Exploring the non-conventional yeast species *Candida ethanolica* for biotechnological applications: a physiological and genomic analysis

Marta Cristina Beja Ferreira
amarta.beja.ferreira@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

July 2018

Abstract

Non-*Saccharomyces* yeast species are emerging as highly interesting cell factories in the biotechnology field where they can be used to produce add-value chemicals. However, the rational utilization of these yeasts is difficulted by a generalized lack of knowledge of their genetics and, thereby, of their physiology. The present work is focused on the exploration of one of such poorly characterized non-conventional yeast species, *Candida ethanolica*.Through a phenotypic screening that examined relevant traits of hosts to be explored in industrial biotechnology (such as ability to use different carbon sources, tolerance to environmental stressors or growth under O_2-limited conditions) a strain of *C. ethanolica*, BJK_1A, isolated from an instance of vinegar production, emerged by showing the ability to grow in glucose, fructose, mannose, galactose, glycerol, xylose and arabinose and also by showing increased resilience (comparing with *S. cerevisiae* which was used as a control) to acetic acid, a key inhibitor present in lignocellulosic hydrolysates. *C. ethanolica* BJK_1A also exhibited increased tolerance to levulinic (which is also present as inhibitor in lignocellulosic hydrolysates), and to succinic acids. Consistent with its high resilience to inhibitors present in lignocellulosic hydrolysates, *C. ethanolica* BJK_1A exhibited a robust growth in hydrolysates obtained from wheat straw and oak sawdust. The interesting phenotypic repertoire of *C. ethanolica* BJK_1A prompted us to obtain its genomic sequence and subsequently annotate it, this being the first time that an annotated version of the genome is studied for this yeast species. A preliminary analysis of this genome and its structure is provided in this work.
Keywords: Non-*Saccharomyces* yeasts, *Candida ethanolica*, lignocellulosic hydrolysates, genome sequencing and annotation

1. Introduction

While the model yeast *Saccharomyces cerevisiae* is an excellent host for the conversion of glucose to ethanol, production of other chemicals from alternative substrates often requires extensive strain engineering. To avoid complex and intensive engineering of *S. cerevisiae*, other yeasts are often selected as hosts for bioprocessing based on their natural capacity to produce a desired product: for example, the efficient production and secretion of proteins, lipids, and primary metabolites that have value as commodity chemicals. Good examples of this are the non-conventional yeast species *Kluyveromyces lactis*, *K. marxianus*, *Scheffersomyces stipitis*, *Yarrowia lipolytica*, *Hansenula polymorpha* and *Pichia pastoris* that have been developed as eukaryotic hosts because of their desirable phenotypes, including thermotolerance, assimilation of diverse carbon sources, and high protein secretion. Metabolic engineering in these yeasts is, however, more challenging in comparison with S. cerevisiae, because less is known about their metabolism and genomics, and advanced genetic engineering tools are limited [20] [10].

When exposed to the harsh conditions of fermentation processes in bioreactors yeasts need to survive a large amount of stress conditions, including osmotic stress, ethanol stress, thermal stress and different fermentation inhibitor stresses. Furthermore, other very amenable traits that yeast strains should possess to be interesting for industrial applications include the ability to use cheap carbon sources like lignocellulosic hydrolysates, capability to sustain fermentation under very high gravity conditions, thermotolerance, osmotolerance, ethanol tolerance, acetic acid tolerance and furan derivative tolerance [16].

There are now a few non-conventional yeasts described as resistant to the stresses mentioned

above. This makes a good point for the argument that non-conventional yeasts may possess novel, unique resistance mechanisms which are not present in the model yeast; with the availability of next-generation sequencing technology and highly advanced molecular and genetic engineering tools, the molecular basis of these mechanisms may be uncovered.

Candida ethanolica was first isolated from industrial fodder yeast cultivated on synthetic ethanol as the only source of carbon. This new yeast was at that time considered different from other *Candida* species by not assimilating nitrate, not producing urease and not fermenting sugars. It was described as capable of growing at temperatures up to $42°$C, being its optimal growth temperature $28°$C [15]. The strain of *Candida ethanolica* used in this work, BJK_1A, was isolated from organic apple cider vinegar. Because the microbiota responsible for the aroma profile of the vinegar have to survive the harsh environment of the submerged bioreactor it was expected that this yeast showed the abilities of other non-conventional yeasts [28]. Accordingly to [28] this strain probably originated from the ethanol fermentation to vinegar broth and is resistant against 30g/L of acetic acid, comparable to *Zygosaccharomyces bailii* [23]. It has been identified in many places mainly being a spoilage yeast, mainly in wines or spirits like grappa, where it is considered that this yeast does not contribute to the fermentation process and it's presence can be prejudicial as it is able to use ethanol [14], or in sour rot-damaged grapes that are sources of wine spoilage yeasts [3]. In other environments, for example in the United States and Brazil, it has been found that this yeast can survive in farms in corn silage [26] [4]. Although *C. ethanolica* is spoiling the silage it's survival in this environment shows that this yeast can assimilate lactate and glucose and can survive the presence of acetic acid. Besides being a spoilage yeast in some cases, in others *C. ethanolica* can be really useful. For example, it has been identified in cocoa bean fermentations [29] [6] [7] [8]. It has been verified that this yeast possesses enzymes that improve the quality of the final product. This enzymes act on the degradation and solubilization of the pulp, allowing penetration into the fermenting mass allowing acetic acid bacteria to grow. Regarding the enzymes that improve the final product aroma it has been demonstrated that *C. ethanolica* has glicosidase and pectinase activity that allow the formation of volatile compounds that are known for their aromatic properties [29].

An alternative to landfill sites nowadays is the biological degradation of these plastics. The ester and urethane linkages present in polyester PUs are naturally vulnerable to enzymatic degradation. Polyether PUs contain ether linkages within the polymer backbone that are reported to be far more recalcitrant. Because of the vulnerability to microbial attack PUs can be degraded in large-scale commercial composting facilities. This method has proved to be useful in reducing landfill and meeting recycling goals, also being rapid and inexpensive. An important factor in composting is the temperature that can reach $70°$. Therefore, the fungal population involved in this kind of method can survive is composted of thermotolerant and thermophilic fungi. The following work intends to thoroughly explore the potential of one strain of *C. ethanolica*, BJK_1A, recovered from organic cider vinegar [28], addressing relevant phenotypic traits of the strain including its tolerance to high concentration of organic acids, its ability to use multiple carbon sources, among others. To complement this approach it was also performed a genomic analysis of the strain, followed by a manually curated annotation, this representing the first genomic sequence available for a strain of this species. a series of phenotypical tests and a genomic characterization of this yeast.

2. Material and Methods

2.1. Strains and growth media

A total of eight different yeast strains, comprising two strains of *Saccharomycodes ludwigii* BJK_5C and UTAD17, *Candida ethanolica* BJK_1A, two strains of *C. glabrata* UTAD68 and CBS138, *Issatchenkia orientalis* and *Hanseniaspora guilliermondii* UTAD222. All of the yeast strains were batch-cultured at 30°C, with orbital agitation of 250rpm in the growth media MMB (Minimal Growth Medium), YPD (Yeast Peptone Dextrose Medium) or YPF (Yeast Peptone Dextrose Fermentation Medium), depending on the application. MMB medium contains per litre 20g glucose (Merck), 1.7g yeast nitrogen base without amino acids (Difco) and 1.67g ammonium sulphate (PanReac). YPD contains per litre 20g glucose (Merck), 10g yeast extract (VWR Chemicals) and 20g peptone (Difco). YPF contains per litre 300g glucose (Merck) 10g yeast extract (VWR Chemicals) and 20g peptone (Difco). When different sugars were being tested glucose in the media was replaced for fructose, mannose, galactose, glycerol, xylose or arabinose reaching a concentration per litre of 20g of the used sugar. For *I. orientalis* SD108, 20mg/L of uracil were required since this strain is auxothrophic for this amino acid. All the media were sterilized by autoclaving for 15 minutes at $121°$C and 1atm.

2.2. Growth assessment in solid media

Growth of the strains in solid YPD or MMB media was assessed based on spot assays. For this, pre-inoculum was performed by cultivating the cells overnigt at 30°C, with orbital agitation of 250rpm, in 100mL shake flasks containing 25mL of medium. On the next day, an appropriate volume of the culture was inoculated in fresh medium yielding a cell suspension with an OD_{600nm} of 0.1. The cultures were then incubated at 30°C and 250rpm until exponential phase (OD_{600nm} over 0.4) and then transferred to eppendorfs to obtain a cell suspension having an OD_{600nm} of 0.05 in a total volume of 1mL. Two dilutions of this initial cell suspension (1:5 and 1:25) were made. Using the three cell suspensions, a drop of 4μL from each one was deposited on the solid medium and incubated at 30°C for 2 or 5 days, depending on growth inhibition.

2.3. Assessment of tolerance to organic acids at a low pH

To assess tolerance of the strains to different organic acids, spot assays were performed in solid MMB medium supplemented with inhibitory concentrations of the chosen organic acids. The pH of the stock solutions of the acids used to supplement the growth medium solutions was set at 4.5 in order to assure the acids were in a protonated form, facilitating diffusion through the cell wall. The medium's pH was also set to 4.5. The same experimental setup described above was used for the spot assays. The used organic acids and concentrations tested were the following: acetic acid: 60, 80, 100, 120, 140mM; propionic acid: 20, 25, 30, 40mM; levulinic acid: 50, 60, 70 80mM; Itaconic acid: 300, 400, 500, 600mM; succinic and gluconic acid acid: 1, 10, 100mM; acrylic acid: 1, 2.5, 5, 7.5, 10, 100mM; adipic acid: 1, 2.5, 5, 7.5, 10mM.

2.4. Growth curves in liquid medium

C. ethanolica BJK_1A and *S. cerevisiae* Y7221 were profiled for their growth rates and *C. ethanolica* BJK_1A was also profiled for its growth in the presence of acetic acid in liquid MMB medium. For both assays, a pre-inoculum of the strains was performed overnight in MMB medium at 30°C with orbital agitation of 250rpm in 100mL shake flasks containing 25mL of medium. On the next day, an appropriate volume of the culture was inoculated in fresh medium, containing appropriate concentrations of acetic acid for the corresponding assay, yielding a cell suspension with an OD_{600nm} of 0.1. For growth in the presence of acetic acid concentrations of 50, 100, 120, 140, 250, 300, 350 and 400mM were tested and the pH of the medium was set to 4. A culture in MMB medium at pH4 without acetic acid was used as control. These cultures were then incubated at 30°C and 250rpm until stationary phase was reached and OD_{600nm} was measured in appropriate time intervals. Using the values of OD_{600nm} from the exponential phase the specific growth rates were calculated for *C. ethanolica* BJK_1A and *S. cerevisiae* Y7221. For 50 and 100mM of acetic acid added to the medium of *C. ethanolica* BJK_1A cultures, the supernatant removed with every OD_{600nm} measurement was analysed with High Performance Liquid Chromatography (HPLC), separating 10μL of the supernatant in an Aminex HPX- 87H column (Biorad) eluted with 0.05% sulphuric acid at a flow rate of 0.6 mL/min. A UV detector set at 210 nm was used for detection of organic acids while detection of sugars, ethanol and glycerol was performed using an RI detector. Under the experimental conditions used the retention times for the following compounds were obtained: glucose - 9.3min, acetic acid - 14min.

2.5. Batch fermentations of *Candida ethanolica* BJK_1A under VHG conditions

The same experimental setup described above was used for this assay although with YPD medium for the pre-inoculum and YPF medium for the inoculum and the fermentation lasted seven days. During this period 1mL of the culture supernatant was collected and OD_{60nm} measured every 24h.The accumulation of metabolites of interest in the culture broth was measured by HPLC separating 10μL of the supernatant in an Aminex HPX-87H column (Biorad) eluted with 0.05% sulphuric acid at a flow rate of 0.6 mL/min. A UV detector set at 210 nm was used for detection of organic acids while detection of sugars, ethanol and glycerol was performed using an RI detector. Under the experimental conditions used the retention times for the following compounds were obtained: glucose - 9.3min, ethanol - 21min.

2.6. Growth of *Candida ethanolica* BJK_1A at different temperatures and under anaerobiotic conditions

In order to assess the ability to grow under different temperatures (over 30°C) the strain of *C. ethanolica* was cultured in solid YPD agar plates and kept at 37°C and at 40°C for two days. To assess the ability of *Candida ethanolica* BJK_1A to grow under anaerobiotic conditions this strain and *S. cerevisiae* Y7221 were cultured on agar plates containing YPD medium and kept in a closed container with an anaerobisis membrane. The container with the plates was maintained at $30°C$ for two days.

2.7. Growth of *Candida ethanolica* BJK_1A on lignocellulose hydrolysates

To assess the ability of *C. ethanolica* BJK_1A to grow on lignocellulose hydrolysates as carbon sources it was performed a growth using wheat

straw, oak sawdust and MMB medium as a control. *C. ethanolica* BJK_1A was cultured using the same setup as above. The inoculum was made in an appropriate volume of fresh hydrolysate medium, in equal volumes of the hydrolysate and MMB medium, and in fresh MMB medium, yielding a cell suspension with an OD_{60nm} of 0.1. This culture was incubated at 30°C and with an orbital agitation of 250rpm during 2 days. During this period 0.5mL of the culture supernatant was collected in appropriated time intervals for HPLC analysis and the culture viability was determined by the CFU method. Accumulation of metabolites of interest and glucose consumption in the culture broth was measured by HPLC separating 10μL of the supernatant in an Aminex HPX- 87H column (Biorad) eluted with 0.05% sulphuric acid at a flow rate of 0.6 mL/min. A UV detector set at 210 nm was used for detection of organic acids while detection of sugars, ethanol and glycerol was performed using an RI detector. Under the experimental conditions used the retention times for the following compounds were obtained: glucose - 9.3min.

2.8. Genomic DNA extraction

C. ethanolica BJK_1A cells were grown in YPD until an OD_{600nm} above 3.0. Afterwards cells centrifuged at 5000rpm, during 5min at 4°C and the supernatant discarded. Pellet was resuspended in 1ml Sorbitol 1M (Sigma) and EDTA (tetrasodium salt dehydrate, Sigma-Aldrich) 0.1M at pH 7.5 solution and transferred to an eppendorf. Afterwards, 10 mg/mL zymolease (Zymo research) was added and the solution was incubated at 37°C until protoplast formation. The spheroplasts were centrifuged at 5000rpm for 5min, and the pellet resuspended in 1mL Tris-HCL with pH 7.4 50 mM (Sigma-Aldrich), and EDTA 20 mM solution. After this step, 30μL of SDS 10% was added to the mixture. After an incubation step of 30min at 65°C, 250μL of Potassium Acetate (5M, Merck) was added to induce protein precipitation, this being followed by a 1h incubation on ice. Afterwards the solution was centrifuged at 10000rpm for 10min and the supernatant transferred to 2 new eppendorfs. 1 volume of cold isopropanol was used to wash the pellet followed by centrifugation at 5000rpm for 15min. Supernatant was discarded and the resulting pellet was incubated in 1mL ethanol 70% during 5min, and wash with ethanol 70% twice. The pellet was dried in speed vacuum and resuspended in 200μL TE (pH 7.4). The final step, addition of 0.5μl of RNase (10 mg/ml) followed by 1h incubation at 37°C followed. Mixture was centrifuged at 10000rpm during 15min and the supernatant was preserved at 4°C till further use.

2.9. Genome Sequencing and Annotation

The genome of *Candida ethanolica* BJK_1A isolate was obtained at the next-generation sequencing (NGS) laboratory of Stab Vida, using Illumina HiSeq sequencing technology. Two rounds of deep sequencing were performed and the reads obtained were analysed using the software CLC Genomics Workbench. The reads obtained in the sequencing step were trimmed based on quality by the laboratory so no trimming step was needed before the assembly step. The trimmed reads were assembled using *de novo* assembly and mapped against the reference genome of *Candida ethanolica* M2 strain.The annotation was performed using the algorithm GeneMark 27 and analysed using the Pedant-pro 3.2.52 to allow comparative feature analysis.

3. Results

4. Ability of a selected set of Non-*Saccharomyces* species to assimilate different carbon sources

This work was focused on the exploration of several strains of different Non-*Saccharomyces* species including *S'codes ludwugii* BJK_5C and UTAD17, *C. ethanolica*, *C. glabrata* UTAD68 and CBS138, *S. cerevisiae* Y7221, *H. guilliermondii* UTAD222 and *I. orientalis* SD108. The first phenotypic trait that was studied in the strains was their ability to grow in the presence of various carbon sources including glucose, fructose, mannose, galactose, glycerol, xylose and arabinose. This phenotypic assessment was performed using spot assays using mineral (MM) or rich (YP) basal medium. Growth was evaluated after 2-5 days of growth, depending on the severity of inhibition observed.

As a result from the assimilation of different carbon sources the strains of *C. ethanolica* and *I. orientalis* were the ones that could use a large number of sugars. Arabinose was the sugar that less yeast strains were able to use, only being observed a robust growth for *I. orientalis* SD108 and *C. ethanolica* BJK_1A.

5. Resistance to organic acids

Resilience of the different yeast species to stress imposed by organic acids was also tested since this is a relevant trait in multiple biotechnological applications. As such, it was tested resilience of the strains to inhibitory concentrations of acetic and levulinic acids since these two acids are known to be present in lignocellulosic hydrolysates. It was also tested resilience to succinic, itaconic, gluconic and acrylic acids since one of the possibilities of exploration of non-conventional yeasts is their use in production of these add-value acids. Necessarily, more tolerant strains will be more interesting as hosts.

For adipic, gluconic and levulinic acids all the strains showed ability to grow in all the tested concentrations of these acids, probably showing that the range tested was not inhibitory enough. For acetic acid, *C. ethanolica* BJK_1A and *I. orientalis* SD108 were the most resistant strains, being able to grow in all tested concentrations. Both of *S'codes ludwigii* strains were also able to survive all the concentrations, but showing a less robust growth for 120 and 140mM. All of the other strains tested showed little tolerance to this acid. When cultured with different concentrations of succinic acid all strains but the *S'codes ludwigii* ones, which failed to grow in 100mM of this acid, were able to grow. Only *I. orientalis* SD108 was able to thrive in all concentrations of acrylic acid. The less resistant strains to itaconic acid were *C. ethanolica* BJK_1A and both strains of *C. glabrata*. All of the other strains were able to grow in the presence of this acid. Against propionic acid-induced stress, *C. ethanolica* BJK_1A and *I. orientalis* SD108 were shown to be the more tolerant strains.

5.1. Phenotypic characterization *Candida ethanolica* BJK_1A

According to [15], *C. ethanolica* ranges from ellipsoidal to elongated shape, single or in chains. When plate cultured it was characterized by marginal dendroidal pseudomycelium. In order to assess the morphology of *C. ethanolica* BJK_1A's cells, these were cultured in agar plates containing YPD medium and observed under an optical light microscope confirming the expected shape of *C. ethanolica*. It showed an ellipsoidal yeast that could form pseudohyphal networks, also visible in the images gathered with the higher magnification. An interesting feature that emerged from the microscopic observation of the cells, was the observation of what appeared to be dark structures located inside the cells. To verify if this dark spots could be multiple nucleus or even bacteria living within the cells in a symbiotic relation, something that had been described before **??**, the cells were stained with 4',6-diamidino-2-phenylindole (DAPI), a fluorescent stain that binds strongly to adenine-thymine rich regions in DNA. Results showed that the dark spherical bodies were not visible using fluorescence microscopy, therefore excluding the possibility of this representing live bacteria or multiple nuclei. The most probable hypothesis is that these structures represent lipid bodies since *C. ethanolica* was previously described to harbor such structures, as referred by [30] and [11].

In order to assess the specific growth rate (μ) of *C. ethanolica* BJK_1A, this strain was cultured in MMB medium and the OD_{600nm} was measured every 2 hours for 8 hours and then at 24 and 26 hours to confirm that the cells had reached the stationary phase.

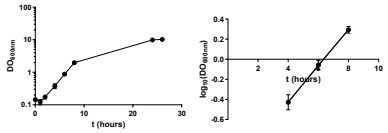

Figure 1: Left - Evolution of $\log_{10} DO_{600nm}$ from *C. ethanolica* BJK_1A cultured on MMB medium with an initial OD_{600nm} of 0.1. Right - Points from from the exponential phase used to calculate the specific growth rate. The equation of the linear regression applied to those points is $y = 0.1804x - 1.146$ with a R^2 of 0.99. The slope of the regression corresponds to the specific growth rate, $\mu = 0.1804h^{-1}$.

From figure 1 it was possible to observe the growth curve of *C. ethanolica* BJK_1A It can grow at a higher rate than *S. cerevisiae* Y7221, $0.1804h^{-1}$ comparing with $0.1218h^{-1}$. The maximum OD_{600nm} for *C. ethanolica* BJK_1A was 10.26 at 26 hours of growth and *S. cerevisiae* Y7221 only reached an OD_{600nm} of 8.52 with the same amount of hours of growth. This result shows that under the experimental conditions used, *C. ethanolica* BJK_1A can grow until up to 20% more than *S. cerevisiae* Y7221.

The results obtained in the phenotypic screening carried out demonstrated the high capability of *C. ethanolica* BJK_1A to tolerate acetic acid stress. Since this trait was previously linked to the ability of yeast cells to co-consume acetate even when glucose is present in the growth medium **??**, this hypothesis was assessed for *C. ethanolica* BJK_1A. For this, the cells were cultivated in MM medium supplemented with increasing amounts of acetic acid (up to 400 mM) and the variations in the concentrations of glucose and acetic acid in the culture broth were accompanied by HPLC.

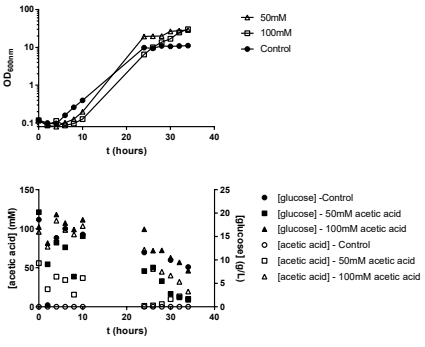

Figure 2: (Left) Evolution OD_{600nm} from *C. ethanolica* BJK_1A cultured on MMB medium at pH4 with an initial OD_{600nm} of 0.1 and varying concentrations of acetic acid. The concentrations of acetic acid used were 50 and 100mM and a culture of *C. ethanolica* BJK_1A without acetic acid was used as control. (Right) Evolution of glucose and acetic acid concentrations obtained from HPLC analysis of supernatant from a culture of *C. ethanolica* BJK_1A on MMB medium with an initial acetic acid concentration of 50mM, 100mM and a control.

Figure 2 illustrates the evolution of the concentration of acetic acid and glucose in the broth of the cultures cultivated in MMB medium supplemented with 50 or 100 mM acetic acid. It is clear that the concentration of both compounds show a similar consumption profile, being consumed almost totally during the time that the experiment took place. It thus seems that *C. ethanolica* BJK_1A is able to use both acetic acid and glucose as carbon sources. In fact, this can explain the observation that cultures incubated in the presence of acetic acid reached higher OD_{600nm} than the culture incubated in the absence of the acid.

According with [15] this *Candida* species differed from other by its lack of capability to ferment sugars and according to [11] this species is able to ferment in addition to glucose, fructose and sucrose. Therefore the next logical step for the physiological characterization of *C. ethanolica* BJK_1A was to evaluate the ability to grow under anaerobic conditions. To do that this strain was cultured on agar plates containing solid YPD medium and kept in a closed container with an anaerobiosis membrane at $30°C$ for two days. As a control *S. cerevisiae* Y7221, was cultivated under the same conditions. As a result of this test it was observed that both species tested showed growth meaning that *C. ethanolica* BJK_1A is able to ferment glucose achieving more biomass than *S. cerevisiae* Y7221 under the same conditions, being both able to grow under anaerobic conditions.

Besides tolerance to inhibitors, tolerance to high temperatures is also a relevant phenotypic trait for production of bioethanol because high temperature conditions allow enzymatic saccharification of the biomass prior to fermentation [24] [25]. As such, growth of *C. ethanolica* BJK_1A was also profiled in MMB medium at a different range of temperatures, as high as 37 and 40°C. It was to observe growth at all temperatures tested, proving *C. ethanolica* BJK_1A's thermotolerance and possible suitability to be used in biotechnological applications such as bioethanol production.

6. Growth of *C. ethanolica* BJK_1A in lignocellulose hydrolysates

The overall results of the phenotypic profiling of *C. ethanolica* BJK_1A described above indicates that this strain could be highly interesting for bioethanol production, specially using hydrolysates as substrates. In that context, in this part of the work it was assessed how *C. ethanolica* BJK_1A grow in two hydrolysates obtained from wheat straw or oak sawdust. The hydrolysate coming from wheat straw had a concentration of 3.046g/L of acetic acid and 16.606g/L of glucose and the one coming from oak sawdust had a higher concentration of acetic acid and glucose, 40,40 and 51,70g/L respectively. Besides cultivating the strains in the hydrolysate itself, it was also assessed the effect of cultivating the cells in the hydrolysate supplemented with minimal medium. This supplementation was performed based on the hypothesis that the nutritional content of the hydrolysate could not be high enough to sustain robust growth of the cells.

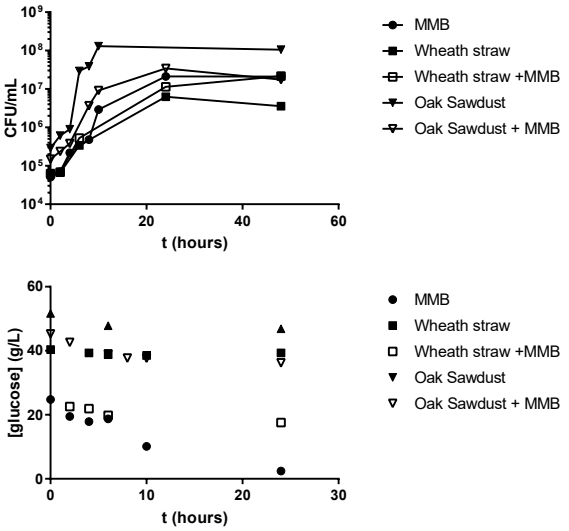

Figure 3: (Top) Evolution of CFU/mL from *C. ethanolica* BJK_1A cultured in lignocellulose hydrolysates with an initial OD_{600nm} of 0.1. (Bottom) Evolution of glucose concentration from *C. ethanolica* BJK_1A cultured in lignocellulose hydrolysates.

The results shown in figure 3 led to conclude that *C. ethanolica* BJK_1A was able to grow to a high number of CFU using oak sawdust. This hysrolysate had the highest concentration of glucose as well as the highest concentration of acetic acid. This result contributes to the conclusion that this strain of *C. ethanolica* BJK_1A is tolerant to high concentrations of acetic acid. The less efficient medium was wheat straw, confirming the importance of the glucose concentration. This hydrolysate had a lower glucose concentration than MMB and also has a lower acetic acid concentration than oak sawdust, being expected that the number of CFU obtained using this medium was lower than when oak sawdust or even MMB was used. When wheat straw was complemented with an equal volume of MMB its initial glucose concentration was similar to the concentration of MMB medium. The close glucose concentration of these two media make the close results expectable, using only MMB or combining equal volumes of MMB and wheat straw lead to approximately the same level of growth.

7. Very high gravity fermentation of *C. ethanolica* BJK_1A

In order to assess capability to grow under osmotic stress, *C. ethanolica* BJK_1A was cultured in YPF medium, containing $300g/L$ of glucose and aliquots were taken from the culture every 24 hours to measure OD_{600nm} and to collect supernatant for the HPLC analysis. The evolution of OD_{600nm} as well as the evolution of glucose and ethanol concentrations are represented in figure 4.

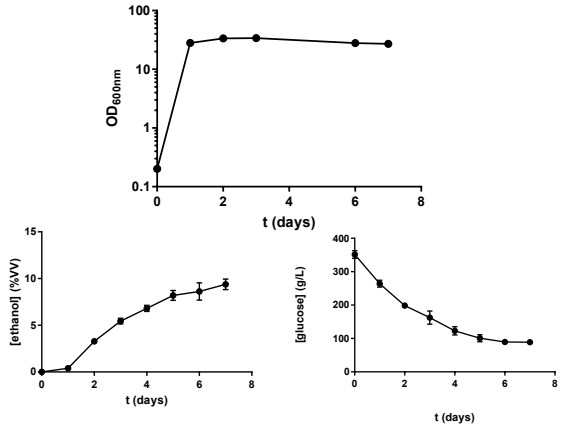

Figure 4: (Top) Evolution of $\log_{10} DO_{600nm}$ from *C. ethanolica* BJK_1A cultured on YPF medium with an initial OD_{600nm} of 0.1. (Bottom right) Concentration of glucose (g/L) on the culture supernatant over the 7 days of fermentation. (Bottom left) Concentration of ethanol (%VV) on the culture supernatant over the 7 days fermentation.

This result showed that *C. ethanolica* BJK_1A is mildly tolerant to osmotic stress caused by presence of high concentrations of glucose in the medium, being able to produce about 10% of ethanol (v/v). Nonetheless, it can resist a 300g/L glucose concentration, being able to grow in YPF medium for at least seven days with an average ethanol production of 9.38%(VV). The fact that the glucose concentration stabilized for the last days of the fermentation at an average concentration of 88.9g/L when ethanol concentration was at the highest suggests that a high ethanol concentration can in fact be inhibitor to *C. ethanolica* BJK_1A's growth. The ability to produce ethanol also corroborates the ability of *C. ethanolica* BJK_1A to ferment glucose.

8. Genome structure of *Candida ethanolica* BJK_1A and comparative analysis with the genome of *Candida ethanolica* M2

Considering the interesting phenotypic properties exhibited by the *C. ethanolica* BJK_1A strain it was decided to proceed with genome sequencing of this strain since up to now only 1 Bioproject focused on *C. ethanolica* (strain M2) was available at NCBI [1] and this did not provide information concerning annotation and eventual gene/genomic analysis. As such, the genome of *C. ethanolica* BJK_1A was obtained using two independent rounds of paired end HiSeq Illumina sequencing which resulted in ca 102 million reads. Assembly of the resulting reads was performed using the *de novo* assembly tool of CLC Genomic Workbench resulting in 4,974 contigs, yielding a total of 85,158,017 assembled bases.

The M2 strain of *C. ethanolica* used as reference was isolated from rotten pineapple and started being studied for bioattraction of pests. The predicted genome for M2 strain is 23.58Mbp with an average GC content of 26.4% [1]. The GC content and size of the genome was significantly different between the two strains. In order to perform an initial comparison between both genomes an alignment using Progressive Mauve was performed. It was visible that a large part of the sequenced genome of strain BJK_1A had no correspondence with the M2 genome. Using the move contigs tool of Mauve it was possible to reorder the contigs to get a better alignment.

To perform the annotation of the contigs of our strain it was decided to first start with the annotation of the genome that was available for the M2 strain. This choice was supported by the idea that having more than one annotated genome of *C. ethanolica* strains available would end up rendering the final result much more reliable. Furthermore, the sequencing of the genome of the BJK_1A strain was also quite laborious and time-consuming due to problems with the company that made the sequencing. Thus, gene prediction was first performed using the genome of the M2 strain as a reference and then these gene models were mapped in the contigs obtained for the BJK_1A strain. The ab initio gene detection that was performed for the M2 strain was manually refined resulting in a predicted ORFeome for the *C. ethanolica* M2 strain estimated in 10,504 protein-encoding genes. The average protein length ranges between 10 proteins having a size of over 3000 amino acids and 5 proteins having a size smaller than 50 amino acids, being the remaining 10,489 of a size between 3000 and 50 amino acids. The gene structure analysis predicted that 86.4% of the genes (9080) have 1 exon, 12.4% (1304) have 2 exons, 1% (105) have 3 exons, and 15 genes have 4 or more exons.

The predicted ORFeome of *C. ethanolica* BJK_1A had initially 5,422 protein-encoding genes. It was still possible to observe a considerable region of the BJK_1A genome that had no match in the M2 genome. This result was expected because the strains come from different environments and could mean that BJK_1A have genes that are not present in the M2 genome. A protein BLAST using

the predicted proteins for both strains was made, using CLC Genomics Workbench. All the protein-encoding genes predicted by the automatic annotation were used and the protein-encoding genes from the annotated M2 genome were used as a reference sequence. From this BLAST resulted that of the 5,422 protein-encoding genes predicted for BJK_1A 1,154 were aligned with a resulting e-value superior to 1×10^{-20}. These genes from BJK_1A were verified and for 1,104 genes it was discovered that they were the result of sequencing bacterial contaminations. It is unclear the origin of this contamination although after several attempts we could confirm that this is not present in our DNA sample suggesting that it may result from a contamination in the DNA sequencing facility setting.

An initial manually refinement of the automatic annotation allowed the removal of the genes that were a result of contamination during the sequencing step. Therefore, as a result of the automatic annotation the predicted ORFeome of *C. ethanolica* BJK_1A has 4,278 protein-encoding genes. The average protein length ranges between 6 proteins having a size of over 3000 amino acids and 20 proteins having a size smaller than 50 amino acids, being the remaining 4,252 of a size between 3000 and 50 amino acids. The gene structure analysis predicted that 64.9% of the genes (2776) have 1 exon, 20.7% (886) have 2 exons, 8.3% (356) have 3 exons, and 260 genes have 4 or more exons.

Table 1: General features of the *C. ethanolica* BJK_1A genome obtained after the manually curated annotation.

	Annotation statistics	
	BJK_1A	M2
Total number of contigs	4,974	2,193
Total number of contigs with genes	1,815	1982
Maximum contig length (bp)	104,226	88,862
Minimum contig length (bp)	1,000	1,000
Average contig size (bp)	3,321.1	10,609.3
Assembly size (bp)	16,519,089	23,266,223
Average GC content (%)	35.0	26.4
Total number of CDS	4,278	10,504
mRNAs	7	4
Genes with 1 exon	2776	9,080
Genes with 2 exons	886	1304
Genes with more than 2 exons	616	120

9. Discussion

The first part of this work consisted in a screening of phenotypic characteristics applied to a small cohort of non-conventional yeast strains belonging to the *S'codes ludwigii*, *C. ethanolica*, *C. glabrata*, *S. cerevisiae*, *H. guilliermondii* and *I. orientalis* species. Firstly a group of sugars were tested to understand which ones could be assimilated and used by the strains. For glucose, fructose and mannose all of the strains showed assimilation capability. According with the literature, all of the species were able to grow on glucose [18]. For fructose assimilation it was also expected to observe growth capability for all strains [11][21][9][31][12][19]. The more relevant differences observed concerning this matter were observed in metabolization of galactose, which was not observed to occur for *H. guilliermondi* UTAD222 and *S'codes ludwigii* UTAD17. Differences between strains started to be noticed with galactose assimilation. Differences concerning the utilization of arabinose and xylose was also observed, with all of the strains but *S'codes ludwigii* UTAD17 showing assimilation capability, *H. guilliermondii* UTAD222 and *S'codes ludwigii* UTAD17 not being able to assimilate xylose and only *C. ethanolica* BJK_1A, *I. orientalis* SD108 and *C. glabrata* UTAD68 being the sole strains able to grow using arabinose. On YPD agar plates suplemented with xylose all of the species were able to grow, meaning that *S'codes ludwigii* UTAD17 and *H. guilliermondii* UTAD222 were using other compounds present in the rich and non-defined YPD medium as carbon source. This behaviour was also noticed for arabinose where all but the *S'codes ludwigii* strains were able to grow. Regarding glycerol, all but *H. guilliermondii* UTAD222 were able to assimilate this carbon source. Although it is not a fermentable sugar, glycerol is a by-product of biodiesel production that can be used as a carbon source for oleaginous yeast, promoting lipid production, a product that is only economically feasible is it can be produced from a cheap substrate like glycerol [22] [5] [17]. Therefore, in order to be used for single cell oils production, it is important that a yeast can assimilate this carbohydrate. As a result of the assimilation test, the species that stood out were *C. ethanolica* BJK_1A and *I. orientalis* SD108 because they were the ones that could assimilate all of the sugars tested. Resistance to organic acids is another desirable trait in the context of industrial biotechnology, being relevant for fermentation of hydrolysates but also if one is interested in the production of these molecules which are themselves recognized add-value compounds [2]. Under the conditions that we used, the concentrations of levulinic acid used (in the range of 50-80mM) did not significantly inhibited growth of any of the strains tested, including of the *S. cerevisiae* strain which was used as a control. Regarding acetic acid, concentrations of 60, 80, 100, 120 and 140mM were tested and only *C. ethanolica* BJK_1A, *I. orientalis* SD108 and both strains of *S'codes ludwigii* could sustain growth in such high concentration , in the range of those that are found in lignocellulosic hydrolysates. The strains of *S'codes ludwigii* were those exhibiting higher resilience being able to tolerate the higher concentration of this acid, 100mM. All of the other species could stand concentrations of 1, 10 and 100mM of succinic acid. Acrylic, gluconic, itaconic, propionic

and adipic acids are also recognized commodity chemicals and therefore tolerance of the strains to different concentrations of these organic acids was also assessed. The results obtained showed that *I. orientalis* was the most tolerant species to all acids tested, with *C. ethanolica* BJK_1A ranking as the second more tolerant strain being able to grow in the presence of 2.5mM acrylic acid, 10mM adipic acid and 100mM gluconic acid.

Considering the good results obtained in the phenotypic screening undertaken it was decided to take a closer look into the *C. ethanolica* BJK_1A strain. In specific, a more thorough physiological analysis was performed aiming to better characterize this species and this strain in particular. The first observation was the morphological observation of *C. ethanolica* BJK_1A cells, this showing that these are ellipsoidal-shapped cells with ability to form pseudohyphae. This kind of organization is used to better decompose substrates, being also associated with species potentially pathogenic for plants and humans [13]. Under a 1000x magnification it was possible to see dark bodies within the cells. DAPI staining and observation with fluorescence microscopy made possible to conclude that those bodies were not nucleic acids, leaving the hypothesis of being lipid bodies. [30] and [11] referred *C. ethanolica* as a yeast capable of producing and accumulate lipids. This hypothesis should be further confirmed in order to assess in *C. ethanolica* BJK_1A can in fact produce lipids that can be used in microbiological produced biodiesel. The extreme phenotype of resistance to acetic acid exhibited in the phenotypic screening performed was further confirmed, the *C. ethanolica* BJK_1A cells being able to grow even in the presence of 400mM acetic acid. HPLC analysis of the supernatant collected over time showed that the acetic acid concentration in the supernatant was reduced to almost zero. It was referred by [28] that this strain of *C. ethanolica* was able to resist concentrations of acetic acid up to 30g/L and these results could provide an explanation for this. These results suggest that *C. ethanolica* BJK_1A is able to use acetate as the carbon source, even when glucose is present in the medium.

The fact that *C. ethanolica* BJK_1A exhibited several traits that are relevant in the context of ethanol production from lignocellulosics prompted us to examine how this strain would behave in two real hydrolysates obtained from wheat straw and oak sawdust. These hydrolysates had different compositions in sugar and also in inhibitors, specially in acetic acid. Results showed that *C. ethanolica* BJK_1A was able to grow using both hydrolysates even when not supplemented with MMB medium. This result complemented the one from the growth with acetic acid in the medium, a higher concentration of this compound led to higher biomass amount. Under very high gravity conditions *C. ethanolica* BJK_1A was able to grow on 300g/L of glucose producing approximately 10%(vv) of ethanol. Although this result shows that *C. ethanolica* BJK_1A was able to survive in VHG conditions it was also elucidative about the toxicity of ethanol to the cells. From the final glucose concentration on the medium it was possible to conclude that the corresponding ethanol levels were sufficiently inhibitory to stop biomass production.

Genomic analysis of *C. ethanolica* BJK_1A was the final step of this work. The sequencing step of this genomic analysis was made by STAB Vida using two rounds of paired end HiSeq Illumina sequencing. The resulting reads were assembled using CLC Genomic Workbench resulting in 4,974 contigs with a total genome size of ca 16Mbp and an average GC content of 35%. A list of 2,193 contigs from *C. ethanolica* M2 that were already publicly available at the NCBI database was annotated and used as a reference to annotate BJK_1A. Using Progressive Mauve it was possible to see that BJK_1A had a large region that did not align with the contigs of M2. This was either due to genes that were only present in BJK_1A but not in M2 or contigs coming from contaminations of the sequencing step. The annotation of BJK_1A using the M2 as reference allowed to understand, comparing protein sequences from both strains, that 1,154 protein-encoding genes had a e-value higher than 1×10^{-20}. After verifying said genes it was possible to conclude that they were in fact the result of contamination from the sequencing step. After removing the wrongly sequenced genes and using Progressive Mauve it was possible to observe that the non-aligning region had decreased considerably. The result of the annotation for BJK_1A after the initial curation predicted 4,278 genes, with a genome size of ca 16Mbp and a GC content of 35.0%. This annotation resulted in a smaller number of genes than those predicted for M2, 10,504. This fact could be explained be the fact that M2 is probably an hybrid, having repeated contigs that were not assembled in a single contig but in two with small differences. The KEGG Koala blast tool allowed to use the annotated genes from BJK_1A to reconstruct the metabolic pathways resulting in 187 genes that encode enzymes involved in carbohydrate metabolism, including glycolysis and gluconeogenesis, TCA cycle, pentose phosphate pathway, fructose, mannose, galactose, starch and sucrose metabolism, among others. The presence of enzymes involved with the metabolism of said sugars comes in line with the sugars assimilation test, confirming the strain's ability to grow on glucose,

fructose, mannose and galactose and supports the tests in [11] that conclude that *C. ethanolica* can assimilate sucrose, confirming that *C. ethanolica* can use a large range of carbon sources. The presence of genes used in the metabolism of a large range of carbon sources sustained the initial tests of carbon assimilation that were used to chose *C. ethanolica* BJK_1A from the initial group of yeasts. It also allowed to predict that this strain has 94 genes associated with lipid metabolism, mainly biosynthesis of triacylglycerols usually associated with oleaginous yeasts [27], placing *C. ethanolica* BJK_1A as a potential lipid producing yeast consisting with the results shown by [30] and [11].

The physiological and genomic analysis of *C. ethanolica* BJK_1A contributed to a better knowledge about this strain. It was possible to compile useful information about carbon sources that this strain can use and also about concentrations of organic acids that it can stand. Being able to grow in lignocellulosic hydrolysates and under VHG conditions is also an advantage for this strain, not only for bioethanol production but also for the production of lipids. The genome annotation allowed to understand which genes are present in *C. ethanolica* BJK_1A and as for future wok it would be interesting to complete the validation of the predicted gene models of *C. ethanolica* BJK_1A. This analysis could provide information about genes present that are responsible for the production of add-value products but also which ones can be engineered to form a better performing strain. Having been identified as a possible oleaginous yeast, the production of lipids from this strain of *C. ethanolica* is also an interesting feature to be studied.

References

[1] [candida] ethanolica m2 strain: Genome sequencing and assembly. Accessed: 2018-06-24.

[2] R. A. J. ao, M. Tobias, P. Anneli, H. Barbel, L. Gunnar, and G. M. F. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by saccharomyces cerevisiae. *Journal of Chemical Technology & Biotechnology*, 82(4):340–349.

[3] A. Barata, S. González, M. Malfeito-Ferreira, A. Querol, and V. Loureiro. Sour rot-damaged grapes are sources of wine spoilage yeasts. *FEMS Yeast Research*, 8(7):1008–1017, 2008.

[4] C. B.F., A. C.L.S., K. P.M., B. L.R., P. M.N., and S. R.F. Occurrence of mycotoxins and yeasts and moulds identification in corn silages in tropical climate. *Journal of Applied Microbiology*, 120(5):1181–1192.

[5] G. P. da Silva, M. Mack, and J. Contiero. Glycerol: A promising and abundant carbon source for industrial microbiology. *Biotechnology Advances*, 27(1):30 – 39, 2009.

[6] I. M. da Veiga Moreira, M. G. da Cruz Pedrozo Miguel, W. F. Duarte, D. R. Dias, and R. F. Schwan. Microbial succession and the dynamics of metabolites and sugars during the fermentation of three different cocoa (theobroma cacao l.) hybrids. *Food Research International*, 54(1):9 – 17, 2013.

[7] H.-M. Daniel, G. Vrancken, J. F. Takrama, N. Camu, P. De Vos, and L. De Vuyst. Yeast diversity of ghanaian cocoa bean heap fermentations. *FEMS Yeast Research*, 9(5):774–783, 2009.

[8] G. V. de Melo Pereira, V. T. Soccol, and C. R. Soccol. Current state of research on cocoa and coffee fermentations. *Current Opinion in Food Science*, 7:50 – 57, 2016. Food chemistry and biochemistry Food bioprocessing.

[9] R. Estrada-Martínez, N. Pacheco López, T. González, A. Sanchez, J. Morales, and I. Rodriguez Buenfil. Sugar assimilation profile of candida glabrata for alcohol production in synthetic medium. 06 2013.

[10] J. F T Spencer, A. L Ragout de Spencer, and C. Laluce. Non-conventional yeasts. 58:147–56, 03 2002.

[11] A. Fatma, A. Samia, T. I. Eva, A. George, J. Atef, N. Moncef, and M. Tahar. Newly isolated yeasts from tunisian microhabitats: Lipid accumulation and fatty acid composition. *Engineering in Life Sciences*, 17(3):226–236.

[12] C. Guillaume, P. Delobel, J.-M. Sablayrolles, and B. Blondin. Molecular basis of fructose utilization by the wine yeast saccharomyces cerevisiae: a mutated hxt3 allele enhances fructose fermentation. *Appl Environ Microbiol*, 73(8):2432–2439, Apr 2007. 2269-06[PII].

[13] S. Hohmann. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev*, 66(2):300–372, Jun 2002. 0007[PII].

[14] L. Iacumin, M. Manzano, F. Cecchini, S. Orlic, R. Zironi, and G. Comi. Influence of specific fermentation conditions on natural microflora of pomace in “grappa” production. *World Journal of Microbiology and Biotechnology*, 28(4):1747–1759, Apr 2012.

[15] R. Johanna, Štros F., and K. Anna. Candida ethanolica n. sp. *Zeitschrift fur allgemeine Mikrobiologie*, 20(9):579–581.

[16] T. K. Sugar and salt tolerant yeasts. *Journal of Applied Bacteriology*, 74(2):101–110.

[17] S. Kitcha and B. Cheirsilp. Screening of oleaginous yeasts and optimization for lipid production using crude glycerol as a carbon source. *Energy Procedia*, 9:274 – 282, 2011. 9th Eco-Energy and Materials Science and Engineering Symposium.

[18] C. Kurtzman, J. Fell, and T. Boekhout. *The Yeasts: A Taxonomic Study*. Elsevier Science, 2011.

[19] D. J.-h. LIU Yang, LI Hong. Non-alcoholic beer production by saccharomycodes ludwigii. *FOOD SCIENCE*, page 186, 0.

[20] A.-K. Lobs, C. Schwartz, and I. Wheeldon. Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. *Synthetic and Systems Biotechnology*, 2(3):198 – 207, 2017.

[21] C. M. and F. F. Selective sugar consumption by apiculate yeasts. *Letters in Applied Microbiology*, 28(3):203–206.

[22] P. A. E. P. Meesters, G. N. M. Huijberts, and G. Eggink. High-cell-density cultivation of the lipid accumulating yeast cryptococcus curvatus using glycerol as a carbon source. *Applied Microbiology and Biotechnology*, 45(5):575–579, Jun 1996.

[23] N. P. Mira, M. Münsterkötter, F. Dias-Valada, J. Santos, M. Palma, F. C. Roque, J. F. Guerreiro, F. Rodrigues, M. J. Sousa, C. Leão, U. Güldener, and I. Sá-Correia. The genome sequence of the highly acetic acid-tolerant zygosaccharomyces bailii-derived interspecies hybrid strain isa1307, isolated from a sparkling wine plant. *DNA Res*, 21(3):299–313, Jun 2014. dst058[PII].

[24] S. Nonklang, B. M. A. Abdel-Banat, K. Chaaim, N. Moonjai, H. Hoshida, S. Limtong, M. Yamada, and R. Akada. High-temperature ethanol fermentation and transformation with linear dna in the thermotolerant yeast kluyveromyces marxianus dmku3-1042. *Appl Environ Microbiol*, 74(24):7514–7521, Dec 2008. 1854-08[PII].

[25] D. Radecka, V. Mukherjee, R. Q. Mateo, M. Stojiljkovic, M. R. Foulquié-Moreno, and J. M. Thevelein. Looking beyond saccharomyces: the potential of non-conventional yeast species for desirable traits in bioethanol fermentation. *FEMS Yeast Research*, 15(6):fov053, 2015.

[26] M. Santos, C. Golt, R. Joerger, G. Mechor, G. B. M. ao, and L. Kung. Identification of the major yeasts isolated from high moisture corn and corn silages in the united states using genetic and biochemical methods. *Journal of Dairy Science*, 100(2):1151 – 1160, 2017.

[27] P. Seraphim and A. George. Lipids of oleaginous yeasts. part i: Biochemistry of single cell oil production. *European Journal of Lipid Science and Technology*, 113(8):1031–1051.

[28] A. Stornik, B. Skok, and J. Trcek. Comparison of cultivable acetic acid bacterial microbiota in organic and conventional apple cider vinegar. *Food Technol Biotechnol*, 54(1):113–119, Mar 2016. FTB-54-113[PII].

[29] S. Visintin, V. Alessandria, A. Valente, P. Dolci, and L. Cocolin. Molecular identification and physiological characterization of yeasts, lactic acid bacteria and acetic acid bacteria isolated from heap and box cocoa bean fermentations in west africa. *International Journal of Food Microbiology*, 216:69 – 78, 2016.

[30] G. Wang, L. Liu, and W. Liang. Single cell oil production from hydrolysates of inulin by a newly isolated yeast *Papiliotrema laurentii* am113 for biodiesel making. *Applied Biochemistry and Biotechnology*, 184(1):168–181, Jan 2018.

[31] H. Xiao, Z. Shao, Y. Jiang, S. Dole, and H. Zhao. Exploiting *Issatchenkia Orientalis* sd108 as a new platform organism for organic acids production. 11 2013.