams, per time, in hours, of fermentation. The error bars represent the mean and standard deviation of the three replicates performed and were calculated by the software GraphPad Prism. The weight loss was calculated by the difference between the weight of the flask at a given time-point and the initial weight (at time-point 0). It is considered that the rate of weight loss is proportional to the carbon dioxide released hence it is a measure of the fermentation rate. When compared to *S.cerevisiae* strains, *B. solioliva* fermentation is very slow [89]. The fermentation profiles observed were higher when fructose and sucrose were added as carbon source, followed by glucose, raffinose, galactose and trehalose, in descending order. In the case of arabinose, xylose and maltose, the weight losses were minimal.

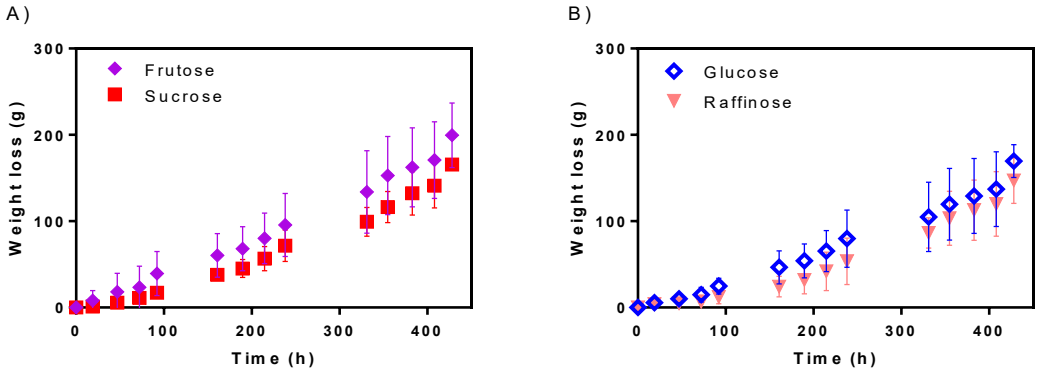

Figure 17 – Fermentation Assays. Weight lost by the flask, in grams, inoculated with the strain of the novel species Blastobotrys solioliva IST508, per fermentation time, in hours, with (A) Fructose and Sucrose, (B) Glucose and Raffinose as sole carbon source in each case. Cells were grown in fermentation basal media, supplemented with 20g/L of each one of the sugars, at 30ºC, 250 rpm.

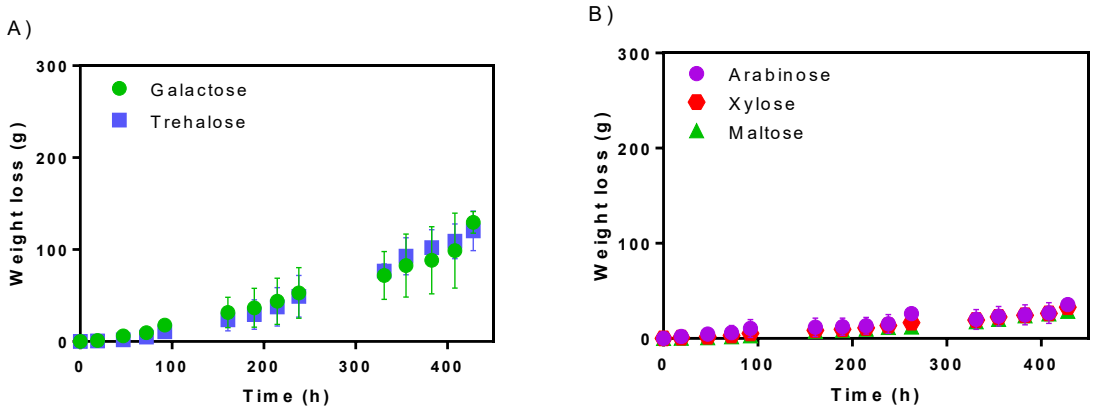

Figure 18 - Fermentation Assays. Weight lost by the flask, in grams, inoculated with the strain of the novel species Blastobotrys solioliva IST508 per fermentation time, in hours, with (A) Galactose and Trehalose, (B) Arabinose, Xylose and Maltose, as sole source of carbon in each case. Cells were grown in fermentation basal media, supplemented with 20g/L of each one of the sugars, at 30ºC, 250 rpm.

3.2.3 Growth curve, Sugar Consumption and Extracellular Metabolites Produced by *Blastobotrys solioliva* IST508

The growth curve of *Blastobotrys solioliva* IST508 was monitored for 145h of incubation with either glucose or xylose as carbon sources (Figure 19). The type strains *B. adeninivorans* PYCC 4638, *B. proliferans* PYCC 5116 were used for comparison. The error bars, on Figures 19-23 represent the mean and standard deviation of the three replicates performed and were calculated by the software GraphPad Prism. In general, after 50 hours of incubation cells leave the exponential phase and entered the stationary phase. *Blastobotrys solioliva* IST508 reaches smaller OD_{600nm} in comparison to the other species tested.

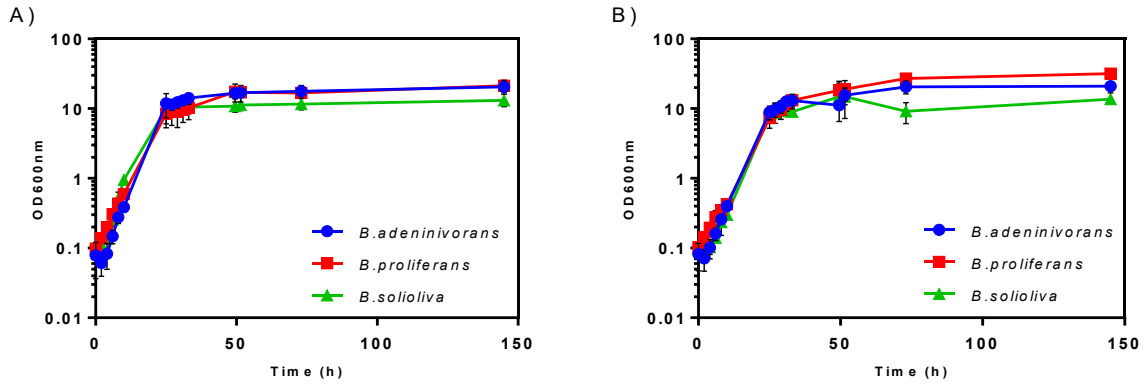

Figure 19 – Growth curves of B. adeninivorans PYCC4638, B.proliferans PYCC 5116 and Blastobotrys solioliva IST508 with 20 g/L of either glucose (A) or xylose (B) as carbon sources, cells were grown in YNB at 30°C, 250 rpm for 145 hours.

HPLC analysis of the composition of the culture media during cultivation was also performed, allowing the description of the sugar consumption profiles for each species. In Figure 20, it is observed that both *B. adeninivorans* PYCC4638 and *B. proliferans* PYCC 5116 are able to consume almost entirely the sugar sources present, in opposition to *Blastobotrys solioliva* IST508, where almost half of the initial sugar (glucose or xylose) concentration are left in the growth medium due to growth stoppage.

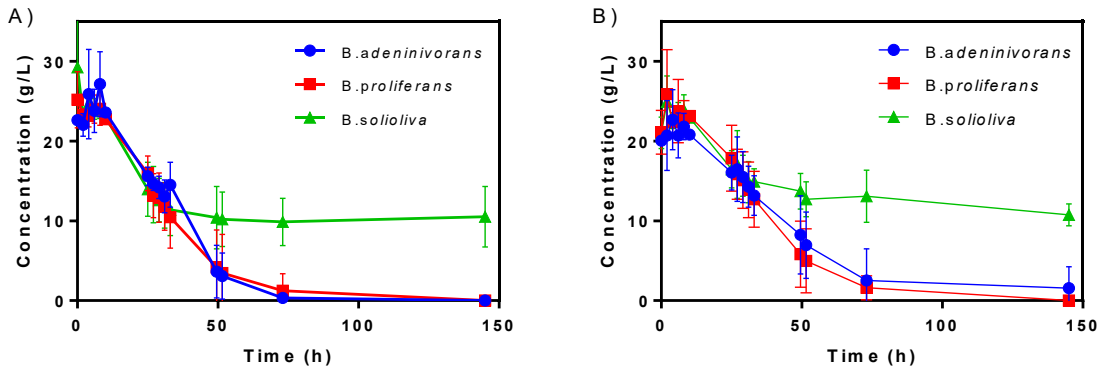

Figure 20 – Consumption of Glucose (A) and Xylose (B) by B. adeninivorans PYCC4638, B.proliferans PYCC5116 and Blastobotrys solioliva IST508 during 145 hours of culture, cells were grown in YNB at 30°C, 250 rpm for 145 hours.

Besides the analysis of glucose and xylose consumption, the production of glycerol, ethanol and organic acids (oxalic acid, citric acid, formic acid, acetic acid, malic acid and succinic acid) were also assessed

for each species tested, since the extracellular pH of the media was around pH 2.0, only after a few hours of incubation. Glycerol or ethanol production was not detected. Figure 21 shows the concentrations of the organic acids produced by *B. adeninivorans* either with glucose (A) or xylose (B) as carbon sources. After 8 hours of growth in glucose medium, formic acid and citric acid start to be produced, reaching maximum concentrations of almost 4 and 2 mM, respectively. After 30 hours, citric acid concentration reaches its peak, and starts decreasing, whereas formic acid reaches the peak after 70 hours, simultaneously with glucose depletion in the media, and also, the diversity of acids produced is bigger in comparison to xylose medium, where only malic acid (0,6 mM) and formic acid (2 mM) are produced.

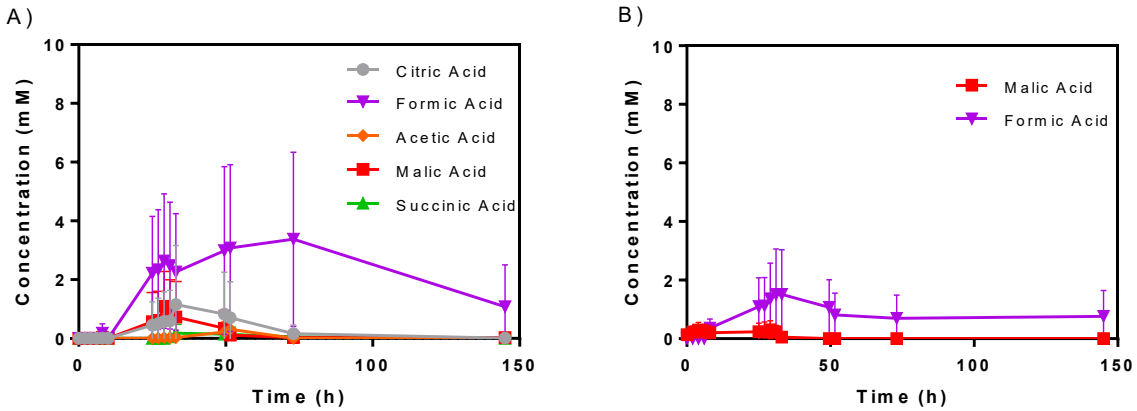

Figure 21 – Extracellular metabolites production by B.adeninivorans PYCC4638 during 145 h of culture, determined by HPLC analysis, with 20 g/L of etheir Glucose (A) or Xylose (B) as carbon sources, cells were grown in YNB at 30ºC, 250 rpm for 145 hours.

In the case of the organic acids produced by *B. proliferans* (Figure 22), formic acid is, by far, the most produced one (glucose as carbon source) during 25 to 50 hours of incubation reaching concentrations of approximately 45 mM in the growth medium. Malic acid is actively produced during the 25th to 30th hour, reaching maximum concentrations of 20 mM. After these periods, their concentration in the growth medium is significantly reduced. When xylose was used as a carbon source, the maximum concentrations of these two acids were much smaller than the ones observed in glucose medium. Formic acid (3 mM) and succinic acid (1 mM) are the most produced acids, in the case of the last acid the production was similar when glucose was the carbon source.

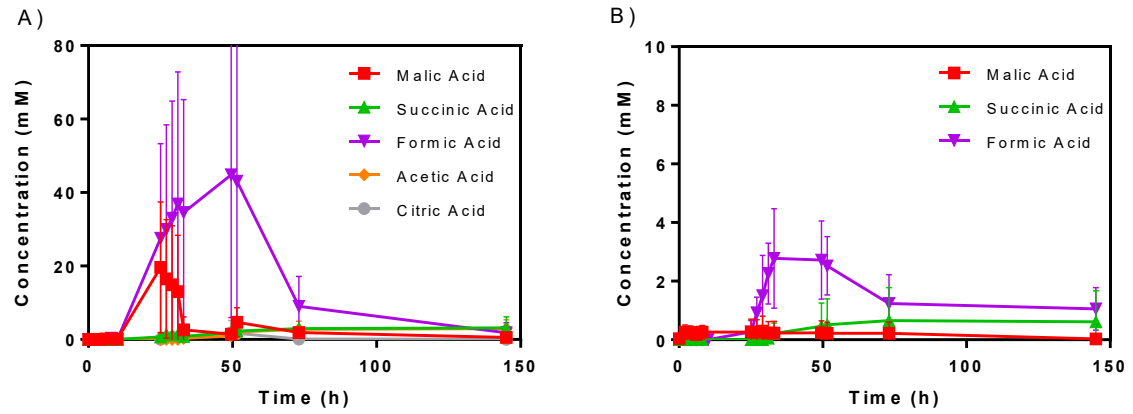

Figure 22 - Extracellular metabolites production by B.proliferans PYCC5116 during 145 h of culture, determined by HPLC analysis, with 20 g/L of either Glucose (A) or Xylose (B) as carbon sources, cells were grown in YNB at 30ºC, 250 rpm for 145 hours

Blastobotrys solioliva IST508 produces similar compounds when growing on glucose and xylose, however, at different concentrations (Figure 23). When glucose is the carbon source, formic acid is the most produced acid, reaching max concentrations of approximately 2 mM, followed by malic and acetic acids, with concentrations around 1 mM. There is also a residual presence of succinic acid. When xylose is the carbon source, formic acid is also the most produced acid (3 mM), followed by succinic acid, malic acid and acetic acid (concentrations around 1 mM), and residual concentrations of oxalic acid are detected. Also, these concentrations started to increase only after 50h of incubation, reaching its maximum in the last point taken, in opposition to what occurs with glucose, where concentrations are almost constant after 20 h.

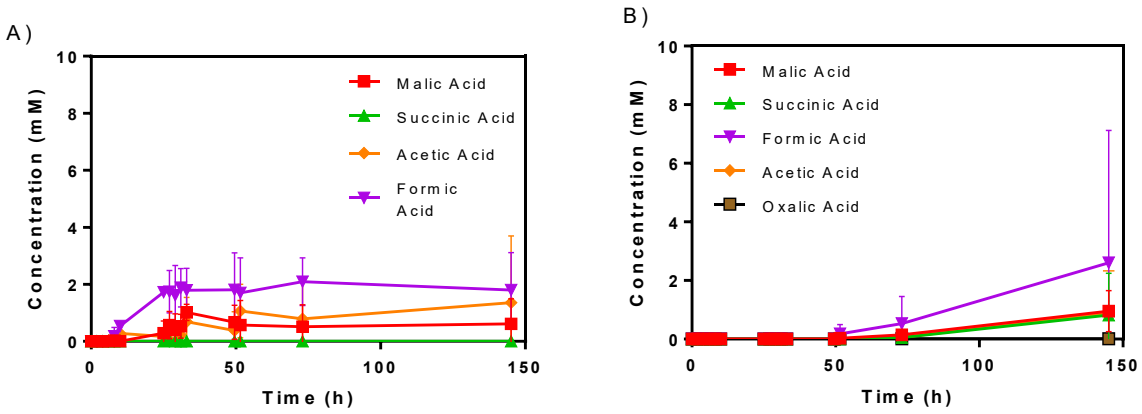

Figure 23 - Extracellular metabolites production by Blastobotrys solioliva IST508 during 145 h of culture, determined by HPLC analysis, with 20 g/L of either Glucose (A) or Xylose (B) as carbon sources, cells were grown in YNB at 30ºC, 250 rpm for 145 hours

In all three species tested formic acid is the most produced acid. In opposition, oxalic acid production is extremely rare.

3.2.4 DNA Content Analysis by Flow Cytometry

In order to study the nuclear DNA content of the new species, flow cytometry was performed. In this technique, a fluorescent dye that binds stoichiometrically to the DNA is added to a suspension of permeabilized single cells [90]. Then, the stained cells are measured by a flow cytometer, and the emitted fluorescent signal yields an electronic pulse with a height proportional to the total fluorescence emission from the cell, that should be proportional to the amount of DNA present. Reference cells should be used, with different DNA contents, in order to measure the DNA content of the unknown cells, since this method is undirect [90]. In this case, *Saccharomyces cerevisiae* BY4741 (haploid strain) and *Saccharomyces cerevisiae* BY4743 (diploid strain) [75] were used as references, through which a linear relation was created between the mean fluoresce emitted and the number of megabases (y=11505x+38892) that allowed to estimate the DNA content of *Blastobotrys adeninivorans* PYCC4638*, Blastobotrys proliferans* PYCC5116 and *Blastobotrys solioliva* IST508. FlowJo (RRID:SCR_008520) was used to analyze the data obtained and the results are shown on Table 8 and Figure 24.

Figure 24 shows the cell cycle analysis of each one of the strains studied, whereas Table 8 shows the intensities of each one of the peaks obtained for the phase G1 or phase G2 and the corresponding DNA contents. The *Blastobotrys solioliva* IST508 strain appears to be haploid, since the cell cycle analysis is more similar to BY4741, with 14.24 Mb of DNA, however it seems that all the cells are either in the G1 phase or in the S phase (Figure 24) and there is no correlation of the DNA content between the two life cycle phases. *B. proliferans* PYCC5116 and *B. adeninivorans* PYCC4638 also seem to be haploid strains, taking into account the DNA content and the cell cycle analysis (Figure 24).

Table 8 – Mean fluorescence and respective DNA Content in Megabases (Mb) of the strains Blastobotrys adeninivoras, Blastobotrys proliferans and Blastobotrys solioliva IST508. (highlighted in blue), and the reference haploid strain Saccharomyces cerevisiae BY4741 and diploid strain Saccharomyces cerevisiae BY4743, before (n) and after (2n) DNA replication.

Species	Mean Fluorescence	DNA Content (Mb)
S.cerevisiae BY4741(n)	167770	12.16
S.cerevisiae BY4741(2n)	333139	24.32
S.cerevisiae BY4743 (2n)	320808	24.32
S.cerevisiae BY4743(4n)	593000	48.64
B. solioliva IST508 (n)	202679	14.24
B. solioliva IST508 (2n)	606000	49.29
B.proliferans PYCC5116 (n)	173512	11.70
B.proliferans PYCC5116 *(*2n*)*	341382	26.29
B.adeninivorans PYCC 4638 *(*n*)*	168939	11.30
B.adeninivorans PYCC 4638 *(*2n*)*	323639	24.75

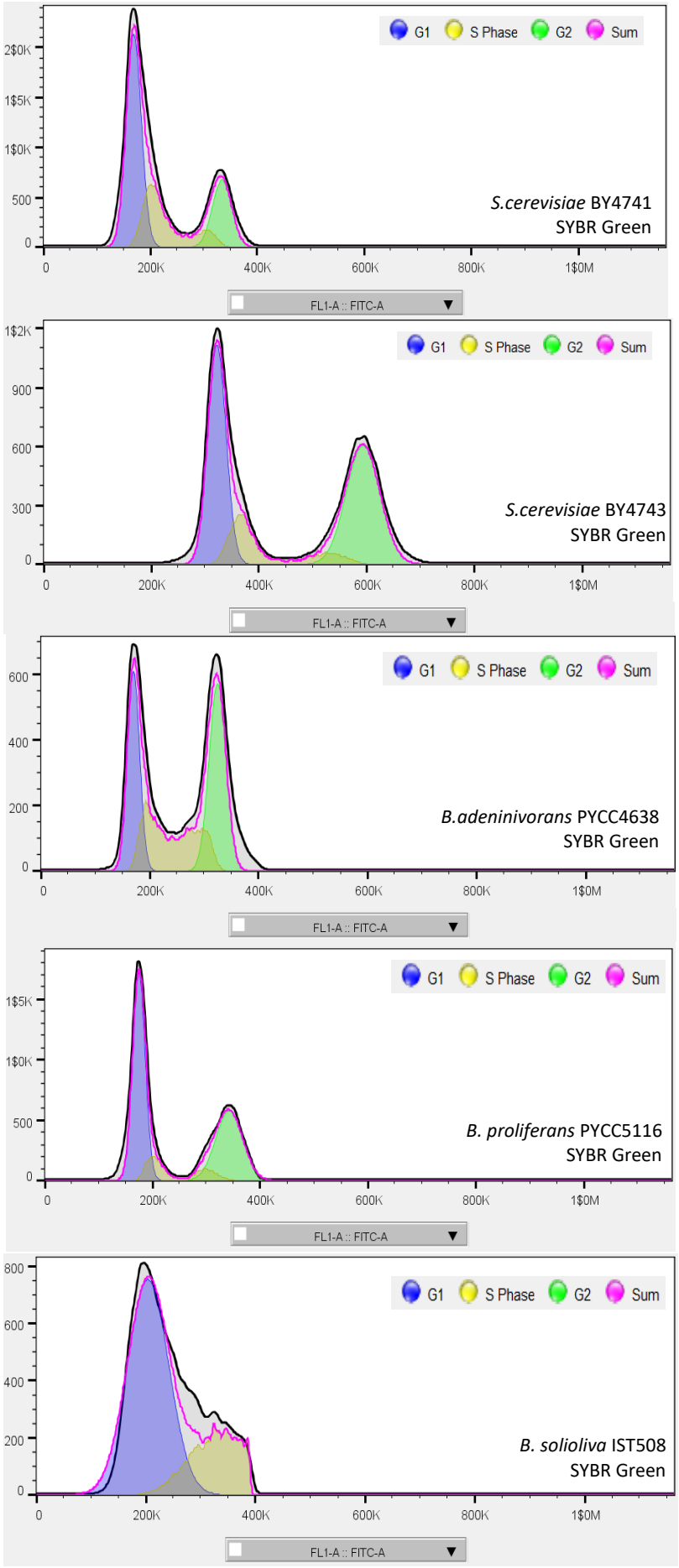
G1
S Phase
G2
Sum
2$0K
1$5K
1$0K
500
0
0
200K
400K
600K
800K
1$0M
FL1-A :: FITC-A
S.cerevisiae BY4741
SYBR Green
1$2K
900
600
300
0
S.cerevisiae BY4743
SYBR Green
B.adeninivorans PYCC4638
SYBR Green
B. proliferans PYCC5116
SYBR Green
800
B. solioliva IST508
SYBR Green

Figure 24 - Cell cycle analysis of cells from strains S.cerevisiae BY4741, S.cerevisiae BY4743, B. adeninivorans PYCC4638, B.proliferans PYCC5116 and B.solioliva IST508 stained with SYBR Green.

3.2.5 Riboflavin Tests

When performing the growth curve analysis, it was observed that the flasks of *Blastobotrys solioliva* IST508 presented a bright yellow color after a few hours of incubation, wheres the other species tested (*Blastobotrys adeninivorans* PYCC4638 and *Blastobotrys proliferans* PYCC5116) did not show the same intensity (Figure 25). The yellow color was observed either when glucose or xylose was used as carbon source, although the color was more intense in glucose medium, but the final ODs and the growth curves were similar in both.

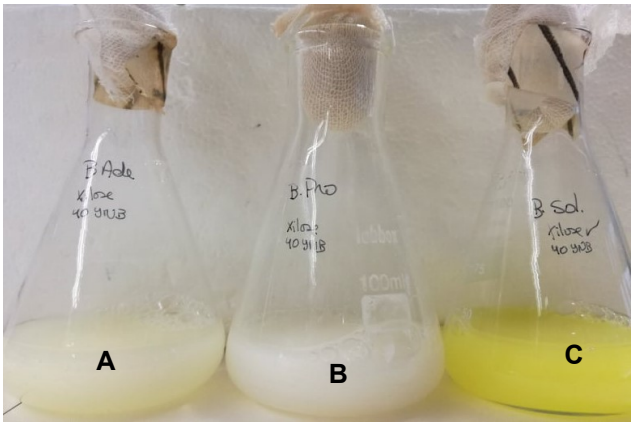

Figure 25 – Differences in the culture media (YNB) color between Blastobotrys adeninivorans PYCC4638 (A), Blastobotrys proliferans PYCC5116 (B) and Blastobotrys solioliva IST508 (C) grown with xylose as carbon source, after 2 days of incubation, at 30°C, 250 rpm.

The compound responsible for the alteration of the color of the growth medium was secreted into the extracellular media by the cells. After some research on the literature, riboflavin was pointed out as a good candidate compound based on its color and water solubility.

In order to confirm this, samples of *Blastobotrys solioliva* IST508 in YPD and YNB culture media were taken and compared with samples from non-inoculated sterile media in order to understand the differences. A positive control of riboflavin (1 g/L) was used. Figure 26 shows the spectra of the different samples analyzed superimposed. No differences between the spectra of YPD with (blue spectrum) or without (yellow) *Blastobotrys solioliva* IST508 growth were observed. However, in the case of YNB, significant differences were seen between the inoculated (green) and the sterile media (black). The magenta spectrum belongs to the positive control that contained riboflavin, and the green spectrum is from the culture media of *Blastobotrys solioliva* IST508 in YNB. There is a clear resemblance between the spectra, indicating that riboflavin is being produced by this species in this specific media.

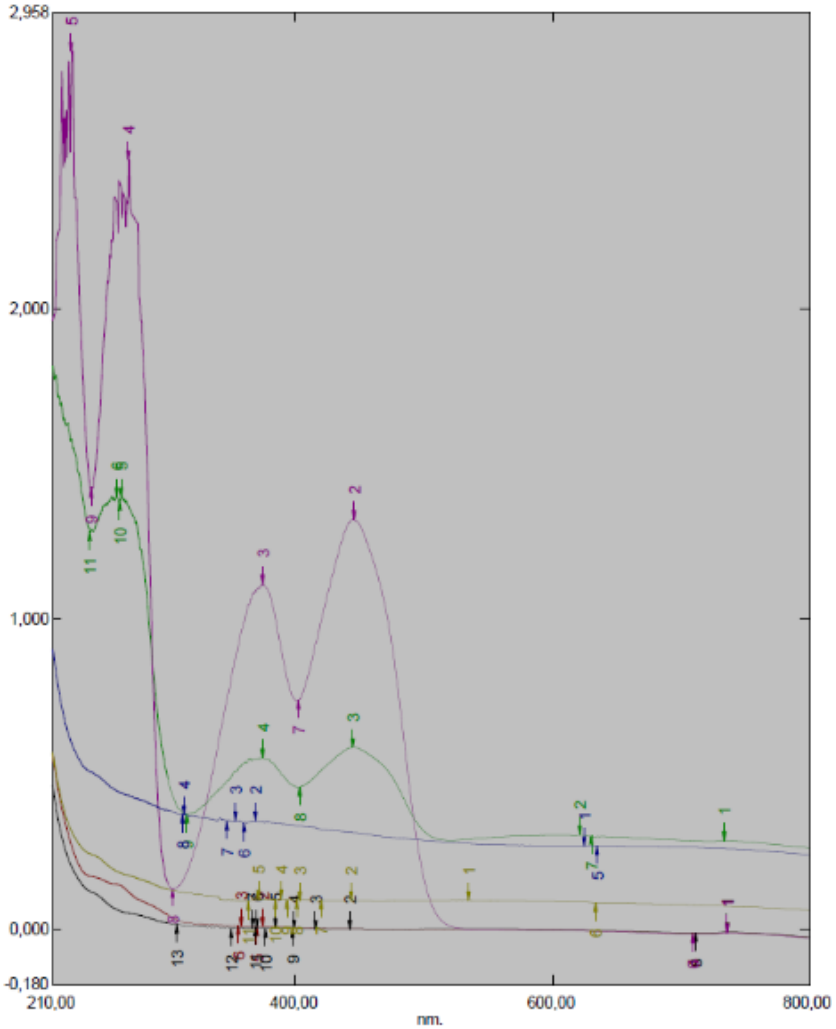

Figure 26 – UV Spectrum from 210 to 800 nm of: a solution with 1g/L of riboflavin (magenta spectrum), culture media of Blastobotrys solioliva IST508 in YNB (green spectrum), culture media of Blastobotrys solioliva IST508 in YPD (blue spectrum), sterile YNB media (black and red spectrums) and sterile YPD media (yellow spectrum).

3.2.6 Analysis of the *Blastobotrys solioliva* IST508 genes potentially involved in riboflavin biosynthesis

To better understand the possible applications of the novel yeast species in biotechnology, with a focus on its ability to produce riboflavin, the complete genome of *Blastobotrys solioliva* IST508 was sequenced. The total DNA content of the scaffolds generated yielded 11.5 Mb. Genes directly associated with riboflavin production pathway were searched for, but also genes previously identified as having an indirect role in riboflavin production in overproducing strains [55,80]. Through BLAST, it was possible to see if the novel species contained copies of these important genes. In total, sixty-five genes were analyzed.

Table 9 shows all the genes directly involved in the main pathway of riboflavin biosynthesis [91], shows also the genes that have been indirectly associated with overproducing strains of *A.gossypii* by Kato *et al.* (2020), Schwehheimer *et al.* (2016) and Lyzak *et al.* (2017) [55,80,81], the corresponding homologs in *S.cerevisiae*, their functions and in some cases the consequences of the duplication of the genes. It is possible to observe genes that encode proteins involved in several different biological processes, like

the TCA cycle, purine pathway, mitochondrial functions, amino acid metabolism, sulfur metabolism and DNA mismatch repair [55,80,81].

A BLAST search against the complete genome database of *Blastobotrys solioliva* IST508 was performed, in order to understand if the strain contained duplications or not of the genes with interest and the resultant matches are presented in the last columns of Table 9. All the genes directly involved in the riboflavin production pathway were found encoded in the strain's genome. The genes that presented matches with good e-values and identities between 30 and 80% with more than one different gene of *Blastobotrys solioliva* IST508 are: *RIB2 (BSOL0I01870 and BSOL0A01330)*, *MET2 (BSOL0M00220 and BSOL0C01770)*, *GLY1 (BSOL0A03900 and BSOL0G01250)*, *SHM2 (BSOL0AA00130 and BSOL0E04520)*, *SEF1 (BSOL0F03070 and BSOL0N00360*) and *DLD1 (BSOL0E03820 and BSOL0B01100)*, indicating that they could be duplicated in its genome. All of these genes are involved with amino acids metabolism, except DLD1, that is related to flavoproteins metabolism.

After gathering this information, the genes with potential duplicates were analyzed using the online tools ScanProsite, Pfam and InterProScan to search for motifs and identify the corresponding protein family of each one of the gene pairs, and to compare with the homologous *Ashbya gossypii* proteins (Table 10). The putative gene copies of *RIB2* and *DLD1* showed no hits with the databases and only two of the genes studied presented similar motifs to the *A. gossypii* proteins and between themselves: *SEF1*, a putative transcription factor related to riboflavin production pathway and the similar motif found is the Zn(2)-C6 fungal-type DNA-binding domain, and *MET2*, an enzyme, with alpha/beta hydrolase folds, involved in the methionine biosynthetic pathway [80,81], indicating that they are duplicated in *Blastobotrys solioliva* IST508 genome.

Table 9 – Genes with potential, direct or indirect, role in riboflavin production, the biological process where they are involved and their function. The homologs in S.cerevisiae of the genes studied (from A.gossypii) are presented. In some cases, there was information about the gene effect on the production of the vitamin, when overexpressed (signals: +, increase in riboflavin production, - decrease, n, no data). BLAST results were obtained using the CLC Genomics Workbench, where each gene was compared against the annotated genome of Blastobotrys solioliva IST508. The gene name in B. solioliva IST508 and the corresponding S.cerevisiae homolog identified are presented. Ref: Reference. [55,80,81,91].

Biological Process	Gene Name (in *A. Gossypii*)	Homolog in *S.cerevisiae*	Function	Consequence of the Overexpression	Ref	BLAST Result		Annotation Results	
						Max %Identity	E-value	Homolog in *S.cerevisiae*	Gene Name (*Blastobotrys solioliva* IST508)
Riboflavin Main Pathway	AGOS_ADL296C	*RIB1*	GTP cyclohydrolase II (first step of the riboflavin biosynthesis pathway)	+	[80,91]	63.7	5.56E-113	*RIB1*	BSOL0F01880
	AGOS_AER037C	*RIB7*	2,5-diamino-6-(ribosylamino)-4(3H)-pyrimidinone 5'-phosphate reductase	+	[80,91]	45.53	6.69E-62	*RIB7*	BSOL0Q00810
	AGOS_AEL091C	*RIB2*	Bifunctional mitochondrial enzyme with DRAP deaminase (third step in riboflavin biosynthesis) and tRNA:pseudouridine synthase activity;	+	[80,91]	57.04	0.00E+00	*RIB2*	BSOL0I00440
				+	[80,91]	36.67	2.71E-46		BSOL0A01340
	AGOS_ADR118C	*RIB3*	3,4-dihydroxy-2-butanone-4-phosphate synthase (DHBP synthase), required for riboflavin biosynthesis from ribulose-5-phosphate	+	[80,91]	55	7.51E-64	*RIB3*	BSOL0H00960
	AGOS_AGR396W	*RIB4*	Lumazine (DMRL) synthase (fourth step of the riboflavin biosynthesis pathway)	+	[80,91]	60.49	4.07E-64	*RIB4*	BSOL0H01990
	AGOS_AGR241W	*RIB5*	Riboflavin synthase (last step of the riboflavin biosynthesis pathway)	+	[80,91]	51.88	1.35E-76	*RIB5*	BSOL0E00740
	AGOS_ABL109W	*FMN1*	Riboflavin kinase, phosphorylates riboflavin to form riboflavin monophosphate (FMN),	n	[80,91]	45.71	1.44E-49	*FMN1*	BSOL0I00570
	AGOS_AER198W	*FAD1*	Flavin adenine dinucleotide (FAD) synthetase, performs the second step in synthesis of FAD from riboflavin	n	[80,91]	38.12	6.74E-33	*FAD1*	BSOL0D03620
Aminoacid Metabolism	AGOS_ABR209W	*TRP2*	Anthranilate synthase, catalyzes the initial step of tryptophan biosynthesis,	n	[80]	58.35	0.00E+00	*TRP2*	BSOL0H00540
	AGOS_AFR485C	*TRP5*	Tryptophan synthase, catalyzes the last step of tryptophan biosynthesis;	n	[80]	68.18	0.00E+00	*TRP5*	BSOL0M01340
	AGOS_ABR212C	*MET6*	Cobalamin-independent methionine synthase, involved in methionine biosynthesis and regeneration	n	[80]	72.08	0.00E+00	*MET6*	BSOL0H00510

AGOS_ACL059C	*STR3*	Peroxisomal cystathionine beta-lyase; converts cystathionine into homocysteine	n	80	49.61	6.64E-101	Protein	BSOL0L00710
AGOS_ADL031W	*MET17*	Methionine and cysteine synthase, required for sulfur amino acid synthesis	n	80	65.75	1.25E-171	*MET17*	BSOL0D01940
					31.71	1.35E-42	Protein	BSOL0Q00430
AGOS_AFR682C	*MET2*	L-homoserine-O-acetyltransferase, catalyzes the conversion of homoserine to O-acetyl homoserine (first step of the methionine biosynthetic pathway)	n	80	51.99	1.31E-143	*MET2*	BSOL0K01940
					36.04	1.82E-34	Protein	BSOL0D01770
AGOS_AFR692C	*SAM2*	S-adenosylmethionine synthetase, catalyzes transfer of the adenosyl group of ATP to the sulfur atom of methionine	n	80	77.63	0.00E+00	*SAM2*	BSOL0B00910
AGOS_AFR692C	*MET10*	Subunit alpha of assimilatory sulfite reductase, converts sulfite into sulfide	n	80	77.63	0	*MET10*	BSOL0L01090
AGOS_AFR297W	*BAS1*	Myb-related transcription factor involved in regulating basal and induced expression of genes of the purine and histidine biosynthesis pathways	n	80	47.06	3.50E-53	*BAS1*	BSOL0E02580
AGOS_AFR366W	*GLY1*	Threonine aldolase and catalyzes the formation of glycine from threonine	+	55	54.55	2.15E-133	Protein	BSOL0A03910
					55.16	8.49E-128	*GLY1*	BSOL0G01250
AGOS_ACR215C	*SHM2*	Serine hydroxymethyltransferase involved in glycine metabolic process and L-serine metabolic process	-	55	80.25	0.00E+00	Protein	BSOL0AK00130
					60	0.00E+00	Protein	BSOL0C00480
AGOS_AER221W	*ADE1*	N-succinyl-5-aminoimidazole-4-carboxamide ribotide synthetase; required for 'de novo' purine nucleotide biosynthesis	+	55	69.84	2.68E-139	*ADE1*	BSOL0B04280
AGOS_ABL186W	*ADE12*	Adenylosuccinate synthase; catalyzes the first step in synthesis of adenosine monophosphate from inosine 5'monophosphate during purine nucleotide biosynthesis	-	55	73.78	0.00E+00	*ADE12*	BSOL0Q01060
AGOS_AER117W	*IMD3*	Inosine monophosphate dehydrogenase; catalyzes the rate-limiting step in the de novo synthesis of GTP	+	55	74.12	2.00E-179	Protein	BSOL0A01990
AGOS_ADL066	*ICL1*	Isocitrate lyase; catalyzes the formation of succinate and glyoxylate from	-	80	66.04	0.00E+00	Protein	BSOL0J01300

			isocitrate, a key reaction of the glyoxylate cycle						
	AGOS_AGR369W	*SEF1*	Putative transcription factor	+	81	61.29	1.83E-174	Protein	BSOL0F03070
						36.21	6.98E-24	Protein	BSOL0O00360
	AGOS_AFR198W	*OCT1*	Mitochondrial intermediate peptidase, cleaves destabilizing N-terminal residues of a subset of proteins upon import, may contribute to mitochondrial iron homeostasis	-	80	41.66	1.36E-172	*OCT1*	BSOL0G00450
	AGOS_AGL272C	*ATP1*	Alpha subunit of the F1 sector of mitochondrial F1F0 ATP synthase, which is a large, complex required for ATP synthesis	n	80	88.16	0.00E+00	*ATP1*	BSOL0E01630
						27.66	1.76E-28	*VMA2*	BSOL0D02940
	AGOS_AGR073C	*SEN2*	Subunit of the tRNA splicing endonuclease, contains the active site for tRNA 5' splice site cleavage	n	80	31.77	2.55E-53	*SEN2*	BSOL0K01640
	AGOS_ADL287C	*ARO2*	Bifunctional chorismate synthase and flavin reductase, catalyzes the conversion of 5-enolpyruvylshikimate 3-phosphate (EPSP) to form chorismate, which is a precursor to aromatic amino acids	n	80	78.65	0.00E+00	*ARO2*	BSOL0H01890
	AGOS_AEL305C	*ILV2*	Acetolactate synthase; catalyzes the first common step in isoleucine and valine biosynthesis and is the target of several classes of inhibitors, localizes to the mitochondria	n	80	72.98	0.00E+00	Protein	BSOL0C00160
	AGOS_FAAL021W	*ILV6*	Regulatory subunit of acetolactate synthase, which catalyzes the first step of branched-chain amino acid biosynthesis; enhances activity of the Ilv2p catalytic subunit, localizes to mitochondria	n	80	61.15	3.05E-115	*ILV6*	BSOL0R01040
Pyrimidine Metabolism	AGOS_AGL123W	*CDD1*	Cytidine deaminase; catalyzes the modification of cytidine to uridine in vitro	n	80	84.62	4.15E-23	*CDD1*	BSOL0C03540
	AGOS_FAFR052C	*FUR1*	Uracil phosphoribosyltransferase, synthesizes UMP from uracil; involved in the pyrimidine salvage pathway	n	80	73.15	6.01E-104	*FUR1*	BSOL0C02490
						35.03	1.42E-34	*URK1*	BSOL0A02220
	AGOS_ABR055C	*SOK2*	Nuclear protein that negatively regulates pseudohyphal differentiation; plays a	+	80	87.74	1.55E-57	Protein	BSOL0O01170

			regulatory role in the cyclic AMP (cAMP)-dependent protein kinase (PKA) signal transduction pathway; relocalizes to the cytosol in response to hypoxia						
	AGOS_ACR052W	*SDH1*	Minor succinate dehydrogenase isozyme; Flavoprotein subunit of succinate dehydrogenase, which couples the oxidation of succinate to the transfer of electrons to ubiquinone as part of the TCA cycle and the mitochondrial respiratory chain	n	80	80.03	0.00E+00	*SDH1*	BSOL0L01480
	AGOS_ACL065C	*SDH2*	Iron-sulfur protein subunit of succinate dehydrogenase	n	80	79.17	1.15E-135	*SDH2*	BSOL0B02640
	AGOS_AFR207C	*SDH3*	Cytochrome b subunit of succinate dehydrogenase	n	80	50	3.98E-38	*SDH3*	BSOL0G02920
	AGOS_AGR323C	*PDX1*	Dihydrolipoamide dehydrogenase (E3)-binding protein (E3BP) of the mitochondrial pyruvate dehydrogenase (PDH) complex	n	80	45	8.91E-65	*PDX1*	BSOL0D00160
						36.17	2.21E-23	*LAT1*	BSOL0G01850
Flavoprotein Related Genes	AGOS_AFR447C	*NDI1*	NADH:ubiquinone oxidoreductase; transfers electrons from NADH to ubiquinone in respiratory chain, upon apoptotic stress, is activated in mitochondria by N-terminal cleavage, then translocates to cytoplasm to induce apoptosis	n	80	51.23	2.16E-142	Protein	BSOL0A04230
	AGOS_AER321W	*DLD1*	Major mitochondrial D-lactate dehydrogenase; oxidizes D-lactate to pyruvate, transcription is heme-dependent	n	80	48.91	3.15E-162	Protein	BSOL0B03960
						43.46	6.42E-129	Protein	BSOL0L01100
						25.52	1.28E-34	*DLD2*	BSOL0J01790
	AGOS_ADL087W	*CBR1*	Microsomal cytochrome b reductase	n	80	51.58	2.88E-88	*CBR1*	BSOL0P00520
						39.71	2.42E-55	*PGA3*	BSOL0G02520
						38.53	9.09E-39	*MCR1*	BSOL0X00100
	AGOS_AGR196W	*GLR1*	Cytosolic and mitochondrial glutathione oxidoreductase, converts oxidized glutathione to reduced glutathione; mitochondrial but not cytosolic form has a role in resistance to hyperoxia	n	80	57.62	1.24E-163	*GLR1*	BSOL0D02680
						30.35	2.64E-39	*LPD1*	BSOL0A05080

	AGOS_FAFR255W	*MTO1*	Mitochondrial protein, performs the 5-carboxymethylaminomethyl modification of the wobble uridine base in mitochondrial tRNAs	n	80	54.52	0.00E+00	*MTO1*	BSOL0A01010
	AGOS_ABL077W	*MET5*	Sulfite reductase beta subunit, involved in amino acid biosynthesis	n	80	52.57	0.00E+00	*MET5*	BSOL0E02990
	AGOS_AGL165W	*PUT1*	Proline oxidase, nuclear-encoded mitochondrial protein involved in utilization of proline as sole nitrogen source	n	80	45.37	2.59E-115	*PUT1*	BSOL0F02570
	AGOS_AER085C	*FAS1*	Beta subunit of fatty acid synthetase, which catalyzes the synthesis of long-chain saturated fatty acids	n	80	57.2	0.00E+00	*FAS1*	BSOL0Z00320
	AGOS_AAR021W	*HEM14*	Protoporphyrinogen oxidase, a mitochondrial enzyme that catalyzes the seventh step in the heme biosynthetic pathway	n	80	29.71	3.13E-60	*HEM14*	BSOL0M00620
	AGOS_ACR175W	*ERV2*	Flavin-linked sulfhydryl oxidase localized to the endoplasmic reticulum lumen, involved in disulfide bond formation within the ER	n	80	47.1	3.33E-41	*ERV2*	BSOL0B00270
	AGOS_ADL348W	*ERO1*	Thiol oxidase required for oxidative protein folding in the endoplasmic reticulum	n	80	34.33	3.94E-87	*ERO1*	BSOL0G00830
Purine Biosynthetic Pathway	AGOS_ACL077C	*RKI1*	Ribose-5-phosphate ketol-isomerase, catalyzes the interconversion of ribose 5-phosphate and ribulose 5-phosphate in the pentose phosphate pathway; participates in pyridoxine biosynthesis	n	80	66.39	5.95E-100	*RKI1*	BSOL0Q01160
	AGOS_AER083C	*PRS1*	5-phospho-ribosyl-1(alpha)-pyrophosphate synthetase, synthesizes PRPP, which is required for nucleotide, histidine, and tryptophan biosynthesis;	n	80	63.91	0	*PRS1*	BSOL0B03260
						49.49	1.92E-61	Protein	BSOL0A03090
						52.87	9.21E-27	*PRS5*	BSOL0Q01030
	AGOS_AFR254C	*ADE5,7*	Bifunctional enzyme of the 'de novo' purine nucleotide biosynthetic pathway,	n	80	63.82	0.00E+00	*ADE5,7*	BSOL0A01090
	AGOS_AAR120C	*ADE8*	Phosphoribosyl-glycinamide transformylase, catalyzes a step in the 'de novo' purine nucleotide biosynthetic pathway	n	80	52.75	1.63E-57	*ADE8*	BSOL0E03680
	AGOS_ACR210C	*ADE2*	Phosphoribosylaminoimidazole carboxylase, catalyzes a step in the 'de	n	80	63.14	0.00E+00	*ADE2*	BSOL0N00540

			novo' purine nucleotide biosynthetic pathway						
	AGOS_AGL334W	*ADE4*	Phosphoribosylpyrophosphate amidotransferase (PRPPAT); catalyzes first step of the 'de novo' purine nucleotide biosynthetic pathway	+	55	66.85	0.00E+00	*ADE4*	BSOL0A03380
	AGOS_AFR297W	*BAS1*	Regulator of purine nucleotide pathway	-	55	48.37	2.91E-54	*BAS1*	BSOL0E02580
Mismatch DNA repair	AGOS_AAL093C	*MSH2*	Protein that binds to DNA mismatches; initiates the mismatch repair process	n	80	32.11	1.94E-25	*MLH3*	BSOL0D00850
						20.84	1.38E-22	*PMS1*	BSOL0A00740
						30.59	4.32E-18	*MLH1*	BSOL0C01140
	AGOS_ADR168C	*MSH3*	Mismatch repair protein; forms dimers with Msh2p that mediate repair of insertion or deletion mutations and removal of nonhomologous DNA ends	n	80	38.13	4.16E-167	Protein	BSOL0B02210
						35.06	3.39E-35	*MSH1*	BSOL0Z00200
	AGOS_AGR116W	*MSH6*	Protein required for mismatch repair in mitosis and meiosis	n	80	52.06	0.00E+00	Protein	BSOL0B04570
						33.33	5.91E-31	*MSH1*	BSOL0Z00200
	AGOS_AFL199C	*MLH1*	Protein required for mismatch repair in mitosis and meiosis as well as crossing over during meiosis	n	80	52.96	8.46E-129	*MLH1*	BSOL0C01140
						27.93	1.23E-33	*PMS1*	BSOL0A00740
	AGOS_AFR226C	*MLH2*	Protein involved in mismatch repair and meiotic recombination	n	80	26.36	1.23E-16	*MLH1*	BSOL0C01140
	AGOS_AAL093C	*MLH3*	Protein involved in DNA mismatch repair and crossing-over during meiotic recombination	n	80	32.11	1.94E-25	*MLH3*	BSOL0D00850
						20.84	1.38E-22	*PMS1*	BSOL0A00740
						30.59	4.32E-18	*MLH1*	BSOL0C01140
	AGOS_AER421W	*PMS1*	ATP-binding protein required for mismatch repair in mitosis and meiosis	n	80	38.98	3.22E-66	*PMS1*	BSOL0A00740
						30.59	3.75E-33	*MLH1*	BSOL0C01140
ATP Metabolism	AGOS_ADR102W	*VMA1*	Subunit A of the V1 peripheral membrane domain of V-ATPase	-	55	83.15	0	*VMA1*	BSOL0G03460
						40	2.63E-18	*VMA2*	BSOL0D02940

Table 10 – Presence of motifs and protein family in each of the genes with matches with more than one gene of Blastobotrys solioliva IST508. Identity is the result of the pairwise alignment of the two genes. ScanProsite, Pfam and InterProScan were used to search for motifs and identify the corresponding protein family of the genes of the novel species and comparison with the homologous Ashbya gossypii proteins [86–88].

	Blastobotrys solioliva *IST508*						*Ashbya gossypii*			
Gene	Gene Name (*B.solioliva*)	Pairwise alignment Identity (%)	Scan ProSite	Pfam	InterProScan		Scan Prosite	Pfam	InterProScan	
					Protein Family	GO Terms			Protein Family	GO Terms
RIB2/ PUS9	BSOL0I00440	20,4	Pyruvate kinase	Pyruvate kinase, barrel domain (E=1.8e-168)	Pyruvate kinase	Glycolytic process Pyruvate kinase activity, Potassium ion binding, Magnesium ion binding, Catalytic activity	S4 RNA-binding domain profile Cytidine and deoxycytidylate deaminases domain profile Rlu family of pseudouridine synthase	Invertebrate-AID/APOBEC-deaminase E=7.4e-40 RNA pseudouridylate synthase E=1.8e-30	Pseudouridine synthase, RluC/RluD	RNA modification, Pseudouridine synthesis RNA binding, Pseudouridine synthase activity, Catalytic activity
	BSOL0A01340		no hit	no hit	None predicted	No Go Terms				
GLY1	BSOL0A03910	29,6	no hit	4'-phosphopantetheinyl transferase superfamily (E=1.2e-08)	None predicted	Holo-[acyl-carrier-protein] synthase activity, Magnesium ion binding	no hit	Beta-eliminating lyase E=8.7e-81	Threonine aldolase	Cellular amino acid metabolic process Catalytic activity, Lyase activity
	BSOL0G01250		no hit	Beta-eliminating lyase (E=8.7e-77)	Threonine aldolase	Cellular amino acid metabolic process Catalytic activity, Lyase activity				
SHM2	BSOL0AK00130	21,2	Serine hydroxymethyltransferase pyridoxal-phosphate attachment	Serine hydroxymethyltransferase (E=4.7e-206)	Serine hydroxymethyltransferase	Tetrahydrofolate interconversion, Glycine biosynthetic process from serine Glycine hydroxymethyltransferase activity, Pyridoxal phosphate binding, Catalytic activity	Serine hydroxymethyltransferase pyridoxal-phosphate attachment site	Serine hydroxymethyltransferase E=1.0e-211	Serine hydroxymethyltransferase	Glycine biosynthetic process from serine, Tetrahydrofolate interconversion Pyridoxal phosphate binding, Glycine hydroxymethyltransferase activity, Catalytic activity

	BSOL0C00480		Serine carboxypeptidases, serine active site Serine carboxypeptidases, histidine active site	Serine carboxypeptidase E=7.6e-132 Carboxypeptidase Y pro-peptide E=9.6e-22	Peptidase S10, serine carboxypeptidase	Proteolysis Serine-type carboxypeptidase activity Vacuole				
SEF1	BSOL0F03070	26	Zn(2)-C6 fungal-type DNA-binding domain	Fungal specific transcription factor domain E=1.1e-17 Fungal Zn(2)-Cys(6) binuclear cluster domain E=8.1e-09	None predicted	Transcription, DNA-templated, regulation of transcription, DNA-templated Zinc ion binding, DNA binding, DNA-binding transcription factor activity, RNA polymerase II-specific Nucleus, host cell nucleus	Zn(2)-C6 fungal-type DNA-binding domain	Fungal specific transcription factor domain E=2.2e-26 Fungal Zn(2)-Cys(6) binuclear cluster domain E= 8.0e-11	None predicted	Regulation of transcription, DNA-templated, transcription, DNA-templated DNA-binding transcription factor activity, RNA polymerase II-specific, Zinc ion binding, DNA binding Host cell nucleus, Nucleus
	BSOL0O00360		Zn(2)-C6 fungal-type DNA-binding domain	Fungal Zn(2)-Cys(6) binuclear cluster domain E= 6.9e-07	None predicted	Regulation of transcription, DNA-templated Zinc ion binding, DNA binding, DNA-binding transcription factor activity, RNA polymerase II-specific Host cell nucleus				
DLD1	BSOL0B03960	19,8	no hit	no hit	None predicted	No Go Terms	PCMH-type FAD-binding domain	FAD linked oxidases, C-	None predicted	Oxidation-reduction process

	BSOL0L01100		PCMH-type FAD-binding domain	FAD linked oxidases, C-terminal domain E=1.3e-49 FAD binding domain E=9.5e-31	None predicted	Oxidation-reduction process Catalytic activity, Flavin adenine dinucleotide binding, Oxidoreductase activit), FAD binding		terminal domain E=7.4e-57 FAD binding domain E=1.4e-35		Oxidoreductase activity, Flavin adenine dinucleotide binding, Catalytic activity, FAD binding
MET2	BSOL0K01940	26,8	no hit	alpha/beta hydrolase fold E=1.1e-48	Homoserine/serine acetyltransferase MetX-like	Biosynthetic process Transferase activity, Transferring acyl groups other than amino-acyl groups	no hit	Alpha/beta hydrolase fold E=4.0e-63	Homoserine/serine acetyltransferase MetX-like	Biosynthetic process Transferase activity, transferring acyl groups other than amino-acyl groups
	BSOL0D01770		no hit	alpha/beta hydrolase fold E=2.1e-44	Homoserine/serine acetyltransferase MetX-like	Biosynthetic process Transferase activity, Transferring acyl groups other than amino-acyl groups				

4. Discussion

The initial objective of this study was the isolation and identification of new non-conventional yeast strains able to grow in xylose with the purpose to screen among them for oleaginous yeasts. For that, several samples were collected from nature and different colonies were isolated, DNA was extracted, and specific regions were amplified and sequenced. Also, taxonomic identification of nine previously isolated strains by the group was repeated, since the former identification was ambiguous.

In total, fifty strains were isolated in this work, from twenty-six different species. Of these, thirteen strains were isolated in xylose, indicating that these strains are able to properly grow using this carbon source. This study also confirmed that yeasts are extremely diverse, well-adapted and widespread in the environment [1,4], since the samples studied were from different and distant geographical locations, like the soil and sea samples from the *Berlengas Island* and the soil from *Alentejo*, and it was possible to isolate yeast strains from most of them.

The sample where more different strains were isolated was the walnut green husk. One possible explanation for such high diversity is the fact that strains were directly isolated from the sample, in order to obtain the majority of yeast strains present, in opposition to other samples, were steps of differential enrichment were made. To the best of our knowledge, there are no other reports of the yeast's biodiversity present in the walnut green husk. On the contrary, no isolates were obtained from olive tree soil 2, juniper berries and physalis. This result was unexpected because the soil and fruits are rich in yeasts [6]. This probably occurred because the yeasts present were not cultivable in the lab, or the methods of isolation were not able to select them, or, particularly in the case of olive tree soil, the excessive presence of filamentous fungi impaired the growth of yeasts.

The taxonomic and phylogenetic analyses are based on the study and comparison of sequences of appropriate genes, that can show close and distant relations [29,92]. In fungi, the sequences used are from ribosomal DNA, which has a fundamental role in protein synthesis, having its organization conserved. The most used sequences for the identification of yeast species are those from the internal transcribed spacer (ITS) region, present between the 18S, 5.8S and 28S rRNA, and from the D1/D2 region, at the 26S nuclear ribosomal large subunit (LSU) rRNA gene [29,92]. To taxonomically identify each isolate, one of these two regions was sequenced, amplified and compared to the sequences present in the NCBI Database. For specific isolates, the results obtained were not conclusive, that is, the identity percentage obtained in the BLAST analysis was below 98% or several different species shared the same identity percentage with the isolate under study. So, to confirm the isolate identity it was necessary to amplify and sequence a different DNA region than the one used initially [92]. Even after re-sequencing, some results were ambiguous, since matches with high scores, with two different species, were obtained. This might be due to the overlapping of signals in the sequences obtained, or we could be dealing with hybrid strains [93], or problems during sequencing might have occurred, or we could be facing a new species, as

it was the case of the isolate IST508, where an identity of 95% with its closest match, *Blastobotrys proliferans*, was obtained for the ITS region and no ambiguities were observed in the DNA sequence.

The isolate IST508 was obtained from the soil surrounding an olive tree, in *Alentejo, Portugal.* Olive trees (*Olea europaea*) are one of the oldest domesticated trees and one of the most important agro-ecosystems in the Mediterranean Basin. Despite being of extreme importance to the economy of these regions, the olive trees microbiome carries a huge fungal diversity yet to be discovered [9,10]. González *et al.* (2019) studied the microbiome of the upper layer of soil close to olive trees and, after extracting and sequencing the DNA content directly from the sample, they observed a high percentage of unclassified DNA sequences that probably belong to unknown microorganisms [94]. This indicates that the olive tree microbiome has an important source of new species, with possible applications in biotechnology. A difficulty that might arise when studying these microbiomes is the fact that microbial communities vary according to the plant compartment and life cycle stage under study, but also vary according to the geographical origin of the sample [94,95]. The prevalent fungi classes present in the olive root endosphere, consisting of more than 85% of the species, were *Sordariomycetes* (38.1%), *Eurotiomycetes* (23%), *Agaricomycetes* (13.2%) and *Dothideomycetes* (11.5%). The remaining classes found were clearly less abundant, with many not even reaching 1% of relative abundance [94]. Until now, and to the best of our knowledge, *Blastobotrys sp.* has never been associated with the olive tree soil microbiome.

In order to confirm if IST508 isolate was a new species, a set of tests to characterize its phylogeny, morphology, fermentation and assimilation abilities were performed. Phylogenetic analysis showed that the closest relative of the new species is *Blastobotrys proliferans.* The media used to assess morphology was the same described by Kurtzman *et al.* (2011) [6], but under the microscope and after 7 days of growth, the morphology of the new species was different from the other *Blastobotrys* species described, even from *Blastobotrys proliferans*, since most of them present pseudohypha or hyphal growth. When *Blastobotrys solioliva* IST508 was grown in sporulation media, the cells' morphology tended to change. Under starvation conditions, like the ones imposed by the sporulation media, meiosis occurs, usually followed by production of a four-spore ascus [96], however no meiotic spores were formed by *Blastobotrys solioliva* IST508. The absence of meiospore formation was expected, since *Blastobotrys sp.* are anamorphic members (asexual phase of the fungal life cycle) of the *Saccharomycetales* [6]. The study of Malak *et al*. (2016) on *Blastobotrys raffinosifermentans* indicated that different morphological forms (budding cells, pseudomycelia and mycelia) are observed under different temperatures [14]. When *Blastobotrys solioliva* IST508 was cultured at 40ºC, a supra-optimal temperature, differences in cells' morphology were observed. Some cells appear to be dead, while others continued to divide by budding in the beginning of incubation. However, it seems that the process stopped, since several cells are at the final stages of budding. This is in agreement with the poor growth observed at this temperature.

In *Blastobotrys solioliva* IST508 asexual reproduction occurs by budding. This evidence was confirmed by observations in the microscope using visible light or also by staining on the cells with Calcofluor White. Calcofluor white has been used for the specific staining of chitin along the yeast and fungal cell

walls, monitoring budding patterns and aging [92,97,98]. Chitin has an important role in cytokinesis in *S. cerevisiae*, since after linking with other cell wall components, forms the septum that separates mother and daughter cells [92,97]. In fact, in the cells stained with calcofluor white, regions with higher fluorescence intensity were observed. These regions were present between the mother cell wall and the bud. Also, higher fluorescence intensity was observed in some regions of the cell wall of single cells, which are the bud scars. Bud scars are formed after the division, where the remaining components (rests of chitin and other cell wall components) form a ring-structured bud scar composed of chitin on the surface of the mother cell [92,97], which under the microscope and after staining with calcofluor white emits fluorescence.

Based on the anaerobic fermentation tests for *Blastobotrys solioliva* IST508, it can ferment different sugars, including fructose and sucrose. However, the fermentation profiles are slow when compared to *S.cerevisiae* strains, the most commoly used yeats in fermentation processes [26,89] which indicates that this new yeast is not industrially competitive for fermentation. In terms of assimilation, which is the aerobic growth on different substrates [6], *Blastobotrys solioliva* IST508 showed the ability to assimilate several compounds, such as fructose, xylose, and arabinose [99]. This strain is also able to grow on vitamin-free media, indicating that it does not need expensive supplements added to the media.

Assimilation and fermentation rely on the ability to internalize the molecules [6], and on the presence of enzymes able to metabolize it. The cases where the strain was not able to grow or ferment a particular substrate were probably caused by the lack of proper plasma membrane transporters or specific enzymes. For example, the ability to utilize sucrose or raffinose indicates that, such as *S. cerevisiae*, this yeast must harbor an extracellular invertase that hydrolyzes sucrose into glucose and fructose, and raffinose into fructose, galactose and glucose, which are transported into the cell by hexose transporters and metabolized through glycolysis. In *S. cerevisiae*, the invertase may be encoded by one or several *SUC* genes (*SUC1–SUC5 and SUC7*) [100,101]. The analysis of *Blastobotrys solioliva* IST508 genome sequence confirmed the presence of two genes (BSOL0P01240 and BSOL0Q00110) sharing homology with *S. cerevisiae SUC* genes (results not shown).

The analysis of the culture media content, during aerobic growth, was also monitored. Strains of *Blastobotrys solioliva* IST508, *B. proliferans* PYCC5116 and *B. adeninivorans* PYCC4638, were monitored. These last two species were chosen to be also studied because *B. proliferans* is the closest relative of the new species and *B. adeninivorans* is one of the species more studied in the genus *Blastobotrys*. The HPLC analysis showed that *Blastobotrys solioliva* IST508 did not consume all the sugars available (xylose and glucose), whereas in the cases of *B. proliferans* PYCC5116 and *B. adeninivorans* PYCC4638, this was not observed. The growth curves of these three species are similar, but *Blastobotrys solioliva* IST508 reaches slightly smaller OD_{600nm}. This could indicate that cells loose viability before consuming all the sugar, which could be caused by a lack in a vital nutrient or by the accumulation of toxic compounds, like the organic acids [102,103]. Probably this strain is more susceptible to the presence of these compounds in comparison to the others tested. In the future, cell viability assays should be performed, since measurements of the optical density of the media only give insights on media turbidity and give no information on cell viability [104].

Ethanol and glycerol, which is a by-product of the fermentation of sugar to ethanol in a redox-neutral process [105], were never observed in any of the strains tested. Indicating that they are Crabtree negative, in contrast to *S. cerevisiae,* since they appear to be favoring respiration over fermentation. Yeasts that display a Crabtree effect ferment in the presence of oxygen and in high concentrations of glucose [106], which was not observed in any of the strains tested.

The production of organic acids was also studied, due to the high acidity of the culture media. *Blastobotrys solioliva* IST508 is capable to produce different organic acids, like formic acid, succinic acid, and acetic acid, in small concentrations and both in the presence of xylose or glucose. *Blastobotrys proliferans* PYCC5116 produces malic acid and formic acid, in 10-fold higher concentrations when compared to the other species tested, with glucose as carbon source. However the concentrations obtained in all the cases are considerably smaller (always less than 3 mM in IST508 case) when compared to other strains already used in industry, for example *Corynebacterium glutamicum* can produce up to 146 g/L (1.23 M) of succinic acid or *Aspergillus niger* that produces 240 g/L (1.25 M) of citric acid [107,108]. Also, in *B. adeninivorans* PYCC4638, it was observed that when glucose was depleted from the media, the concentrations of formic acid started to decrease, indicating that the strain was using it as carbon source.

The determination of the DNA content in a species provides information regarding molecular and cellular genetics of the cell and also provides data for the research of its genetic evolution [109]. Usually, the method used for determining cell ploidy (number of chromosome sets) is to assess the number of chromosomes in a single cell. However, this method tends to be time-intensive and laborious, and the results may occasionally be inexact. Flow cytometry has been used to determine the ploidy level and DNA content and appears to be an efficient, pinpoint and powerful method [109,110]. Measurements through flow cytometry allows to reveal the cell distribution within the major phases of the cell cycle (G0/G1 versus S versus G2/M) and disclose the DNA ploidy of the measured cell population [111]. With this analysis, *Blastobotrys solioliva* IST508 showed to have a more similar DNA content to the control haploid strain, which should mean that the novel species is probably haploid. The fluorescence observed in *Blastobotrys solioliva* IST508 cells, through flow cytometry, was equivalent to 14.24 Mb of DNA. The total DNA content of the scaffolds generated by the complete sequencing of its genome yielded 11.5 Mb. These differences could be due to problems involved with the cytometry analysis, since cells in the G2 phase were not identified. This work should be repeated by inducing the cells to arrest at the G2 phase, using nocodazole [112] . In the case of *Blastobotrys adeninivorans* PYCC4638 and *B. proliferans* PYCC5116, the cell cycle analysis is also similar to the control haploid strain, indicating that these are also haploid.

During culture of the novel species in YNB medium, it was observed that the liquid media turned bright yellow after a few hours of incubation and this color became more intense throughout growth, indicating that this yeast was producing and releasing a compound to the extracellular media. This unknown compound had to be water soluble and emit the yellow color and riboflavin was a candidate that reunited these characteristics. After running the UV spectrum of *Blastobotrys solioliva* IST508 culture media

(YNB) against the UV spectrum of riboflavin, it was observed that riboflavin had absorption peaks nearly at the same wavelengths of the compound present in the *Blastobotrys solioliva* IST508 growth medium. Production of this vitamin was only observed in a specific media, YNB, whereas no absorption peaks were detected in the YPD UV spectrum. This could indicate that the different media compositions affect this vitamin production. Stahmann *et al.* (2000) demonstrated that the presence of iron impaired riboflavin production [53]. In fact, YPD medium contains 25 μM of iron, while YNB medium has 1.2 μM, which accounts for a difference of over 20 times more iron in the first media [6,113]. Besides, YNB medium contains pyridoxine hydrochloride, which is converted into the active form pyridoxal 5'-phosphate that is an essential cofactor for the enzymatic reaction that transforms serine into glycine, which was also proved to be a positive influencer in riboflavin's production [114,115].

Besides, formic acid and riboflavin production appear to be related. The first reaction in the pathway of riboflavin biosynthesis leads to production of formic acid and this compound was the most produced organic acid in all the three species studied. It could indicate that all of them are able to produce this vitamin. However, in *B. adeninivorans* PYCC4638 and *B. proliferans* PYCC5116 the media did not change its color to yellow, so probably the vitamin is being consumed, in opposition to what occurs with *Blastobotrys solioliva* IST508 that excretes riboflavin to the external media. In addition, riboflavin's overproduction has been related to oxidative stress and aging of cells. This vitamin acts as a cell reducing agent in the cells [80,81]. It is likely that riboflavin overproduction in *B. solioliva* IST508 is related with the oxidative stress generated by the dissociation of the weak acids, like formic acid, in the cytosol of the cells that influence the free radical production and inhibiting cell metabolic activity, elucidating also why the cells stop consuming the sugar after some hours of incubation.

To gain more knowledge about the strain *Blastobotrys solioliva* IST508, its complete genome was sequenced, and the presence of genes involved with riboflavin production were verified. The rapid developments in DNA sequencing technologies and functional genomics, allows the investigation of the links between genotype and phenotype [108]. Initially the genes directly involved in the riboflavin biosynthesis pathway were studied. In the case of the novel *Blastobotrys* species, all the genes of this main biosynthesis pathway were found. Regarding *RIB2*, two *RIB2* gene homologs were detected during the BLAST analysis against *Blastobotrys solioliva* IST508 genome sequence. According to Ledesma-Amaro *et al*. (2015) the single gene overexpression of this gene can increase the vitamin's production [91]. However, protein motifs were only observed in one of the genes, probably due to their low identities and also the presence of gene copies alone might not affect the vitamin's production. Future work should involve the analysis of the complete gene expression, for example through RNA-sequencing [116] to better understand which genes are being expressed under an overexpression situation of the vitamin.

Recently, Kato *et al*. (2020) published an extensive study regarding the genomic analysis of a riboflavin *A. gossypii* overproducing mutant, in order to understand the metabolism changes underlying increased productivity [80]. From these findings, several genes were chosen to be studied in the genome of *Blastobotrys solioliva* IST508. *MET2*, *GLY1*, *SHM2*, *SEF1* and *DLD1* matched with two *Blastobotrys solioliva* IST508 genes, indicating that they could be duplicated. The protein motifs present on each one

of these genes were also studied. However, only for two of them, *SEF1* and *MET2*, there was a similarity in the protein family and in the motifs found. *MET2* is a homoserine acetyltransferase that catalyzes the first step unique to methionine biosynthesis, converting L-homoserine to O-acetyl-L-homoserine using acetyl-CoA as an acetyl group donor. This enzyme is also involved in the regulation of homoserine in other pathways, being vital to cell growth and viability [117]. If this duplication results in the increased concentration of O-acetyl-L-homoserine, this molecule could be channelled into glycine production, thus increasing its concentration in the cell, which is known to increase riboflavin production [52,53,118], however more tests should be done to confirm this hypothesis. *SEF1*, in *Candida famata*, is responsible for riboflavin synthesis regulation. It is believed that this gene is involved in the positive regulation of this pathway, since its overexpression is connected to an increase in this vitamin's production [81]. The product of *SEF1* gene is a transcriptional regulator that belongs to Gal4 family of transcriptional activators and possesses zinc finger domain at its N-terminus. This regulator activates *RIB1* and *RIB7* promoters. Besides this, *SEF1* is also involved in iron homeostasis in *C. famata.* Levels of *SEF1* depend on iron concentration, where an iron deficiency leads to its increase, and consequently of riboflavin production. In this case, the vitamin could reduce practically insoluble Fe^{3+} ions to more soluble Fe^{2+} ions providing iron to the cells [81].

In conclusion, this work allowed the description of a new yeast strain, *Blastobotrys solioliva* IST508, and the sequencing of its complete genome. It was found that this strain is able to assimilate and metabolize several substrates, like sucrose, fructose, arabinose present in a wide variety of feedstocks used in industry, like sugar cane, fruits, sugar beet, and lignocellulosic materials [119]. This strain, while growing on xylose, can produce riboflavin, which is an aspect that requires further studies in order to unveil its real biotechnological potential.

5. Conclusion and Future Perspectives

The increasing concerns with overpopulation, energetic needs and global warning renders the need to shift towards a circular economy, where wastes are re-used. The isolation and study of new non-conventional yeasts can help the change towards more sustainable bioprocesses. For these reasons, there has been an increase in the interest in non-*Saccharomyces* yeasts. From new metabolic pathways and new added-value products, to desirable characteristics with applications in industry, that will allow increasing yield and productivity of specific processes, can be unveiled by studying new non-conventional yeast strains.

This study confirmed the great diversity of non-conventional yeasts in the environment, since over twenty different species were isolated from 11 different origins, only accounting for the lab-cultivable strains, since many others could have been present but could not be isolated under the experimental conditions used [120]. Besides the isolation work of different strains, the novel species found, *Blastobotrys solioliva* IST508 represents an interesting candidate to be further explored. This new species could have potential for biotechnological processes, due to its ability to grow on a wide variety of compounds, including the ones present in agro-industrial wastes, like the pentoses: xylose and arabinose, fructose and others.

The most striking feature of *Blastobotrys solioliva* IST508 is its ability to produce riboflavin. Future work will involve the study of this yeast's capacity of producing this vitamin and other added-value products from it. The complete sequencing of this strain genome and its analysis allowed to confirm the presence of all the genes involved in the main pathway of this vitamin's production and also the presence of other genes that can affect riboflavin biosynthesis. Besides being present, some of these relevant genes seem to have paralogs in the genome that may be involved in riboflavin production. However, the transcriptional analysis of these and other genes, either through RT-PCR or RNA-sequencing, will be essential to understand which players are involved in riboflavin overproduction.

To conclude, the work performed in this thesis gave contributions to the field of non-conventional yeasts, especially in the case of strains able to consume xylose, and also allowed the description of a novel *Blastobotrys* species.

6. References

1. Gamero, A., Quintilla, R., Groenewald, M., Alkema, W., Boekhout, T. & Hazelwood, L. High-throughput screening of a large collection of non-conventional yeasts reveals their potential for aroma formation in food fermentation. *Food Microbiology* **60,** 147–159 (2016).

2. Cherubini, F. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Conversion and Management* **51,** 1412–1421 (2010).

3. Hong, K.-K. & Nielsen, J. Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. *Cellular and Molecular Life Sciences* **69,** 2671–2690 (2012).

4. Nandy, S. K. & Srivastava, R. K. A review on sustainable yeast biotechnological processes and applications. *Microbiological Research* **207,** 83–90 (2018).

5. de Souza Varize, C., Maria Christofoleti-Furlan, R., de Souza Miranda Muynarsk, E., Vinícius de Melo Pereira, G., Dantas Lopes, L. & Carlos Basso, L. in *Yeasts in Biotechnology* (IntechOpen, 2019). doi:10.5772/intechopen.83035

6. Kurtzman, C., Fell, J. W. & Boekhout, T. *The yeasts: a taxonomic study*. *Elsevier* **1,** (Elsevier B.V., 2011).

7. Klopotek, A. Blastobotrys nivea gen.nov., sp.nov. *Archiv for Mikrobiologie* **58,** 92–96 (1967).

8. Kurtzman, C. P. Blastobotrys americana sp. nov., Blastobotrys illinoisensis sp. nov., Blastobotrys malaysiensis sp. nov., Blastobotrys muscicola sp. nov., Blastobotrys peoriensis sp. nov. and Blastobotrys raffinosifermentans sp. nov., novel anamorphic yeast species. *International Journal of Systematic and Evolutionary Microbiology* **57,** 1154–1162 (2007).

9. Nouri, H., Moghimi, H., Geranpayeh Vaghei, M. & Nasr, S. Blastobotrys persicus sp. nov., an ascomycetous yeast species isolated from cave soil. *Antonie van Leeuwenhoek* **111,** 517–524 (2018).

10. Barretto, D. A., Avchar, R., Carvalho, C., Sampaio, J. P., Vootla, S. K. & Baghela, A. Blastobotrys bombycis sp. nov., a d-xylose-fermenting yeast isolated from the gut of the silkworm larva Bombyx mori. *International Journal of Systematic and Evolutionary Microbiology* **68,** 2638–2643 (2018).

11. Crous, P. W., Wingfield, M. J., Richardson, D. M., Leroux, J. J., Strasberg, D., Edwards, J., Roets, F., Hubka, V., Taylor, P. W. J., Heykoop, M., Martín, M. P., Moreno, G., Sutton, D. A., Wiederhold, N. P., Barnes, C. W., Carlavilla, J. R., Gené, J., Giraldo, A., Guarnaccia, V., Guarro, J., Hernández-Restrepo, M., Kolařík, M., Manjón, J. L., Pascoe, I. G., Popov, E. S., Sandoval-Denis, M., Woudenberg, J. H. C., Acharya, K., Alexandrova, A. V., Alvarado, P., Barbosa, R. N., Baseia, I. G., Blanchette, R. A., Boekhout, T., Burgess, T. I., Cano-Lira, J. F., Čmoková, A., Dimitrov, R. A., Dyakov, M. Y., Dueñas, M., Dutta, A. K., Esteve-Raventós, F., Fedosova, A. G.,

Fournier, J., Gamboa, P., Gouliamova, D. E., Grebenc, T., Groenewald, M., Hanse, B., Hardy, G. E. S. J., Held, B. W., Jurjević, Ž., Kaewgrajang, T., Latha, K. P. D., Lombard, L., Luangsa-ard, J. J., Lysková, P., Mallátová, N., Manimohan, P., Miller, A. N., Mirabolfathy, M., Morozova, O. V., Obodai, M., Oliveira, N. T., Ordóñez, M. E., Otto, E. C., Paloi, S., Peterson, S. W., Phosri, C., Roux, J., Salazar, W. A., Sánchez, A., Sarria, G. A., Shin, H.-D., Silva, B. D. B., Silva, G. A., Smith, M. T., Souza-Motta, C. M., Stchigel, A. M., Stoilova-Disheva, M. M., Sulzbacher, M. A., Telleria, M. T., Toapanta, C., Traba, J. M., Valenzuela-Lopez, N., Watling, R. & Groenewald, J. Z. Fungal Planet description sheets: 400–468. *Persoonia - Molecular Phylogeny and Evolution of Fungi* **36,** 128 (2016).

12. Wang, D., Xu, Y., Hu, J. & Zhao, G. Fermentation Kinetics of Different Sugars by Apple Wine Yeast Saccharomyces cerevisiae. *Journal of the Institute of Brewing* **110,** 340–346 (2004).

13. Kunze, G., Gaillardin, C., Czernicka, M., Durrens, P., Martin, T., Böer, E., Gabaldón, T., Cruz, J. A., Talla, E., Marck, C., Goffeau, A., Barbe, V., Baret, P., Baronian, K., Beier, S., Bleykasten, C., Bode, R., Casaregola, S., Despons, L., Fairhead, C., Giersberg, M., Gierski, P. P., Hähnel, U., Hartmann, A., Jankowska, D., Jubin, C., Jung, P., Lafontaine, I., Leh-Louis, V., Lemaire, M., Marcet-Houben, M., Mascher, M., Morel, G., Richard, G.-F., Riechen, J., Sacerdot, C., Sarkar, A., Savel, G., Schacherer, J., Sherman, D. J., Stein, N., Straub, M.-L., Thierry, A., Trautwein-Schult, A., Vacherie, B., Westhof, E., Worch, S., Dujon, B., Souciet, J.-L., Wincker, P., Scholz, U. & Neuvéglise, C. The complete genome of Blastobotrys (Arxula) adeninivorans LS3 - a yeast of biotechnological interest. *Biotechnology for Biofuels* **7,** 66 (2014).

14. Malak, A., Baronian, K. & Kunze, G. Blastobotrys (Arxula) adeninivorans : a promising alternative yeast for biotechnology and basic research. *Yeast* **33,** 535–547 (2016).

15. Antonopoulou, G., Kampranis, A., Ntaikou, I. & Lyberatos, G. Enhancement of Liquid and Gaseous Biofuels Production From Agro-Industrial Residues After Thermochemical and Enzymatic Pretreatment. *Frontiers in Sustainable Food Systems* **3,** (2019).

16. Do, D. T. H., Theron, C. W. & Fickers, P. Organic Wastes as Feedstocks for Non-Conventional Yeast-Based Bioprocesses. *Microorganisms* **7,** 229 (2019).

17. Li, M. & Borodina, I. Application of synthetic biology for production of chemicals in yeast Saccharomyces cerevisiae. *FEMS Yeast Research* (2014). doi:10.1111/1567-1364.12213

18. Gao, R., Li, Z., Zhou, X., Bao, W., Cheng, S. & Zheng, L. Enhanced lipid production by Yarrowia lipolytica cultured with synthetic and waste-derived high-content volatile fatty acids under alkaline conditions. *Biotechnology for Biofuels* **13,** 3 (2020).

19. Raud, M., Kikas, T., Sippula, O. & Shurpali, N. J. Potentials and challenges in lignocellulosic biofuel production technology. *Renewable and Sustainable Energy Reviews* **111,** 44–56 (2019).

20. dos Santos, S. C. Yeast toxicogenomics: genome-wide responses to chemical stresses with impact in environmental health, pharmacology, and biotechnology. *Frontiers in Genetics* **3,**

(2012).

21. Nandal, P., Sharma, S. & Arora, A. Bioprospecting non-conventional yeasts for ethanol production from rice straw hydrolysate and their inhibitor tolerance. *Renewable Energy* **147,** 1694–1703 (2020).

22. Scordia, D., Cosentino, S. L., Lee, J.-W. & Jeffries, T. W. Bioconversion of giant reed (Arundo donax L.) hemicellulose hydrolysate to ethanol by Scheffersomyces stipitis CBS6054. *Biomass and Bioenergy* **39,** 296–305 (2012).

23. Jeffries, T. W., Grigoriev, I. V, Grimwood, J., Laplaza, J. M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., Jin, Y.-S., Passoth, V. & Richardson, P. M. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis. *Nature Biotechnology* **25,** 319–326 (2007).

24. Kurtzman, C. P. in *The Yeasts* 521–543 (Elsevier, 2011). doi:10.1016/B978-0-444-52149-1.00042-2

25. Martins, G. M., Bocchini-Martins, D. A., Bezzerra-Bussoli, C., Pagnocca, F. C., Boscolo, M., Monteiro, D. A., Silva, R. da & Gomes, E. The isolation of pentose-assimilating yeasts and their xylose fermentation potential. *Brazilian Journal of Microbiology* **49,** 162–168 (2018).

26. Kumar, A., Singh, L. K. & Ghosh, S. Bioconversion of lignocellulosic fraction of water-hyacinth (Eichhornia crassipes) hemicellulose acid hydrolysate to ethanol by Pichia stipitis. *Bioresource Technology* **100,** 3293–3297 (2009).

27. Hurst, C. J. in *Understanding Terrestrial Microbial Communities. Advances in Environmental Microbiology* 289–405 (Springer, 2019). doi:10.1007/978-3-030-10777-2_9

28. Shenoy, B. D., Jeewon, R. & Hyde, K. D. Impact of DNA sequence-data on the taxonomy of anamorphic fungi. *Fungal Diversity* **26,** 1–54 (2007).

29. Kurtzman, C. P. & Robnett, C. J. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **73,** 331–371 (1998).

30. Peter, G., Dlauchy, D., Price, N. P. J. & Kurtzman, C. P. Diddensiella caesifluorescens gen. nov., sp. nov., a riboflavin-producing yeast species of the family Trichomonascaceae. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* **62,** 3081–3087 (2012).

31. Marvanová, L. Two new Blastobotrys species. *Transactions of the British Mycological Society* **66,** 217–222 (1976).

32. Middelhoven, W. J., Niet, M. C. H.-T. & Rij, N. J. W. K.-V. Trichosporon adeninovorans sp. nov., a yeast species utilizing adenine, xanthine, uric acid, putrescine and primary n-alkylamines as the sole source of carbon, nitrogen and energy. *Antonie van Leeuwenhoek* **50,** 369–378 (1984).

33. Middelhoven, W. J., de Jong, I. M. & de Winter, M. Arxula adeninivorans, a yeast assimilating many nitrogenous and aromatic compounds. *Antonie van Leeuwenhoek* **59,** 129–137 (1991).

34. Kaiser, C., Uhlig, S., Gerlach, T., Körner, M., Simon, K., Kunath, K., Florschütz, K., Baronian, K. & Kunze, G. Evaluation and validation of a novel Arxula adeninivorans estrogen screen (nAES) assay and its application in analysis of wastewater, seawater, brackish water and urine. *Science of The Total Environment* **408,** 6017–6026 (2010).

35. Gienow, U., Kunze, G., Schauer, F., Bode, R. & Hofemeister, J. The Yeast Genus Trichosporon spec. LS3; Molecular characterization of genomic complexity. *Zentralblatt für Mikrobiologie* **145,** 3–12 (1990).

36. Thomas, S., Sanya, D. R. A., Fouchard, F., Nguyen, H.-V., Kunze, G., Neuvéglise, C. & Crutz-Le Coq, A.-M. Blastobotrys adeninivorans and B. raffinosifermentans, two sibling yeast species which accumulate lipids at elevated temperatures and from diverse sugars. *Biotechnology for Biofuels* **12,** 154 (2019).

37. Jankowska, D. A., Trautwein-Schult, A., Cordes, A., Bode, R., Baronian, K. & Kunze, G. A novel enzymatic approach in the production of food with low purine content using Arxula adeninivorans endogenous and recombinant purine degradative enzymes. *Bioengineered* **6,** 20–25 (2015).

38. Jean, S., Lainhart, W., Anderson, N. W., Yarbrough, M. L. & Burnham, C.-A. D. Clinical impact of molecular identification of rare yeasts and nonsporulating molds recovered in culture from clinical specimens. *Diagnostic Microbiology and Infectious Disease* **96,** 114945 (2020).

39. Brown, S. P., Cornforth, D. M. & Mideo, N. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends in Microbiology* **20,** 336–342 (2012).

40. Schwarz, C., Brandt, C., Whitaker, P., Sutharsan, S., Skopnik, H., Gartner, S., Smazny, C. & Röhmel, J. F. Invasive Pulmonary Fungal Infections in Cystic Fibrosis. *Mycopathologia* **183,** 33–43 (2018).

41. Huyben, D., Boqvist, S., Passoth, V., Renström, L., Allard Bengtsson, U., Andréoletti, O., Kiessling, A., Lundh, T. & Vågsholm, I. Screening of intact yeasts and cell extracts to reduce Scrapie prions during biotransformation of food waste. *Acta Veterinaria Scandinavica* **60,** 9 (2018).

42. Meier, A. K., Worch, S., Böer, E., Hartmann, A., Mascher, M., Marzec, M., Scholz, U., Riechen, J., Baronian, K., Schauer, F., Bode, R. & Kunze, G. Agdc1p – a Gallic Acid Decarboxylase Involved in the Degradation of Tannic Acid in the Yeast Blastobotrys (Arxula) adeninivorans. *Frontiers in Microbiology* **8,** (2017).

43. Mat, S. ., Daud, M., Rojie, M., Hussain, H. & Rukayadi, R. Effects of Candida sp. and Blastobotrys sp. Starter on Fermentation of Cocoa (Theobroma cacao L.) Beans and Its Antibacterial Activity. *Journal of Pure and Applied Microbiology* **10,** 2501–2510 (2016).

44. Azam, S. M., Yu, D. & Wu, A. in *Food Safety and Mycotoxins* (ed. Wu, A.) 113–135 (Springer Singapore, 2019). doi:10.1007/978-981-32-9038-9

45. Parasuraman, P., Devadatha, B., Sarma, V. V., Ranganathan, S., Ampasala, D. R. & Siddhardha, B. Anti-quorum sensing and antibiofilm activities of Blastobotrys parvus PPR3 against Pseudomonas aeruginosa PAO1. *Microbial Pathogenesis* **138,** 103811 (2020).

46. Ayadi, I., Belghith, H., Gargouri, A. & Guerfali, M. Screening of new oleaginous yeasts for single cell oil production, hydrolytic potential exploitation and agro-industrial by-products valorization. *Process Safety and Environmental Protection* **119,** 104–114 (2018).

47. Viñarta, S. C., Angelicola, M. V., Barros, J. M., Fernández, P. M., Mac Cormak, W., Aybar, M. J. & de Figueroa, L. I. C. Oleaginous yeasts from Antarctica: Screening and preliminary approach on lipid accumulation. *Journal of Basic Microbiology* **56,** 1360–1368 (2016).

48. Vicente, G., Bautista, L. F., Rodríguez, R., Gutiérrez, F. J., Sádaba, I., Ruiz-Vázquez, R. M., Torres-Martínez, S. & Garre, V. Biodiesel production from biomass of an oleaginous fungus. *Biochemical Engineering Journal* **48,** 22–27 (2009).

49. Sitepu, I. R., Garay, L. A., Sestric, R., Levin, D., Block, D. E., German, J. B. & Boundy-Mills, K. L. Oleaginous yeasts for biodiesel: Current and future trends in biology and production. *Biotechnology Advances* **32,** 1336–1360 (2014).

50. Xue, S., Chi, Z.-M., Zhang, Y., Li, Y.-F., Liu, G., Jiang, H., Hu, Z. & Chi, Z.-M. Fatty acids from oleaginous yeasts and yeast-like fungi and their potential applications. *Critical Reviews in Biotechnology* **38,** 1049–1060 (2018).

51. Thakur, K., Tomar, S. K. & De, S. Lactic acid bacteria as a cell factory for riboflavin production. *Microbial Biotechnology* **9,** 441–451 (2016).

52. Liu, S., Hu, W., Wang, Z. & Chen, T. Production of riboflavin and related cofactors by biotechnological processes. *Microbial Cell Factories* **19,** 31 (2020).

53. Stahmann, K. P., Revuelta, J. L. & Seulberger, H. Three biotechnical processes using Ashbya gossypii, Candida famata, or Bacillus subtilis compete with chemical riboflavin production. *Applied Microbiology and Biotechnology* **53,** 509–516 (2000).

54. Burgess, C., O'Connell-Motherway, M., Sybesma, W., Hugenholtz, J. & van Sinderen, D. Riboflavin Production in Lactococcus lactis: Potential for In Situ Production of Vitamin-Enriched Foods. *Applied and Environmental Microbiology* **70,** 5769–5777 (2004).

55. Schwechheimer, S. K., Park, E. Y., Revuelta, J. L., Becker, J. & Wittmann, C. Biotechnology of riboflavin. *Applied Microbiology and Biotechnology* **100,** 2107–2119 (2016).

56. Revuelta, J. L., Ledesma-Amaro, R., Lozano-Martinez, P., Díaz-Fernández, D., Buey, R. M. & Jiménez, A. Bioproduction of riboflavin: a bright yellow history. *Journal of Industrial Microbiology & Biotechnology* **44,** 659–665 (2017).

57. Suzuki, G. T., Fleuri, L. & Macedo, G. A. Influence of Nitrogen and Carbon Sources on Riboflavin Production by Wild Strain of Candida sp. *Food and Bioprocess Technology* **5,** 466–473 (2012).

58. Silva, R., Aguiar, T. Q., Oliveira, R. & Domingues, L. Light exposure during growth increases riboflavin production, ROS accumulation and DNA damage in Ashbya gossypii riboflavin-overproducing strains. *FEMS Yeast Research* (2018). doi:10.1093/femsyr/foy114

59. Solovieva, I. M., Kreneva, R. A., Leak, D. J. & Perumov, D. A. The ribR gene encodes a monofunctional riboflavin kinase which is involved in regulation of the Bacillus subtilis riboflavin operon. *Microbiology* **145,** 67–73 (1999).

60. Schwechheimer, S. K., Becker, J., Peyriga, L., Portais, J.-C. & Wittmann, C. Metabolic flux analysis in Ashbya gossypii using 13C-labeled yeast extract: industrial riboflavin production under complex nutrient conditions. *Microbial Cell Factories* **17,** 162 (2018).

61. Riley, R., Haridas, S., Wolfe, K. H., Lopes, M. R., Hittinger, C. T., Göker, M., Salamov, A. A., Wisecaver, J. H., Long, T. M., Calvey, C. H., Aerts, A. L., Barry, K. W., Choi, C., Clum, A., Coughlan, A. Y., Deshpande, S., Douglass, A. P., Hanson, S. J., Klenk, H.-P., LaButti, K. M., Lapidus, A., Lindquist, E. A., Lipzen, A. M., Meier-Kolthoff, J. P., Ohm, R. A., Otillar, R. P., Pangilinan, J. L., Peng, Y., Rokas, A., Rosa, C. A., Scheuner, C., Sibirny, A. A., Slot, J. C., Stielow, J. B., Sun, H., Kurtzman, C. P., Blackwell, M., Grigoriev, I. V. & Jeffries, T. W. Comparative genomics of biotechnologically important yeasts. *Proceedings of the National Academy of Sciences* **113,** 9882–9887 (2016).

62. Löbs, A.-K., Schwartz, C. & Wheeldon, I. Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. *Synthetic and Systems Biotechnology* **2,** 198–207 (2017).

63. Sharma, K. K. in *Developments in Fungal Biology and Applied Mycology* 201–226 (Springer Singapore, 2017). doi:10.1007/978-981-10-4768-8_11

64. Ansorge, W. J., Katsila, T. & Patrinos, G. P. in *Molecular Diagnostics* 141–153 (Elsevier, 2017). doi:10.1016/B978-0-12-802971-8.00008-0

65. Ambardar, S., Gupta, R., Trakroo, D., Lal, R. & Vakhlu, J. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology* **56,** 394–404 (2016).

66. Allona, I., Kirst, M., Boerjan, W., Strauss, S. & Sederoff, R. Editorial: Forest Genomics and Biotechnology. *Frontiers in Plant Science* **10,** (2019).

67. Sharma, K. K. Fungal genome sequencing: basic biology to biotechnology. *Critical Reviews in Biotechnology* 1–17 (2015). doi:10.3109/07388551.2015.1015959

68. Corre, C. & Challis, G. L. in *Comprehensive Natural Products II* 429–453 (Elsevier, 2010). doi:10.1016/B978-008045382-8.00046-0

69. Zaky, A. S., Greetham, D., Louis, E. J., Tucker, G. A. & Du, C. A New Isolation and Evaluation

Method for Marine-Derived Yeast spp. with Potential Applications in Industrial Biotechnology. *Journal of Microbiology and Biotechnology* **26,** 1891–1907 (2016).

70. Sláviková, E. & Vadkertiová, R. The diversity of yeasts in the agricultural soil. *Journal of Basic Microbiology* **43,** 430–436 (2003).

71. Kurtzman, C. P. & Robnett, C. J. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. *FEMS Yeast Research* **3,** 417–432 (2003).

72. Borneman, J. & Hartin, R. J. PCR Primers That Amplify Fungal rRNA Genes from Environmental Samples. *Applied and Environmental Microbiology* **66,** 4356–4360 (2000).

73. Soukup, A. in *Methods in Molecular Biology* 27–42 (2019). doi:10.1007/978-1-4939-9469-4_2

74. Fortuna, M., João Sousa, M., Côrte-Real, M., Leão, C., Salvador, A. & Sansonetty, F. Cell Cycle Analysis of Yeasts. *Current Protocols in Cytometry* **13,** (2000).

75. Harsch, M. J., Lee, S. A., Goddard, M. R. & Gardner, R. C. Optimized fermentation of grape juice by laboratory strains of Saccharomyces cerevisiae. *FEMS Yeast Research* **10,** 72–82 (2010).

76. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27,** 863–864 (2011).

77. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. & Pevzner, P. A. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19,** 455–477 (2012).

78. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).

79. Proux-Wéra, E., Armisén, D., Byrne, K. P. & Wolfe, K. H. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics* **13,** 237 (2012).

80. Kato, T., Azegami, J., Yokomori, A., Dohra, H., El Enshasy, H. A. & Park, E. Y. Genomic analysis of a riboflavin-overproducing Ashbya gossypii mutant isolated by disparity mutagenesis. *BMC Genomics* **21,** 319 (2020).

81. Lyzak, O. O., Ledesma-Amaro, R., Dmytruk, K. V., Sibirny, A. A. & Revuelta, J. L. in *Biotechnology of Yeasts and Filamentous Fungi* 281–296 (Springer International Publishing, 2017). doi:10.1007/978-3-319-58829-2_9

82. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28,** 27–30 (2000).

83. Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *Journal of*

Molecular Biology **147,** 195–197 (1981).

84. Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. & Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47,** W636–W641 (2019).

85. Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L. & Xenarios, I. New and continuing developments at PROSITE. *Nucleic Acids Research* **41,** D344–D347 (2012).

86. de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A. & Hulo, N. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research* **34,** W362–W365 (2006).

87. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E. & Finn, R. D. The Pfam protein families database in 2019. *Nucleic Acids Research* **47,** D427–D432 (2019).

88. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. & Hunter, S. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30,** 1236–1240 (2014).

89. LIU, R. & SHEN, F. Impacts of main factors on bioethanol fermentation from stalk juice of sweet sorghum by immobilized Saccharomyces cerevisiae (CICC 1308). *Bioresource Technology* **99,** 847–854 (2008).

90. Nunez, R. DNA Measurement and Cell Cycle Analysis by Flow Cytometry. *Current Issues in Molecular Biology* (2001). doi:10.21775/cimb.003.067

91. Ledesma-Amaro, R., Serrano-Amatriain, C., Jiménez, A. & Revuelta, J. L. Metabolic engineering of riboflavin production in Ashbya gossypii through pathway optimization. *Microbial Cell Factories* **14,** 163 (2015).

92. Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., Bolchacova, E., Voigt, K., Crous, P. W., Miller, A. N., Wingfield, M. J., Aime, M. C., An, K.-D., Bai, F.-Y., Barreto, R. W., Begerow, D., Bergeron, M.-J., Blackwell, M., Boekhout, T., Bogale, M., Boonyuen, N., Burgaz, A. R., Buyck, B., Cai, L., Cai, Q., Cardinali, G., Chaverri, P., Coppins, B. J., Crespo, A., Cubas, P., Cummings, C., Damm, U., de Beer, Z. W., de Hoog, G. S., Del-Prado, R., Dentinger, B., Dieguez-Uribeondo, J., Divakar, P. K., Douglas, B., Duenas, M., Duong, T. A., Eberhardt, U., Edwards, J. E., Elshahed, M. S., Fliegerova, K., Furtado, M., Garcia, M. A., Ge, Z.-W., Griffith, G. W., Griffiths, K., Groenewald, J. Z., Groenewald, M., Grube, M., Gryzenhout, M., Guo, L.-D., Hagen, F., Hambleton, S., Hamelin, R. C., Hansen, K., Harrold, P.,

Heller, G., Herrera, C., Hirayama, K., Hirooka, Y., Ho, H.-M., Hoffmann, K., Hofstetter, V., Hognabba, F., Hollingsworth, P. M., Hong, S.-B., Hosaka, K., Houbraken, J., Hughes, K., Huhtinen, S., Hyde, K. D., James, T., Johnson, E. M., Johnson, J. E., Johnston, P. R., Jones, E. B. G., Kelly, L. J., Kirk, P. M., Knapp, D. G., Koljalg, U., Kovacs, G. M., Kurtzman, C. P., Landvik, S., Leavitt, S. D., Liggenstoffer, A. S., Liimatainen, K., Lombard, L., Luangsa-ard, J. J., Lumbsch, H. T., Maganti, H., Maharachchikumbura, S. S. N., Martin, M. P., May, T. W., McTaggart, A. R., Methven, A. S., Meyer, W., Moncalvo, J.-M., Mongkolsamrit, S., Nagy, L. G., Nilsson, R. H., Niskanen, T., Nyilasi, I., Okada, G., Okane, I., Olariaga, I., Otte, J., Papp, T., Park, D., Petkovits, T., Pino-Bodas, R., Quaedvlieg, W., Raja, H. A., Redecker, D., Rintoul, T. L., Ruibal, C., Sarmiento-Ramirez, J. M., Schmitt, I., Schussler, A., Shearer, C., Sotome, K., Stefani, F. O. P., Stenroos, S., Stielow, B., Stockinger, H., Suetrong, S., Suh, S.-O., Sung, G.-H., Suzuki, M., Tanaka, K., Tedersoo, L., Telleria, M. T., Tretter, E., Untereiner, W. A., Urbina, H., Vagvolgyi, C., Vialle, A., Vu, T. D., Walther, G., Wang, Q.-M., Wang, Y., Weir, B. S., Weiss, M., White, M. M., Xu, J., Yahr, R., Yang, Z. L., Yurkov, A., Zamora, J.-C., Zhang, N., Zhuang, W.-Y. & Schindel, D. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* **109,** 6241–6246 (2012).

93. Krogerus, K., Magalhães, F., Vidgren, V. & Gibson, B. Novel brewing yeast hybrids: creation and application. *Applied Microbiology and Biotechnology* **101,** 65–78 (2017).

94. Fernández-González, A. J., Villadas, P. J., Gómez-Lama Cabanás, C., Valverde-Corredor, A., Belaj, A., Mercado-Blanco, J. & Fernández-López, M. Defining the root endosphere and rhizosphere microbiomes from the World Olive Germplasm Collection. *Scientific Reports* **9,** 20423 (2019).

95. Fausto, C., Mininni, A. N., Sofo, A., Crecchio, C., Scagliola, M., Dichio, B. & Xiloyannis, C. Olive orchard microbiome: characterisation of bacterial communities in soil-plant compartments and their comparison between sustainable and conventional soil management systems. *Plant Ecology & Diversity* **11,** 597–610 (2018).

96. Codón, A. C., Gasent-Ramírez, J. M. & Benítez, T. Factors which affect the frequency of sporulation and tetrad formation in Saccharomyces cerevisiae baker's yeasts. *Applied and environmental microbiology* **61,** 630–8 (1995).

97. Arcones, I. & Roncero, C. in *Methods in Molecular Biology* 59–72 (2016). doi:10.1007/978-1-4939-3145-3_5

98. Okada, H. & Ohya, Y. Fluorescent Labeling of Yeast Cell Wall Components. *Cold Spring Harbor Protocols* **2016,** pdb.prot085241 (2016).

99. Peters, D. in *White Biotechnology. Advances in Biochemical Engineering/Biotechnology* 1–30 (Springer Berlin Heidelberg, 2006). doi:10.1007/10_031

100. Batista, A. S., Miletti, L. C. & Stambuk, B. U. Sucrose Fermentation by Saccharomyces cerevisiae Lacking Hexose Transport. *Journal of Molecular Microbiology and Biotechnology* **8,**

26–33 (2004).

101. Hackel, R. A. & Khan, N. A. Genetic control of invertase formation in Saccharomyces cerevisiae. *Molecular and General Genetics MGG* **164,** 295–302 (1978).

102. Piper, P., Calderon, C. O., Hatzixanthis, K. & Mollapour, M. Weak acid adaptation: the stress response that confers yeasts with resistance to organic acid food preservatives. *Microbiology* **147,** 2635–2642 (2001).

103. Teixeira, M. C., Mira, N. P. & Sá-Correia, I. A genome-wide perspective on the response and tolerance to food-relevant stresses in Saccharomyces cerevisiae. *Current Opinion in Biotechnology* **22,** 150–156 (2011).

104. Sutton, S. Measurement of microbial cells by optical density. *J. Validation Techn* (2011). at <http://www.microbiologynetwork.com/content/JVT_2011_v17n1_Measurement-of-Microbial-Cells-by-Optical-Density.pdf>

105. Wang, Z., Zhuge, J., Fang, H. & Prior, B. A. Glycerol production by microbial fermentation. *Biotechnology Advances* **19,** 201–223 (2001).

106. Pfeiffer, T. & Morley, A. An evolutionary perspective on the Crabtree effect. *Frontiers in Molecular Biosciences* **1,** (2014).

107. Becker, J., Lange, A., Fabarius, J. & Wittmann, C. Top value platform chemicals: bio-based production of organic acids. *Current Opinion in Biotechnology* **36,** 168–175 (2015).

108. Chen, Y. & Nielsen, J. Biobased organic acids production by metabolically engineered microorganisms. *Current Opinion in Biotechnology* **37,** 165–172 (2016).

109. Yan, J., Zhang, J., Sun, K., Chang, D., Bai, S., Shen, Y., Huang, L., Zhang, J., Zhang, Y. & Dong, Y. Ploidy Level and DNA Content of Erianthus arundinaceus as Determined by Flow Cytometry and the Association with Biological Characteristics. *PLOS ONE* **11,** e0151948 (2016).

110. Liu, Q., Zhang, C., Zhou, J., Dong, Q., Huo, L. & Dong, H. One simple, rapid and economical method for ploidy detection of Trichogramma dendrolimi Matsumura (Hymenoptera Trichogrammatidae). *Journal of Asia-Pacific Entomology* **23,** 345–349 (2020).

111. Darzynkiewicz, Z., Huang, X. & Zhao, H. Analysis of Cellular DNA Content by Flow Cytometry. *Current Protocols in Immunology* **119,** (2017).

112. Palma, M., Münsterkötter, M., Peça, J., Güldener, U. & Sá-Correia, I. Genome sequence of the highly weak-acid-tolerant Zygosaccharomyces bailii IST302, amenable to genetic manipulations and physiological studies. *FEMS Yeast Research* **17,** (2017).

113. Du, Y., Cheng, W. & Li, W.-F. Expression profiling reveals an unexpected growth-stimulating effect of surplus iron on the yeast Saccharomyces cerevisiae. *Molecules and Cells* **34,** 127–132 (2012).

114. PubChem Database NCBI. Pyridoxine hydrochloride, CID=6019. *National Center for Biotechnology Information* at <https://pubchem.ncbi.nlm.nih.gov/compound/Pyridoxine-hydrochloride>

115. Zou, Y., Chen, T., Feng, L., Zhang, S., Xing, D. & Wang, Z. Enhancement of 5-aminolevulinic acid production by metabolic engineering of the glycine biosynthesis pathway in Corynebacterium glutamicum. *Biotechnology Letters* **39,** 1369–1374 (2017).

116. Svensson, V., Natarajan, K. N., Ly, L.-H., Miragaia, R. J., Labalette, C., Macaulay, I. C., Cvejic, A. & Teichmann, S. A. Power analysis of single-cell RNA-sequencing experiments. *Nature Methods* **14,** 381–387 (2017).

117. Born, T. L., Franklin, M. & Blanchard, J. S. Enzyme-Catalyzed Acylation of Homoserine: Mechanistic Characterization of the Haemophilus influenzae met2 -Encoded Homoserine Transacetylase †. *Biochemistry* **39,** 8556–8564 (2000).

118. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45,** D353–D361 (2017).

119. Balat, M. Production of bioethanol from lignocellulosic materials via the biochemical pathway: A review. *Energy Conversion and Management* **52,** 858–875 (2011).

120. Bodor, A., Bounedjoum, N., Vincze, G. E., Erdeiné Kis, Á., Laczi, K., Bende, G., Szilágyi, Á., Kovács, T., Perei, K. & Rákhely, G. Challenges of unculturable bacteria: environmental perspectives. *Reviews in Environmental Science and Bio/Technology* **19,** 1–22 (2020).

7.Annexes

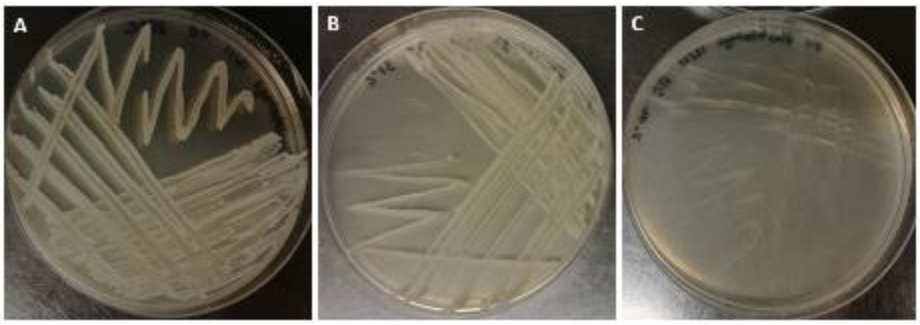

Figure A1 – Temperature effect in Blastobotrys sp. Growth in malt extract agar, after 3 days of incubation, at 25ºC (A), 37ºC (B) and 40ºC (C).