THE UNIVERSITY *of* EDINBURGH

THE UNIVERSITY *of* EDINBURGH
School of Engineering

PRODUCTION OF NOVEL TAXANE INTERMEDIATES OF ANTICANCER DRUG TAXOL USING MICROBIAL CONSORTIA

Thesis presented for the degree of Doctor of Philosophy

Institute for Bioengineering

School of Engineering

The University of Edinburgh

March 2023

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. The work presented is entirely my own except where stated otherwise by reference or acknowledgement.

Any contributions from colleagues in the collaboration, such as diagrams or calibrations, are explicitly referenced in the text.

Some parts of this work have been featured in collaboration projects published in:

Malcl, Koray, Nestor Jonguitud-Borrego, Hugo Van Der Straten Waillet, Urtė Puodžiū Naitė, Emily J. Johnston, Susan J. Rosser, and Leonardo Rios-Solis. 2022. “ACtivE: Assembly and CRISPR-Targeted in Vivo Editing for Yeast Genome Engineering Using Minimum Reagents and Time.” *ACS Synthetic Biology* 11(11):3629–43.

Nowrouzi, Behnaz, Rachel A. Li, Laura E. Walls, Leo d’Espaux, Koray Malcı, Lungang Liang, Nestor Jonguitud-Borrego, Albert I. Lerma-Escalera, Jose R. Morones-Ramirez, Jay D. Keasling, and Leonardo Rios-Solis. 2020. “Enhanced Production of Taxadiene in Saccharomyces Cerevisiae.” *Microbial Cell Factories* 19(1).

Santoyo-Garcia, Jorge H.; Walls Laura E.; Valdivia-Cabrera, Marissa; Malci, Koray, Jonguitud-Borrego, Nestor; Halliday, Karen J; Rios-Solis, Leonardo. 2023. “The Synergetic Effect from the Combination of Different Adsorption Resins in Batch and Semi-Continuous Cultivations of S. Cerevisiae Cell Factories to Produce Acetylated Taxanes Precursors of the Anticancer Drug Taxol.” *BioRxiv : The Preprint Server for Biology*.

To my grandmother Gloria Robles Torres, in loving memory

Abstract

Paclitaxel —commercially known as taxol— is one of the most widely used drugs to treat different types of cancer. Its production using microbial platforms has been challenging due to the complexity of the molecule and the difficulty of expressing functional enzymes. Expression of many heterologous enzymes in microorganisms often leads to an increased metabolic burden that usually causes growth and production decrease. Division of Labour can offer great advantages as the metabolic burden can be divided into two or more specialised organisms.

In this work, the use of microbial consortia to produce oxygenated and acetylated paclitaxel precursors was examined. The aim of Chapter 3 was specialisation by engineering various GAL80 mutant strains which could produce taxanes without the need for galactose for induction, and to further expand the Taxol metabolic pathway by incorporating the enzyme Taxane-10β-Hydroxylase (T10βOH). Using glucose showed a decreased Taxadiene production of up to 50% in comparison to galactose, probably because MIG1 is believed to cause repression in the presence of this sugar. Sucrose and fructose yielded very poor taxadiene results around 90% lower than galactose to produce taxanes. Conversely, using raffinose yielded the best production for Taxa-4, 11-dien-5a-yl acetate (T5αAc) at 72 ± 15 mg/L which represented a 7-fold increase compared to previous literature data. Despite the good production results with raffinose and galactose, the T10βOH product was not detected.

The next section aimed to develop microbial consortia and an inoculum engineering strategy to enable the production of T10β-ol. Different plasmids carrying the T5αOH, TAT and T10βOH enzymes were assembled for expression in an engineered ethanologenic *E. coli* capable of growing *on xylose*. As result, the product of the T10βOH enzyme was detected and quantified for the first time using the consortia. A production of 8 ± 0.30 mg/L of T10βol was found when fermenting glucose with a yeast strain with GAL80 deletion which was only capable of producing taxadiene proving that the pathway was efficiently divided. Finally, as a second approach, two plasmids carrying TAT and T10βOH enzymes were assembled to form a consortium integrated by two *S. cerevisiae* strains. The inoculation ratio of 3:1 (EJ2 10β: EJ2 TAT) resulted in a production of 26.5 ±5 mg/L of Taxa-4, 11-dien-5α-acetoxy-10β-ol (T10βol), the

highest ever reported for this compound. It is important to mention that this consortium was the most ambitious of all as it involved 2 different yeast strains plus 3 different bacteria strains, which to the best of our knowledge is one of the most successful and complex consortia including five microorganisms diving a synthetic pathway and achieving true synergy to divide the metabolic burden among the consortia.

Following the detection of T10βol, we aimed to test a novel enzyme candidate (T1βOH) to produce the next paclitaxel intermediate taxadiene-5a-acetoxy-1a,10b-diol (T1β-ol) using optimized co-cultures described in chapter 4. After adjusting inoculation ratios and other parameters based on the evidence obtained from previous experiments we managed to detect and quantify the new T1β-ol compound with an estimated production of 45 ±3 mg/L.

The strategies of inoculum engineering, microbial consortia and conversion towards constitutive expression have shown to be very suitable to expand and optimize the paclitaxel metabolic pathway. These strategies also proved effective for the detection and production of a new compound highlighting the potential for its application in other metabolic routes practically and rapidly.

Lay summary

Paclitaxel is a very efficient drug used for cancer treatment. However, the molecule of this drug is very complex which has limited its production using microorganisms.

Traditionally, the genome of an organism can be altered to make such an organism produce a specific compound (vaccines, antibiotics, or diverse chemicals). However, repeatedly altering the genome can have a detrimental effect not only on production but also on the growth of the organism. To avoid the repeated alteration of the genome of a microorganism the use of microbial consortia is suggested.

Using microbial consortium –a group of microorganisms—, it is possible to delegate specific tasks to different microorganisms to increase the production of specific compounds. This concept, known as division of labour, offers the advantage of reducing the detrimental effect of repeatedly altering the genome of one microorganism. Instead, different DNA modifications are split in different organisms, securing optimised growth and production.

In this work, the use of microbial consortia to produce paclitaxel precursors was examined. In chapter 3 we modified the genome of a microbe so it could grow in different carbon sources. By using raffinose the best products for the precursor T5αAc was achieved with a production of 72 ± 15 mg/L which represented a 7-fold increase compared to previous literature data.

In the next section, we tested different approaches to mixing microorganisms including bacteria and yeast. The best approach was mixing different yeasts with different plasmids (a circular DNA that can self-reply) and changing how much of each organism was used at the beginning of the experiment. This strategy enabled the production of the precursor T10βol with production levels of ± 0.31 8mg/L mixing *E. coli* and yeast while mixing different yeast strains achieved the highest-ever production of this compound with a total amount of 26.5 ±5 mg/L. In Chapter 5, the same approach was used to study a novel enzyme that could produce a new precursor known as T1βOH. By using this approach, we managed to detect the new compound with a total production of 45mg/L. Finally, we sequenced the whole genome of one of

our strains to verify all the genetic modifications and we were able to confirm their absence or presence.

The strategies of mixing different microorganisms and adjusting how much of each is added at the beginning of each experiment allowed for increasing some paclitaxel precursors and even the discovery of new compounds. The results achieved in this work highlight the potential for its application in the discovery of new compounds or optimisation in the production of others.

Acknowledgements

These four years working on my PhD degree, have allowed me to grow both academically and on a personal level. However, this was a result of the support of different persons that deserve my full respect and gratitude.

First, I would like to thank the National Council of Science and Technology (CONACyT) for supporting my studies during these 4 years so I can pursue my dream to become a doctor.

I would also like to recognise the great support and guidance that I received from my supervisor Leonardo Rios Solis. Thanks to Leo's support and guidance I was able to overcome many of the challenges that I encountered during this journey. Thank you for understanding the huge weight of being a PhD student and having a family during my studies and for always making sure I was in good condition to carry on with my work. I will be forever grateful for all the feedback and suggestions given always with patience and in a very assertive manner like only a true leader and friend can do.

I would like to recognise the support provided by my colleagues from my group and from IBioe: Ainoa, Alex Speakman, Alex Sturtivant, Behnaz, Jorge, Koray, Laura, Mariachiara and Rafa who immensely supported and gave me advice with those difficult experiments. Sorry guys for distracting you with my complaints and jokes and thank you for your friendship.

I would also like to thank the lab staff Mark and Stuart for always being kind, patient and helpful when I needed them to help me with my analysis. Thank you for helping me to analyse the "million" GC-MS samples from my experiments.

I want to especially thank Kata who was a great friend to me since day one. Thanks for making sure nothing was ever missing in the lab, for looking after the equipment when it needed maintenance but more importantly thank you for always listening and giving great advice and being there when I needed it the most. I will always be in debt to you.

I would also like to thank my friends and family in Mexico for cheering me in the distance. Thank you to my mother-in-law Aurora for looking after Elias when I was at the lab doing experiments or writing at the library.

Thank you to my brothers Erick and Javi for always believing in me, cheering me in the difficult moments and being my best friends ever since we were kids.

Special mention to my parents Claudina and Nestor. Without you, I could have never accomplished anything of what I have done. Thank you for staying up late when I was a kid to help me study for my exams and make sure that I always had the support I needed to achieve my goals. I owe you everything.

Finally, and most importantly a special mention to my son Elias and my wife Ana. Thank you, Elias, because although you are too young to realise, your laughter and smile cheered me up like nothing could and gave me the strength to carry on and always give my best when things got difficult. Also, thanks to Ana, the love of my life and my partner in crime. Since I started with this crazy idea of pursuing a PhD you always believed in me and always stood by my side no matter how hard things got. Thank you for always cheering me through the hard times and for helping me to get up when things got difficult. Thank you for being an amazing wife and my best friend. And thank you for being an amazing mother to Elias. This is not my achievement, it is ours.

Index

CHAPTER 1:

INTRODUCTION

Chapter 1: Introduction

1.1 Taxanes production in microbial cells

Ever since the first recombinant protein was first expressed in *E. coli* (Ai et al., 2019)(Itakura et al., 1977) heterologous protein production has significantly increased and expanded to numerous hosts. The first heterologous protein ever expressed in yeast was human leukocyte interferon (Hitzeman et al., 1981) and the first commercially FDA-approved protein was surface antigens of hepatitis B virus (Valenzuela et al., 1982). Ever since, diverse eukaryotic and prokaryotic expression platforms like plant cells, mammalian cells, yeast, and bacteria have been used for the production of diverse therapeutics such as vaccines, enzymes, hormones, antibodies, and several other compounds to treat non-life threatening and life-threatening sickness and diseases (Tripathi & Shrivastava, 2019).

As scientific advances and bioprocesses start to modernize and become more refined, so do all these biotechnological products. However, despite the enormous advances in the areas mentioned, some compounds are still hard and/or very expensive to produce due to their complexity (H. Zhang et al., 2011). Among these, we can mention taxanes, which are a class of diterpene compounds (Kusari et al., 2014). These compounds –usually found in plants— have properties that have been discovered among many other things to work as powerful inhibitors of cell growth hence they've been used as very effective anti-cancerogenic drugs (Kusari et al., 2014).

Among these compounds, we can mention paclitaxel –commercially found as Taxol®— (Kingston, 2001). For a long period of time since its discovery in the 1960s, the main method of obtention was the extraction from *T. brevifolia* which consequently caused a depletion of this tree species (Gallego-Jara et al., 2020). Furthermore, the method of obtention was calculated to be at least 10 times the budget available for clinical trials which as a consequence started a race for the development of a chemical synthesis route an event that finally happened in 1994 (Nicolaou et al., 1994). The newly developed method proved to be too expensive to produce for the National Institute of Cancer of the US until Bristol-Myers Squibb (BMS) took over the patent and started the production with a selling price of $986 per dose which –once the patent expired—dropped to $150 per dose for the generic version (Gallego-Jara et al., 2020).

The current methods of production have not changed over the years as the options still are chemical extraction from plants of the genus *Taxus* or by semi-synthetic production from its precursor 10-deacetylbaccatin III (Gallego-Jara et al., 2020). Different approaches for the production and scale-up of this compound by synthetic biology approaches have been performed but results are still very far from optimal as yields are usually low or its production is still too expensive to be profitable (Sharma et al., 2022). Some of these approaches along with their advantages and disadvantages are discussed in the next paragraphs.

More recently, efforts have been done to improve bioprocessing methods for the production of taxane products where production of taxanes has reached up to 39.0 ± 5.7 mg/L at microscale optimization and up to 98.9 mg/L in mini bioreactors (Walls et al., 2021, 2022). Furthermore, advanced extraction methods have increased the recovery of taxanes increasing the total amount of taxanes a two-fold improvement compared to the traditional dodecane overlay liquid-liquid extraction approach resulting in a recovery of 61 ± 8 mg/L in microscale and 76 ± 19 mg/L using 500mL bioreactors (Santoyo-Garcia et al., 2022). Other methods for improving taxane production in bacteria and yeast are discussed more deeply in the next sections.

Finally, some research has also been done on using microbial consortia towards the production of taxanes. The most relevant on this topic is the one done by (Kang Zhou et al., 2015) (Figure 1). In this work, a consortium of *S. cerevisiae* and *E. coli* was established to enhance the production of paclitaxel precursors. The consortia were able to produce 33 mg/L of oxygenated taxanes, including Taxa-4, 11-dien-5a-acetoxy-10b-ol (T10β-ol) (Kang Zhou et al., 2015). This work is one of the most important ones exemplifying how complex molecules can be produced using microbial consortia.

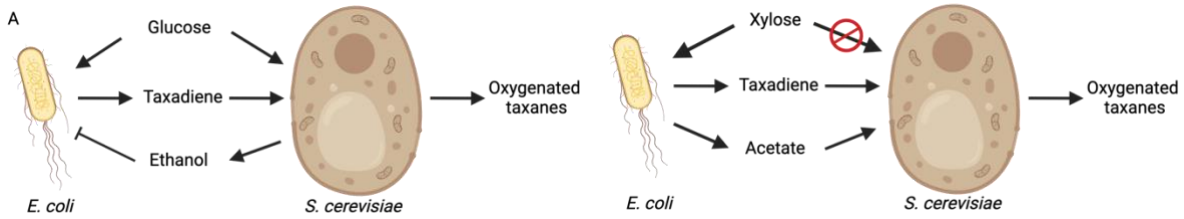


Figure 1.1 Adapted from (Kang Zhou et al., 2015). In the first image, both organisms grow on glucose however, *E. coli* is inhibited by ethanol. Taxadiene is produced by *E. coli* while the product is oxygenated by yeast, image on the right of panel A showed a similar approach for taxanes production, however, xylose (not fermentable by yeast) was consumed by *E. coli* and produced acetate that was used as a carbon source by yeast.

1.2 Optimization and discovery of new Paclitaxel pathway enzymes

Efforts have been made towards adding new enzymatic steps, discovering new enzymes to produce new taxol intermediates, and optimising the heterologous pathways already existing (Q. Huang et al., 2001; Kaspera & Croteau, 2006; Nowrouzi et al., 2020). Also, as the enzymes were originally found in plants the strategy so far has been to codon optimised them for their expression in heterologous hosts (Nowrouzi et al., 2020). However, protein engineering has been important towards the design and improvement of these proteins for their optimisation in their respective hosts. Some of the enzymes already discovered and integrated have some level of promiscuity (Yadav, 2014) and some research has focused on decreasing the level of promiscuity and increasing the level of regioselectivity to decrease the number of by-products and increasing the total yield of the desired product and precursors. The taxol metabolic pathway is shown below in Figure 1.2.

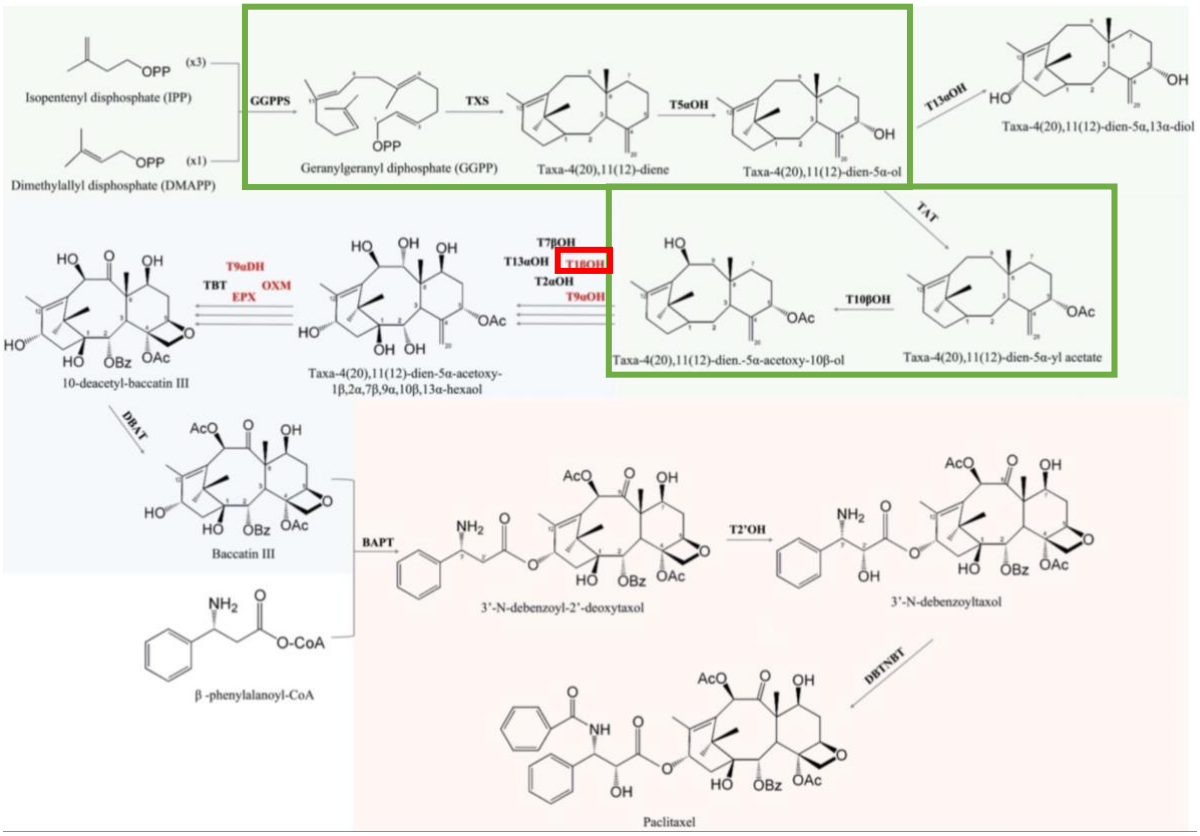


Figure 1. 2 Taxol metabolic pathway (taken and adapted from (Escrich-Montana, 2022)). MVA/MEP pathway is engineered in our S. cerevisiae strain. Enzymes squared in green are inserted in our yeast strain. TASY (also known as TXS) is expressed for taxadiene synthesis. T5αH, TAT and T10βH are also engineered in our strains for expression of Taxa-4, 11-dien-5a-ol (T5α-ol), Taxa-4, 11-dien-5a-ylacetate (T5αAc) and Taxa-4, 11-dien-5a-acetoxy-10β-ol (T10β-ol) respectively. All the potential candidate enzymes are highlighted in red. Furthermore, T1βOH is squared in red.

The first time Taxadiene was produced in *E. coli* it was accomplished by the expression of 3 different genes crucial for taxadiene production: IDP (IDP isomerase), GGDP (Geranylgeranyl diphosphate) and TASY (taxadiene synthase) (Q. Huang et al., 2001). The yield obtained was 0.5 mg/L however it was later improved to 1.3 mg/L by the overexpression of DXP (deoxyxylulose-5-phosphate) an enzyme responsible for the production of IDP (Q. Huang et al., 2001). It is interesting to remark that in this work the author also discusses a strategy for increasing the solubility of TASY by eliminating the first 78 amino acids of the protein.

Furthermore, the following steps in the paclitaxel pathway demand several oxygenation reactions. It is believed that naturally, Yew trees (*Taxus spp.*) employ up to 8 different P450's– which usually work coupled with other reductases known as CPRs (Cytochrome P450 reductase) in different oxygenations steps for paclitaxel production (Kaspera & Croteau, 2006). Hence, later attempts for taxadiene expression in *E. coli* have been focused on these enzymes.

However, because of the different cell membrane configuration in bacteria and the lipophilic N terminal amino acids of P450's (Kitaoka et al., 2015) these enzymes have been difficult to express in *E. coli* (Schuler & Werck-Reichhart, 2003). In plant membranes, the hydrophobic tails help to potentialize electron transfer interactions, however, as *E. coli* lacks such membranes this protein domain can become an obstacle (Biggs et al., 2016). Furthermore, the lack of proper post-translational modifications in prokaryotes can lead to improper folding of the protein potentially negatively affecting its activity and leading to the aggregation of the protein forming "inclusion bodies" (deposition of unfolded or partially misfolded protein) (Gasser et al., 2008).

Some solutions to work around this have been the engineering of the N-terminal transmembrane (TM) domain by deletion of bases codifying for this region and/or creating a chimeric protein with the CYP450 and its corresponding reductase (Ajikumar et al., 2010; Biggs et al., 2016). A good example of this is the creation of an N-terminal TM engineered P450 that was coupled with its reductase partner (CPR). This chimeric protein achieved a 98% conversion rate of taxadiene to the Taxa-4, 11-dien-5a-ol (T5αol) and the by-product 5(12)-Oxa-3(11)-cyclotaxane (OCT) (Ajikumar et al., 2010). Furthermore, this method was combined with a modular approach where the pathway was divided into an upstream and downstream module (the native MEP pathway and the heterologous taxadiene pathway) each tuned differently. With this approach, *E. coli* was capable of producing 1 g/ L of taxadiene (Ajikumar et al., 2010).

Another interesting example is the report of an *E. coli* strain capable of producing ~570 ± 45 mg/L of oxygenated taxanes (Biggs et al., 2016). To achieve this, researchers measured the performance of different chimeric proteins (P450-O-CPR and P450-L-CPR) with different N-terminal modifications.

Alternatively, metabolic engineering efforts have been done in eukaryotic platforms for taxane production, more specifically in *S. cerevisiae* (DeJong et al., 2006; Nowrouzi et al., 2020, 2022; K. Zhou et al., 2016). The first mention of this in the literature was given when researchers expressed *TASY from Taxus Chinensis* alone showing that an increase in the precursor geranylgeranyl diphosphate (produced by GGDP) is relevant for adequate taxadiene production (Engels et al., 2008). To overcome this, a

copy of GGDP was inserted, however, results showed that there was a competition between this enzyme and innate yeast metabolism so finally a GGDP from *Sulfolobus acidocaldarius* and a codon optimised TASY from *Taxus Chinensis* were expressed resulting in a production of 8.770.85 mg/L along with 33.175.6 mg/L of the side product geranylgeraniol (GGOH*)* (Engels et al., 2008)*.*

More recent strategies used for increased taxadiene production in *S. cerevisiae* are focused on expressing the enzymes for paclitaxel production including P450 enzymes and their reductases (Nowrouzi et al., 2020, 2022). Using *S. cerevisiae* as chassis for the expression of P450 can be advantageous as it enables an improved expression of eukaryotic enzymes in comparison to bacteria chassis while maintaining the benefits for scaling-up fermentations (Vickers et al., 2017).

Other strategies for increased taxadiene production have focused on combinatorial approaches of molecular biology with bioprocessing methods (Santoyo-Garcia et al., 2022, 2023; Walls et al., 2021, 2022). By chromosomal integration of multiple TASY copies with multiple solubility tags taxadiene titres were increased up to 57 ± 3 mg/L at 30 °C at a microscale (Nowrouzi et al., 2020). The process was then scaled up to 25 mL and 250 mL at 20 °C where production achieved was 129 ± 15 mg/L and 127 mg/L respectively (Nowrouzi et al., 2020). Furthermore, studies from the same author have also focused on exploring the optimal P450 activity in *S. cerevisiae* for the conversion of taxadiene into T5α-ol. For example, a paper published in 2022 concluded that gene dosages coupled with self-sufficient oxygenases and CPRs accelerated the formation of oxygenated taxane products (Nowrouzi et al., 2022).

It is also important to mention that as the structure of paclitaxel is extremely complex the whole metabolic pathway is not fully elucidated and currently it is believed to include up to 20 different steps that involve terpene cyclization, 9 cytochrome P450-catalyzed hydroxylations, 3 acylations, acetylations, oxetane ring formation, benzoylations, and phenylisoserine side chain attachment (Mutanda et al., 2021). From the whole pathway, it is believed that between 14 (Mutanda et al., 2021) and 16 (T. Wang et al., 2021) enzymes and steps have been fully characterised which means that there is still room for more enzymes and steps to be discovered and described.

As mentioned before, new enzymes continue to be discovered and *in silico* approaches have been used for the proposal of new candidate enzymes in the taxane biosynthetic pathway (Ramírez-Estrada et al., 2016; Sanchez-Muñoz et al., 2020). Genome-wide expression analysis coupled this with DNA-amplified fragment length polymorphism (cDNA-AFLP) and in silico analysis has proven to be useful for elucidating new enzymatic steps in the pathway (Ramírez-Estrada et al., 2016). By this method, 15 putative genes encoding for 5 steps involved in taxane biosynthesis were elucidated among which it is important to remark on the discovery of two new cytochrome P450s tagged as TB574 and TB331 (Ramírez-Estrada et al., 2016). These candidate genes were later postulated as possible taxane hydroxylases and reported to have a very high similarity with other characterised taxane hydroxylases hence it was inferred that the new putative genes would act on carbons 1β and 9α (Sanchez-Muñoz et al., 2020). Finally, a recently published thesis (Escrich-Montana, 2022) seems to confirm the hydroxylation of carbon 1β of the taxadiene core. In addition to the sequence and structural comparisons with other hydroxylases, this author performed an in silico docking analysis with 37 ligands revealing the potential of TB574 as T1βOH gene.

1.3 Microbial consortia for the production of value-added products and population control in microbial consortia

Among the many advantages of using microbial consortia to produce natural products is the division of labour (DOL) (McCarty & Ledesma-Amaro, 2019; Tsoi et al., 2019) (Figure 1.3). DOL can help to ease a heavy metabolic burden on the different organisms in a system, enabling an improved production of the product (McCarty & Ledesma-Amaro, 2019; Tsoi et al., 2019). Another remarkable advantage is that while engineering a single strain to produce a specific chemical can be difficult and time-consuming, microbial consortia can be used to maximise the native pathway of an organism which in combination with the metabolic pathway of another organism results in easier genetic engineering strategies (Tsoi et al., 2019). Furthermore, some of the challenges of bioprocessing can be mitigated by the use of microbial communities (Roell et al., 2019). For instance, the optimal environment established for bioproduction may not be the best for the organism itself, forcing it to endure adverse conditions that may impact productivity so in this aspect the microbial consortia approach takes advantage of different microorganisms allowing to choose

the most suitable ones that can act in synergy not only for bioproduction but also to overcome adverse environments (Tsoi et al., 2019).

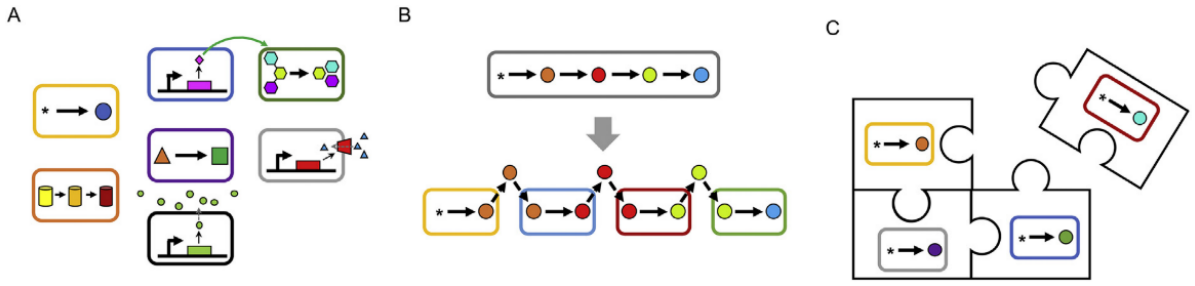


Figure 1. 3 Advantages provided by the division of labour (taken from Tsoi et al., 2019). A) Improved functionality via specialization allows each strain to focus on a specific task of the pathway B) Reduced metabolic burden can allow the microorganisms better growth and/or production C) Modularity of the system can ease manipulation and increase the versatility of the system.

There are several applications of microbial consortia using the concept of DOL and the examples span different areas. For example, in bioremediation, communities have been of particular interest as they have been proven effective for the improvement of soil for plants (Che & Men, 2019; McCarty & Ledesma-Amaro, 2019; Tsoi et al., 2019), degradation of pollutants and complex compounds that cannot be degraded by a single bacteria system such as cellulose (Cao et al., 2022; Flores et al., 2019a), dyes, antibiotics, plastics and petroleum (Cao et al., 2022).

An example of this is the degradation of azo dyes (synthetic dyes used in the food and textile industry) by a consortium of a halophilic bacteria consortium integrated mostly by bacteria of the genera *Marinobactirum, Zobellella, Enterococcus, Tistrella, Clostridiisalibacter and Caenispirillum* among others. This consortium achieved a 98% decolouration of enriched textile wastewater (Tian et al., 2021).

Another bioremediation example was the high degradation of diesel oil by *Dietzie sp.* and *Pseudomonas stutzieri.* Synergically, the degradation of the fuel achieved rates of 85.54% ±6.42% (Hu et al., 2020). In this study, each *P. stutzieri* fed on C_{16} intermediates from *Dietzia sp* while this strain received acetate and glutamate product of a metabolic carbon flux reorganization of the former.

Improvement of fuel production has also been investigated as a potential application of engineered synthetic microbial communities to improve fuel production such as bioethanol and hydrogen (A. Mishra & Ghosh, 2020; S. Wang et al., 2019). In line

with the production of bioethanol, in another recent study (A. Mishra & Ghosh, 2020), a consortium of *Zymomonas mobilis* and *Scheffersomyces shetae* was grown on sugars derived from wild sugarcane grass hydrolysate achieving a production of 25 g/L of ethanol which represents a 78.6% of the maximum theoretical production.

Recent evidence shows that distributing a pathway in a microbial consortium can provide a novel and effective solution for the production of natural products (McCarty & Ledesma-Amaro, 2019; Kang Zhou et al., 2015). However, defining the type of interaction of a microbial consortium is also hugely important as it is our limited capacity in creating ecosystems with a desired dynamic that halts the utilization in consortia for multiple purposes (Kong et al., 2018). 6 different microbial interactions have been proposed (Faust & Raes, 2012): commensalism, amensalism, cooperation (mutualism), competition and predation while some other authors also cite neutralism (the lack of interaction) (Kong et al., 2018).

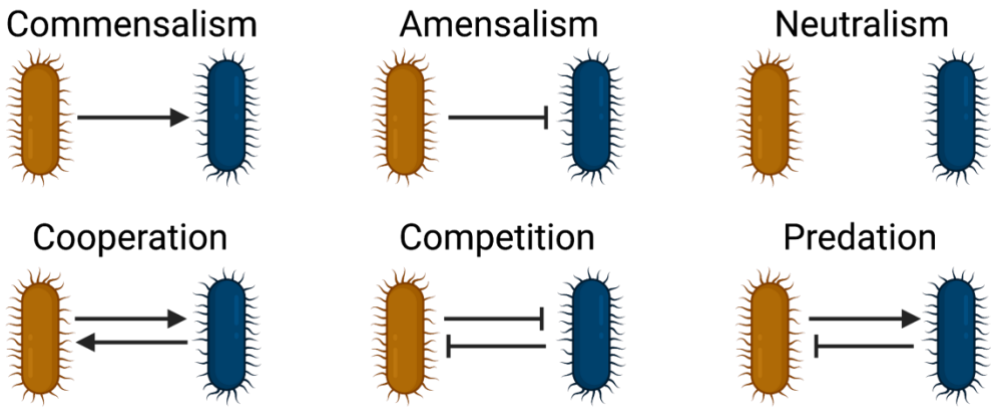


Figure 1. 4. Different types of interactions can be found in natural ecosystems. These interactions can be replicated in controlled laboratory conditions to suit the needs of particular microbial consortia (adapted from Kong et al., 2018)

In the following paragraphs, some examples of consortia, the product generated, their type of interaction and their control strategies will be given. It is important to mention that although predation, competition and amensalism exist in nature, these types of relationships are often non-desirable in synthetic microbial consortia (Kong et al., 2018). The reason for this is that the final goal of consortia itself is the additive effect achieved by the cooperation between the organisms and these types of interactions end with one organism population taking over and outnumbering the others (Kong et al., 2018) hence these interactions will not be the subject of discussion in this work.

The first interaction to be reviewed is neutralism which happens when 2 or more strains cohabit in a space without establishing any social interaction (McCarty & Ledesma-Amaro, 2019). Due to this nature in which the organisms tend to not be affected by the other or its products, neutralism is one of the most widely social interactions used in literature for the production of high-added value products in microbial consortia (McCarty & Ledesma-Amaro, 2019) as will be described in the next paragraphs.

This approach has been explored by researchers from Arizona State University where –to improve the fermentation of lignocellulosic sugar— ethanologenic strains of *E. coli* were engineered to consume xylose and glucose as carbon sources (Flores et al., 2019b). Each of these strains was specialise in metabolizing a particular sugar orthogonally which means independently from each other and without any further interaction. By this approach, there were no interactions between the strains, yet the consortium was able possible to grow together by degrading a glucose-xylose mixture that emulated natural conditions of 2:1 found in lignocellulose (Flores et al., 2019b). As one of the disadvantages of microbial consortia is their inherent unpredictability due to their increased complexity in comparison to monocultures (L. Liu et al., 2021). Orthogonality aims for a simple design in which interaction between organisms is limited and confined only to that of the purpose of the consortium decreasing the unpredictability of the system (Flores et al., 2019b). The results of this experiment showed a co-utilization of 98% of sugars present in the media with a maximum theoretical yield of 90% of ethanol production even more, the results were significantly better when compared to their respective monocultures (Flores et al., 2019b). It is important to mention that the catabolic activities were balanced by tuning the inoculum ratio at the start of the fermentation (Flores et al., 2019b).

Another example of this interaction applied to the food industry was the co-culture done with two different *Bacillus* strains where the objective was to enhance the “de-browning” of soy sauce to make it more commercially attractive (Det-udom et al., 2019). This experiment was divided into 2 different tasks the first one consisting of the consumption of xylose (a precursor in the browning reaction) by an engineered strain and the second consisting of the degradation of melanoidins, (a major pigment in the

soy sauce). Both strategies implemented synergically and orthogonally resulted in an enhanced “de-browning” of the soy sauce (Det-udom et al., 2019).

Commensalism takes place when one organism takes advantage of another without affecting it or barely doing so (Kong et al., 2018). Very interestingly this approach has been explored for energy production as shown in a paper published in 2017 where researchers developed a strategy based on the principles of division of labour and strain specialization (Y. Liu et al., 2017). Firstly, they engineered an *E.coli* strain capable of producing riboflavin and lactate as carbon source and electron donor shuttle for *Shewanella oneidensis* –a bacterium capable of reducing metal ions (Venkateswaran et al., 1999)— to generate energy. Secondly, to reduce the metabolic burden, the labour was divided into two engineered *E. coli* strains, one of them producing riboflavin and the other lactate which in consequence derived in a tenfold increase in the production of the latter product (Y. Liu et al., 2017). Lastly, some further strain specialization was done by keeping the same lactate-producing strain but replacing the other by engineering a *B. subtilis* strain capable of producing improved amounts of riboflavin. As an additional advantage, this strain was also capable of producing a certain amount of lactate which once again, resulting once again in an overall lactate production increase (Y. Liu et al., 2017).

Finally, mutualism can be generally summarised in two; facultative mutualism and obligate mutualism, the former meaning that the relationship provides benefits to the organisms over being isolated while the second meaning that they both depend on each other to survive (Venkataram et al., 2023). In recent work, researchers made use of this approach to engineer a microbial consortium with strains of *Bacillus cereus* and *Brevundimonas naejangsanensis for* hydrogen production (S. Wang et al., 2019). In this community, *B. cereus* was in charge of producing lactate as a carbon source and electron donor for *B. naejangsanensis* while this strain was producing formate for the first strain to produce hydrogen. Hydrogen production in the consortia was 42 % and 58% higher when compared to the amounts produced by monocultures (S. Wang et al., 2019).

Another relevant example of a mutualistic approach focused on a high taxadiene *E. coli* producer which was grown with a modified strain of *S. cerevisiae* to produce oxygenated taxanes (Kang Zhou et al., 2015). During this experiment, xylose was

used to feed *E. coli* which in consequence produced acetate, a substance that is inhibitory for its growth. The acetate secreted was used as a carbon source for *S. cerevisiae* allowing yeast to grow only on this substrate but also eliminating ethanol production—a common by-product of fermentation that also has an inhibitory effect on bacteria. In this paper, they obtained a production of 33mg/L of oxygenated taxanes which was the highest yield ever achieved of these compounds (up to that date) and it was the first report of oxygenated taxane production from a simple substrate in microbes (Kang Zhou et al., 2015). These results show the advantages and usefulness of microbial consortia for the synthesis of complex compounds and exploit a mutualist relationship in which without the acetate produced by *E. coli*, yeast doesn't have any carbon source and without yeast consuming acetate, the product accumulates becoming toxic for the bacteria (Kang Zhou et al., 2015).

Along with the interaction methods already reviewed for control populations, some tools have been exclusively designed for fine-tuning population balances. Among these, we can mention intercellular signalling, population control via exogenous molecules and syntrophic interactions (Kong et al., 2018; Lalwani et al., 2021; X. Li et al., 2022; Zhao et al., 2021). As research has shown that "when utilized together, these tools facilitate the assembly and control of interactive microbial consortia" (McCarty & Ledesma-Amaro, 2019). Next, some examples of this are given.

In 2018 researchers made use of *Lactococcus lactis* antimicrobial peptides nisin and lactococcin to control populations and achieved the desired interactions described (Kong et al., 2018). Using these two molecules, the authors manage to design six two-strain consortia with unique modes of interaction that included: commensalism, amensalism, neutralism, cooperation (mutualism), competition and predation (Kong et al., 2018).

Another great tool that has recently acquired notoriety is the use of optogenetics which is described as the control of gene expression by manipulation of protein domains with photosensitivity to specific light wavelengths (Zhao et al., 2021). Among the advantages of using this method for expression over chemical induction, we can mention that it doesn't alter media composition, it doesn't pose any cytotoxicity and the use of light as a regulator of expression can enhance the bioproduction (Zhao et al., 2018). Furthermore, by using optogenetics circuits researchers were capable of

regulating toxin-antitoxin systems and balancing populations in a consortium of *E. coli* and *S. cerevisiae* (Lalwani et al., 2021).

Although the use of optogenetics seems promising for tuning microbial populations, its use often requires several engineering steps and can be hard laboured if implemented from scratch, especially in new microbial platforms. Thus, there are simpler and more effective ways in which populations can be balanced towards a specific goal, for example, through the interactions mentioned previously or by the control of initial inoculation ratios (X. Li et al., 2022).

To test its stability, a consortium of *E. coli- C. glutamicum* was established and tested with a mutualistic approach (Sgobba et al., 2018). In this study, an L-lysine auxotrophic *E. coli* was engineered with an enzyme to metabolise starch and decompose it into glucose which was fed to *C. glutamicum* and allowed to grow and produce lysine for *E. coli* (Sgobba et al., 2018). By using fluorescence (each strain had different fluorescent markers for visualization and differentiation) the strains were quantified, and it could be noted that the interaction guaranteed the stability of the consortia in a constant 1:3 ratio in favour of *C. glutamicum* (Sgobba et al., 2018). This paper is an example of how choosing a carbon source and organisms that work for the desired goal is essential for balancing populations.

Furthermore, many examples mentioned before in this introduction have been noted to consider initial inoculation ratio as an important factor in their consortia design not only to optimise a particular goal (degradation or production of a compound) but also as a mean to balance populations in the consortia. Some of the studies mentioned previously and some new examples, are summarised in Table 1.1 to highlight the importance of initial inoculation ratios.

Table 1.1. Inoculation ratios used in different consortia. The table summarises the objective of the consortia and the outcome. References are included.

Consortia	Inoculation ratios	Objective	Reference
Ethanologenic *E. coli* strains LYGluc1 and LYXyl3	1:1, 1:5, 1:10, 1:100, 1:500, 1:1000	Conversion of xylose-glucose to ethanol	(Flores et al., 2019b)
***E. coli* P2C and *E. coli* BLNA**	1:1, 1:2, 1:3, 1:4, 1:5	Production of naringerin	(Ganesan et al., 2017)
2 different engineered strains of *Lactococcus lactis* NZ9000	90:1, 30:1, 20:1, 10:1, 3:1, 1:1, 1:3, 1:10, 1:20, 1:30	Design different consortia with defined social interactions	(Kong et al., 2018)
E. coli* and *S. cerevisiae	1:40	Production of taxanes, tanshinone precursors and functionalized sesquiterpenes	(Kang Zhou et al., 2015)
3 different *E. coli* strains with different modules for production of rosmarinic acid	For 2 strain co-culture: 19:1, 9:1, 5:1, 3:1, 1:1, 1:3, 1:5, 1:9 For 3 strain co-culture 1:1:2, 1:2:1, 2:1:1, 1:1:1, 1:2:2	Rosmarinic a id production by optimization of 3 strain co-cultures	(Z. Li et al., 2019)
E. coli*, *B. subtilis* and *S. oneidensis	10: 1: 100	Power generation	(Y. Liu et al., 2017)
2 different *E. coli* strains	9:1, 5:1, 4:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:4, 1:5, 1:9	Production of the indigo dye	(Chen et al., 2021)

In all the studies presented in Table 1.1, the inoculation ratio played a major role not only in the increase in the production of a specific compound but also in the balancing of the microbial populations, additionally these only consider the initial inoculation ratios of the consortia as a mean to balance the populations as there are not many studies focused solely on the impact of inoculation ratios over the diverse number of variables involved in co-culture fermentations. A recent review paper on co-culturing microbial consortia (Kapoore et al., 2022) also supports the claim that by adding an incorrect starting ratio or the addition of one microorganism at different growth than the other might lead to unbalance in the system and as consequence triggering and adverse reaction that might end with one species overtaking the other.

One of the recent works focused exclusively on analysing the impact of inoculation ratio was published in 2021 (Gao et al., 2021). In this work, different inoculation ratios of 1:1000 to 1000:1 were tested in *E. coli- P. putida* co-culture revealing that the final ratio was dependent on the initial ratios and that the initial ratio had a key role in regulating the metabolic capacity of the co-culture(Gao et al., 2021). Furthermore, it was shown that different interaction patterns were established by the co-cultures depending on the inoculation ratios. Finally, the study concluded that initial ratios impacted the reproducibility of the co-culture (Gao et al., 2021).

1.4 Genome engineering for strain specialization

One of the necessary traits of microbial consortia is the need for each strain to be optimized in a particular “step” or process within the consortia. This parallel optimization of organisms allows for strain specialization and increases the modularity of the system (Bittihn et al., 2018; McCarty & Ledesma-Amaro, 2019). Ever since the discovery of genome engineering, strains have been specialized to be used in monocultures to produce hundreds of different compounds as mentioned in section 1.1. However, with the expansion of the tools available for genome editing and cloning more and more complex pathways have been enabled for the production of natural products as will be discussed in section 1.4.

1.4.1 Strain specialization in yeast

One of the most important tools for genome engineering is CRISPR/CAS (Adli, 2018). Ever since its discovery and its first successful implementation for genome editing in 2013 (Ran et al., 2013) hundreds of several different CAS proteins have been discovered and characterised for diverse applications from genome editing to even diagnostics (Jonguitud-Borrego et al., 2022; Malcı et al., 2022). It is important to mention that a new CRISPR/Cas9 method was used for all the genomic integrations/deletions mentioned in this study. Figure 1.5 summarizes the methodology.

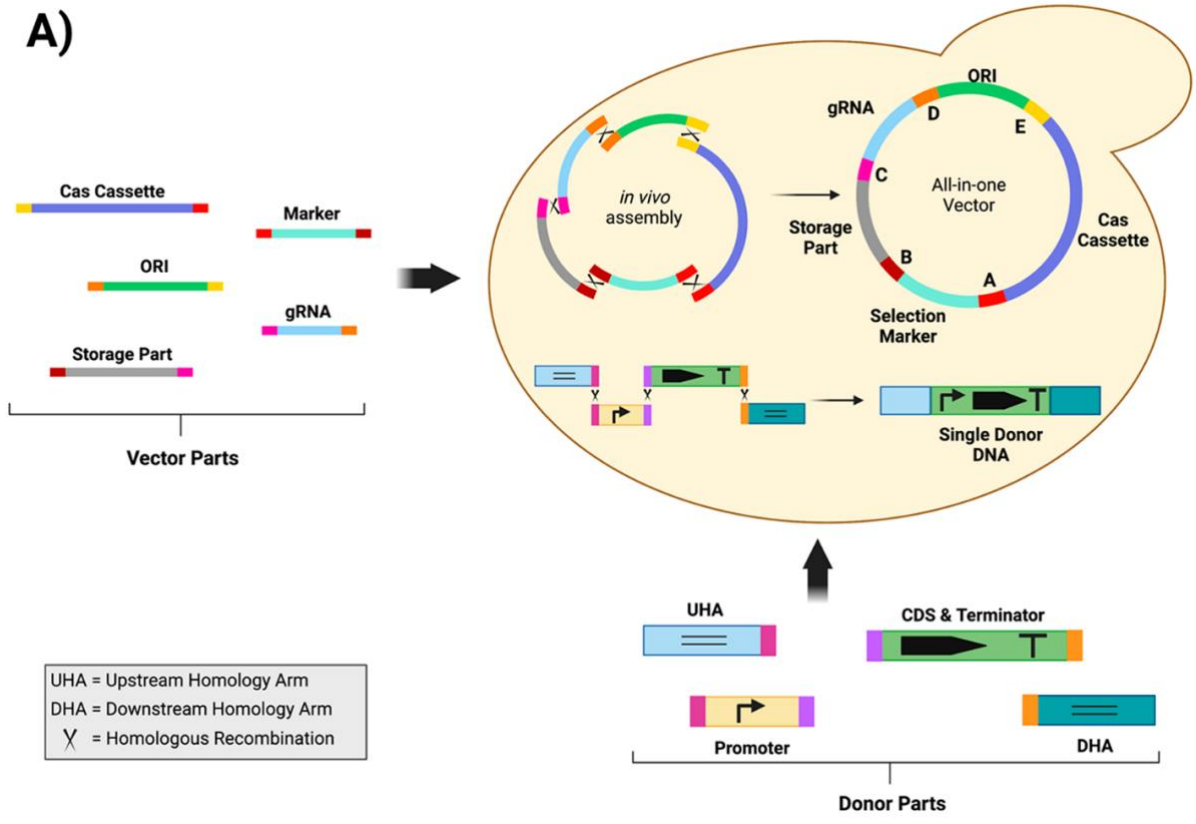


Figure 1. 5. Overview of the method used for genomic insertions/deletions. The CRISPR/CAS plasmid is divided into 5 different parts. These parts have an overlapping sequence for in vivo assembly via homologous recombination. The plasmid is transformed along with the donor DNA (usually consisting of the gene of interest and upstream and downstream homology regions to the target region) (taken from (Malcl et al., 2022)).

The expansion and creation of new CRISPR/CAS9 methods and toolkits have allowed genome engineering to become a powerful strategy towards the optimization of metabolic pathways to produce natural compounds (Malcı et al., 2020). Furthermore, multiplexing approaches for single-locus or multi-locus genomic modifications

approaches will allow speed up strain construction for yeast cell factories (Malcı et al., 2020).

One of the most recent and most impressive examples of genome editing in *S. cerevisiae* was to produce catharanthine and vindoline which are precursors of the anticancer drug vinblastine (J. Zhang et al., 2022). In this work, researchers made more than 56 edits that included the expression of 34 heterologous genes from plants along with deletions and knockdowns and further genomic engineering to overexpress 10 native genes for improvement of the precursors previously mentioned (J. Zhang et al., 2022). The result was “the longest biosynthetic pathway ever refactored into yeast”.

1.5 Aim and Objectives

The general aim of the thesis was to discover and increase the production of novel acetylated and oxygenated paclitaxel intermediates through the establishment of a microbial consortium.

For chapter 3, the main objective was to specialise our strain to increase taxanes production, and the specific objectives were as follows:

- Further engineer yeast strains to increase the production of paclitaxel precursors T/5αAc and T10βol by deletion of the GAL80 transcription repressor and insertion of T10βOH enzyme.
- Evaluate the phenotype of the engineered strains under different carbon sources using high-throughput methods.
- Scale up the production of acetylated and oxygenated taxanes using 500 mL mini bioreactors.

For Chapter 4 the objectives were:

- Design *E. coli* plasmids to produce T5αOH, TAT and T10βOH enzymes and reductases.
- Distribute production of oxygenated and acetylated taxadiene in *E. coli* – *S. cerevisiae* and *S. cerevisiae* – *S. cerevisiae* microbial consortia and evaluate their performance using high-throughput methods.

- Increase the production of T10βOH by combining genome engineering approaches with inoculum design and microbial consortia.

For Chapter 5 the objectives were:

- Testing of a novel next enzyme candidate (T1βOH) to produce the next paclitaxel intermediate using optimized co-cultures.
- Perform full genome sequencing to validate genome knock-ins and knockouts.

CHAPTER 2:

MATERIALS AND METHODS

Chapter 2: Materials and Methods

2.1 Oligonucleotides, Reagents, and Plasmids

All the chemicals were sourced from Sigma-Aldrich unless otherwise stated. All the primer design was done by using Benchling. Designed oligos were ordered as standard DNA oligos from Integrated DNA Tecnologies (IDT). Depending on the use of the parts (PCR, homologous recombination, or CRISPR/CAS assay) the fragment size of the oligos varied in length from 17 bp to 80 bp. All primers used in the study are listed in Table A1 in the Appendix section. The primer TM calculation was done by using the ThermoFisher Scientific TM calculator which can be found online at: https://shorturl.at/rtJTY.

The synthetic parts amplification for the CRISPR/CAS9 plasmid assembly was done by PCR by using Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific F548S). For colony PCR, DreamTaq Green PCR Master Mix (Thermo Fisher Scientific EP0711) was used. The PCR protocol was performed as follows:

1. In a PCR Eppendorf tube 10 µL of Master Mix, 2 µL of template DNA, 1 µL of forward primer, 1 µL of reverse primer and 6 µL of distilled water were added. Under some circumstances—such as low DNA concentration or poor amplification—the total volume was 40 µL and each of the reagents was adjusted accordingly.
2. The PCR tube was placed inside the thermocycler (ProFlex™ from Fisher Scientific) and the thermocycling parameters were adjusted according to the table shown below (Table 2.1).

Table 2.1 Thermocycling conditions used for PCR.

Step	Temperature in C°	Time	Number of cycles
Initial denaturation	95	1 min	1
Denaturation	95	30 s	30
Annealing	Primer specific	30 s	
Extension	72	1 min	
Storage	4	∞	1

3. To verify the DNA amplification 1 µL of the PCR product was loaded onto a 1% agarose gel dyed with 0.75% of SYBR-safe DNA gel stain.
4. The gel was run in 1X TAE buffer for 30 minutes at 100 V.
5. The results were visualised by using a UV light transilluminator.

Following PCR in some cases it was necessary to purify the product and, in such cases, QIAquick PCR Purification Kit was used to purify and clean up PCR-generated fragments while GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) was used for plasmid extraction, on both cases the methodology followed was that indicated by the manufacturer.

After purification, it was necessary to assess not only the DNA quantity but also its quality. This was done as follows.

1. 2 µL of the DNA recovered was analysed with a NanodropTM 2000c spectrophotometer.
2. Absorbance was measured at 260 nm and the DNA concentration is calculated automatically by the equipment. 260/280 ratios of ~1.8-2 and 260/230 ratios of ~2-2.2 are expected in pure DNA samples.
3. When necessary, the DNA sample was diluted in distilled sterile water to achieve the desired concentration.

All the CRISPR plasmid parts (*URA3:* uracil auxotrophic selection marker, 2µ ori: yeast origin of replication, *AmpR*: ampicillin resistance gene for plasmid storage in bacteria and bacterial origin of replication and CAS cassette) were provided by my colleague

Koray Malcı from my research group and amplification of those parts were needed was done with PrimeSTAR® GXL DNA Polymerase (TaKaRa) following the protocol indicated by the manufacturer. Synthetic gRNA cassettes were designed in Benchling following the methodology from (Malcl et al., 2022) and ordered from Twist Bioscience.

To look for conserved and homology regions in T5□OH and T10□OH, the sequences were downloaded in .fasta format from the LeoRios Group Benchling repository and a pair-wise alignment was performed by using MEGA11®.

2.2 Strains

All the strains in this study were the following and derived from:

- ***S. cerevisiae s10 or LRS4*** *CEN.PK2-1C (EUROSCARF),* (MATa, leu2-3, 112::HIS3MX6-GAL1p-ERG19/GAL10p-ERG8;ura3-52::URA3-GAL1p-MvaSA110G/GAL10p-MvaE [codon optimized]; his3Δ1::hphMX4-GAL1p-ERG12/GAL10p-IDI1; trp1-289::TRP1_GAL1p-CrtE(X.den)/GAL10p-ERG-20;YPRCdelta15::NatMX-GAL1p-CrtE(opt)/GAL10p-CrtE; ARS1014::GAL-282 |1p-TASY-GFP; ARS1622b::GAL1p-MBP-TASY-ERG20; ARS1114a::TDH-3p-MBP-TASY-ERG20),
- ***S. cerevisiae S14 or LRS5:*** *LRS4,* ARS511b::GAL1p-T5αOH/GAL3-CPR.
- ***S. cerevisiae KM:*** *LRS5,* RKC3::GAL1p-TAT, RKC4::GAL1p-TAT.
- ***S. cerevisiae s10.1:*** s10, ΔGAL80.
- ***S. cerevisiae S14.1***: S14, ΔGAL80.
- ***S. cerevisiae EJ1***: KM, ΔGAL80.
- ***S. cerevisiae EJ2****:* EJ1, ARS1531::Galp1-T10β
- ***S. cerevisiae EJ2-TAT****:* EJ2 but carrying a plasmid with the TAT gene.
- ***S. cerevisiae EJ2-T10β*** : EJ2 but carrying a plasmid with the T10β gene.
- ***S. cerevisiae* AE1** *KM* but carrying a plasmid with the T10β gene.
- ***E. coli*** LYglc1: frdBC::(Zm frg $celY_{Ec}$); ΔldhA::(Zm frg $casAB_{Ko}$); adhE::(Zm frg $estZ_{Pp}$ FRT); ΔackA::FRT; rrlE::(pdc adhA adhB FRT); ΔmgsA::FRT; LY180 ΔxylR; LYglc lacZ::cat-sacB (CmR).
- ***E. coli*** LYxyl3: frdBC::(Zm frg $celY_{Ec}$); ΔldhA::(Zm frg $casAB_{Ko}$); adhE::(Zm frg $estZ_{Pp}$ FRT); ΔackA::FRT; rrlE::(pdc adhA adhB FRT); ΔmgsA::FRT; ΔptsI; ΔptsG; ΔgalP; glk::kanR (KanR); xylR::xylR* (KanR).

The parental strains LRS4 and LRS5 were provided by Dr Leonardo Rios (Dr Leonardo Rios, the University of Edinburgh/ University of Newcastle) while KM was provided by Koray Malcı.

AE1 strain was kindly donated by Ainoa Escrich from Universidad Pompeu Fabra. The details of the plasmid construction can be found in (Escrich-Montana, 2022), however, at the moment of writing this dissertation the work cited had a 12-month embargo.

E. Coli strains LYglc1 and LYxyl3 were kindly provided by Professor Xuang Wang from Arizona State. More information about these strains can be found in the corresponding paper (Flores et al., 2019a).

2.3 Media

All the chemicals were sourced from Sigma-Aldrich unless otherwise stated. Different media were used for this research and all the ingredients are listed below in Table 2.2. The quantities included are for the preparation of 1L of media.

Table 2.2 Ingredients used for the preparation of different media.

Ingredient	Quantity
Yeast extract	10 g
Peptone	20 g
Carbon source (glucose, xylose, galactose, raffinose, etcetera)	200 g
Agar (optional)	20 g
Yeast nitrogen base	1.7 g
Ammonium sulphate	5 g
CSM-URA	Detailed by the manufacturer

2.3.1 YP Media

1. To prepare 1 L of YP media with a 2% carbon source concentration yeast extract and peptone were added in a Duran bottle according to Table 2.2. Agar was added when Petri dishes were needed.
2. 800 mL of distilled water was added to the bottle.

3. The chosen carbon source was added to the media in a different bottle and filled with 200 mL of distilled water.
4. Both bottles were autoclaved for 20 min
5. In a sterile environment the bottle containing the carbon source was added to the bottle of media to achieve a final concentration of 2%
6. OPTIONAL: if Petri dishes (15 mm x 100mm) were prepared, add ~20 mL to each and leave them open until the media is solidified.
7. Store Petri dishes in a cabinet or at 30 °C in an incubator.

2.3.2 CMS- URA

For the selection of positive *S. cerevisiae* transformants carrying and expressing the URA3 selection marker synthetic defined medium containing a complete supplement mixture minus uracil (CSM-Ura, MP Biomedicals) agar was used.

1. To prepare 1 L of CSM-URA media with a 2% carbon source concentration yeast nitrogen base (YNB) without amino acid and ammonium sulphate, 2% (w/v) glucose, and 2% (w/v). Yeast extract and peptone were added in a Duran bottle according to Table 2.2. Agar was added when Petri dishes were needed.
2. 800 mL of distilled water were added to the bottle.
3. The chosen carbon source was added to the media in a different bottle and filled with 200 mL of distilled water.
4. Before autoclave sterilisation pH was adjusted to 6 using a few drops of 3M NaOH. This step is important as T5□OH activity has been shown to decrease at lower pH (Hefner et al., 1996).
5. Both bottles were autoclaved for 20 min
6. In a sterile environment the bottle containing the carbon source was added to the bottle of media to achieve a final concentration of 2%
7. OPTIONAL: if Petri dishes (15 mm x 100mm) were prepared, add ~20 mL to each and leave open until the media is solidified.
8. Store Petri dishes in a cabinet or at 30 °C in an incubator.

2.4 In-Vivo Assembly and Yeast Plasmid Transformation

The following protocol details all the steps followed for plasmid assembly and plasmid transformation. With some variations, the protocol followed was the one described by (Gietz & Schiestl, 2007). Everything was done in sterile conditions.

1. Overnight cultures were typically made by adding 4.5 mL of YP media in a 50 mL falcon tube and 0.5 mL of 2% glucose for a night at 30°C in a shaking incubator at 200 rpm.
2. Between 80-100 µL from the overnight culture were added to a 50 mL falcon tube containing fresh YPD media and left inside a shaking incubator at 200 rpm and 30°C for 3 to 4 hours.
3. After this, the cells were pelleted by centrifugation at 3000 g for 5 minutes and the supernatant was discarded to be later resuspended with distilled water by gently pipetting up and down the pellet. This procedure was repeated at least 3 times to wash off any residual media.
4. During the final step, the resuspended pellet was transferred to a 1.5 mL Eppendorf tube where it was once more centrifuged and the supernatant was removed.
5. To prepare the transformation mix, 240 µL PEG (50%(w/v)), 36 µL 1.0 M lithium acetate (LiAc) and 50 µL single-stranded (previously denatured at 100° C for 5 min) carrier DNA (2.0 mg/mL) (herring sperm DNA, Promega) were mixed in a 2mL tube added to the pelleted cells.
6. Approximately ~750 fmol – ~1500 fmol from each donor DNA previously quantified in the spectrophotometer were added to the transformation mix.
7. The mixture was briefly vortexed and the solution in the tube was incubated for ~45 minutes at 42 µL using Thermomixer C (this is known as the heat shock method).
8. The cells were centrifuged for 1 minute at 3000 g and the supernatant was discarded.
9. 1mL of sterile water was added to resuspend the pellet and from this, only 200 µL were plated into a petri dish with CSM -Ura plates for 2-3 days at 30° C.

It is important to mention that the intention is that all plasmids resulting from this research will be available through the Addgene Vector Database.

2.5 Preparation of competent *E. coli* cells

For *E. coli* overnight cultures 4.5 mL of LB media (prepared as per instructions of the fabricant) from Sigma-Aldrich was used in combination with 0.5 mL of a 2% glucose or xylose solution all mixed in a 50 mL falcon tube incubated all night at 37°C in a shaking incubator at 200 rpm.

1. 250 mL of fresh LB media was inoculated with 250 μL of the overnight culture and incubated for 3 hours to reach an optical density (OD) of approximately 0.4.
2. The cells were put in ice for 10 minutes and swirled to cool evenly all the media.
3. The cells were pelleted by a 3500 g centrifugation cycle and the supernatant was discarded.
4. The cells were resuspended in 10 mL cold 100 mM $CaCl_2$ solution to then incubate on ice for 20 minutes.
5. The cells were once again centrifuged at 3500 g for 5 minutes and after discarding the supernatant, 5 mL of cold 85mM $CaCl_2$/ 15% (v/v) glycerol solution was added to resuspend the cells.
6. Aliquots of the cells in 2 mL Eppendorf tubes were stored at -80°C.

All the procedure was done in sterility conditions.

2.6 Plasmid Assembly and Bacteria Transformation

Plasmid assembly for *E. coli* was done in vitro using NEBuilder® HiFi DNA Assembly by New England Biolabs® and following the instructions detailed by the manufacturer. Plasmid transformation was done by the heat shock method.

1. The first step was to defrost an aliquot of the competent cells.
2. 50 μL of the cells were separated into a different 1.5 mL Eppendorf tube and mixed with 2 - 5 μL of the assembly and incubated on ice for 30 min.
3. Then the cells were incubated for 45 seconds at 42°C in a water batch (step can be done in Thermomixer C) and immediately after this, the tube was placed on ice for 3 minutes.

4. To heal the cells from the heat shock and reconstitute the cell wall, 500μL of LB media –previously warmed at 37°C—were added to the tube and incubated at 37°C, 300 rpm for 1 hour in the Thermomixer C.
5. Finally, 50-100 μL of the cells were spread onto LB agar plates with ampicillin (100μg/mL final concentration) and incubated at 37°C overnight.

All the procedure was done in sterile conditions. The colonies growing in the plate were selected as positive colonies. It is important to mention that the intention is that all plasmids resulting from this research will be available through the Addgene Vector Database.

2.7 Selection of positive transformants and confirmation of transformation

The transformed colonies were screened to confirm the insertion or deletion of the desired genes by the following procedure.

1. Selected colonies growing in CSM- URA medium were randomly picked with a pipette tip and incubated in 50 µL of 20 nM NaOH for 15 minutes at 98^{o}C using the thermocycler ProFlex™ from Fisher Scientific.
2. 2 µL of the lysate was used as a DNA template for PCR with specific primers (Table A1 and A2 in appendix 3 A) to confirm the genetic modifications. The reagents used for PCR are detailed in the Table 2.1.
3. To remove the CRISPR plasmids from the strains that tested positive by colony PCR, they were subjected to multiple cultivations in YPD agar Petri Dishes.
4. Special primers flanking the modified region were designed and after PCR purification of the product (QIAquick PCR Purification Kit from Qiagen was used for purification while quantification was made by analysing 2 µL with a NanodropTM 2000c spectrophotometer), this was sent to Genewiz from Azenta Life Sciences for Sanger sequencing to further confirm the deletion/insertion.

2.8 Shake flask fermentations for preliminary experiments using ethanol as a carbon source.

To evaluate the production of taxanes when ethanol is used as feedstock the first experiments were done in 250 mL shake flasks previously autoclaved.

1. Prior to the fermentation an overnight culture was prepared with 4.5 mL of YP medium and 0.5 mL of the respective carbon sources used (galactose and ethanol) at 20%.
2. The next day, the OD of the overnight culture was measured by diluting 15 µL in 985 µL of distilled water in a cuvette for the spectrophotometer and measuring absorbance at 600 nm***. The result was later multiplied by 1000 and divided by 15 which resulted in the original OD. This number was used to adjust OD to 1 for motives of standardization.
3. The strain S14 (containing mvaS, TASY (3x), T5αH + CPR, TAT) was fermented using a 250 mL shake flask at one-fifth of their capacity using ethanol at 2% and a combination of ethanol + galactose in a 3:1 ratio (both at 2%).
4. The flasks were stored in a 30ºC shaking incubator at 250 rpms for 3 days. Up to this point, everything was done under sterile conditions.
5. Finally, after the fermentation in flasks, 50 µL of the dodecane layer was separated from the culture by using a pipette and transferred to GC-MS vials for further analysis.

***This dilution method was used throughout the whole research to quantify the OD of the cultures and adjust to 1.

2.9 Microscale fermentations for the production of acetylated and oxygenated taxanes

To evaluate the best strains and conditions for the production of taxanes these were tested using Axygen's polypropylene 24 V-shaped well plates in a microplate's Thermomixers C by Eppendorf.

1. Before the fermentation, all strains were grown in overnight cultures for at least 16 hours in 50 mL Falcon tubes containing 5 mL YPD and were incubated at 30°C inside shaking incubators at 200 rpms.
2. The next day 160 µL (final 2% concentration) of the relevant carbon source was added to the well, OD was measured at 600nm using nanodrop to quantify biomass and an adequate amount of inoculum was set OD to 1 was added to the well.

3. The remaining volume was completed with YP until a volume of 1600 μL was reached.
4. Finally, 400 μL of dodecane were added to the well. To preserve sterility the plates were covered with Thermo Scientific's sterile gas-permeable adhesive films (AB0718).
5. The plates were taken to Eppendorf's Thermo mixers C –with a plate holder adapter— and were agitated at 350 RPMs at 30□C for 3 days. The preparation of the plates was done in a laminar airflow cabinet to preserve sterility. All samples were done in duplicates or triplicates unless specified.
6. Once the fermentation finished, the media was collected in 2mL Eppendorf tubes to measure the final OD.
7. The dodecane layer was separated from the culture by centrifugation at 1000 rpm for 5 minutes and 50 µL was transferred to GC-MS vials for further analysis.

2.10 Up-scale batch fermentations for the production of acetylated and oxygenated taxanes using bioreactors.

The conditions that worked better were replicated at a larger scale using MiniBio500 bioreactor by Applikon Biotechnology. The steps for bioreactor fermentation are described in the next paragraphs.

1. The first step was to prepare the media (as instructed in 2.3.1) inside the bioreactor vessel and take it to the autoclave for sterilisation with the pH and dissolved oxygen (DO) probes already in place.
2. The bioreactor vessel was closed appropriately and taken to the biohood to add the sterile carbon source was added to reach a final concentration of 2%.
3. 50 mL of dodecane were added to the vessel to achieve a total volume of 250 mL. The 20% dodecane overlay was added for *in situ* liquid-liquid extraction of the taxane products.
4. For calibration of the DO probe, the bioreactor was left connected to the stirrer at 300 rpm for a whole night.
5. After preparing an overnight culture yeast inoculation was set so that the initial OD was 1 and added to the bioreactor.

6. To prevent foam formation a few droplets were added of polypropylene glycol P2000 (antifoaming agent by Alfa Aesar) previously diluted in water and sterilised by autoclave.
7. The different parameters such as pH, DO, and temperature was controlled using my-control system software provided with the bioreactors. The bioreactors were set for 30ºC with a 70% setpoint dissolved oxygen saturation and pH was set to be within a range of 5 - 6.5. Whenever pH went out of range, it was adjusted by using a 1M NaOH solution.
8. After a specific time interval, samples were taken by using a syringe connected to a sample tube inside the bioreactor.
9. The samples taken were stored at 4°C for offline measurement of OD, ethanol content, sugar consumption and taxane production. From each sample taken the dodecane layer was separated from the culture by centrifugation at 1000 rpm for 5 minutes and 50 µL were stored in GC-MS vials for further analysis.
10. At the end of the experiment a tiny portion of the sample was analysed using a 100X lens in an inverted brightfield microscope (Leica’s model DM IRB) to check if contamination was present in the culture.

2.11 Biomass, sugars, and ethanol quantification

2.11.1 Biomass quantification

To quantify biomass, dilutions of the cultures were done in cuvettes and the sample was measured at 600 nm using a NanodropTM 2000c spectrophotometer. The result was later multiplied by 1000 and divided by the number of µL used for the dilution to finally obtain the original OD.

Alternatively, for microbial consortia, photographs of the samples using a bright field microscope (Leica’s model DM IRB) were taken to visualise and count the yeast and bacteria cells in the community. The method for counting consisted in dividing the photograph of the sample into 9 equal parts and counting the number of bacterial cells and yeast cells in just one of the squares to later multiply the result by 9. However, this method was followed only to have a rough estimation of the consortia population dynamics and as such it needs statical validation to prove its reproducibility.

2.11.2 Sugars quantification

Reducing sugars were measured by using the dinitrosalicylic acid (DNS) method described by Miller, 1959.

1. The first step was to centrifuge the sample and separate 500 μL into a new Eppendorf tube to which the same amount of DNS reagent (3,5–dinitrosalicylic acid, 1 %; sodium sulphite, 0.05 %; sodium hydroxide, 1 %) was added.
2. The solution was mixed by inversion of the tube several times and incubated for 5 minutes at 100 $^{\circ}$C.
3. While the solution was still warm, 167 μL of Rochelle salt solution was added (40 % potassium sodium tartrate tetrahydrate).
4. The tubes were allowed to cool, and absorbance was measured at 575nm using the nanodrop previously mentioned in this section.
5. Dilutions were made for samples whose measured values were below or above the range of the spectrophotometer ~0.15-1.2 g/L.
6. The reducing sugar concentration was calculated by using standard curves previously prepared with the sugar of interest.

2.11.3 Ethanol quantification

For measuring ethanol in the samples, it was necessary to remove the contaminants by the Carrez clarification method.

1. The samples were centrifuged and 100 µL of the supernatant was diluted into a new tube with 700 μL distilled water.
2. Next, 50 μL of Carrez reagent I (potassium hexacyanoferrate, 3.6 %), 50 μL of Carrez reagent II (zinc sulphate heptahydrate, 7.2 %) and 100 μL of 100mM NaOH were added to each sample tube to add up a total of 1000 µL.
3. Finally, samples were briefly vortexed and centrifuged at 3000 G for 5 minutes.
4. Once the sample was ready, ethanol quantification was done per instructions of the Ethanol Assay Kit by Megazyme (K-ETOH).

2.12 Taxanes detection by Gas Chromatography-Mass Spectrometry (GC-MS)

Taxane identification and quantification were done by GC-MS. Out of the 50μL of dodecane, only 1μL was injected into the TRACE 1300 Gas Chromatograph (Thermo Fisher Scientific, UK) which was coupled to an ISQ LT single quadrupole mass spectrometer by the same company. The column used was Thermos Scientific TG-5MS with 30m x 0.25mm x 0.25um dimensions. Settings were as follows: Injector temp 250ºC, split less injection, split flow 10ml/min, split ration 33.3, carrier flow 2ml/min. The initial GC temperature was 120ºC and held for 1 minute, then increased up to 20ºC/min until a temperature of 300ºC was achieved, and the sample was held for 3 minutes.

To identify and quantify the production of taxanes standard concentrations of taxadiene and GGOH were related to the peak area for each product and concentrations were calculated. To distinguish a peak from the background noise, the mass spectrum for each peak was analysed. Also, only peaks with three times the base high were considered as true compound peaks. The pure standards of taxadiene and GGOH were previously obtained by a donation from the Baran Lab (The Scripps Research Institute, USA) and Sigma Aldrich, respectively. Xcalibur™ Software from Thermo Fisher Scientific was used to analyse GC-MS metabolites and measure peak areas for identification and quantification.

2.13 Genomic characterization of *S. cerevisiae* EJ2

The necessary steps for the genomic characterization of strain EJ2 are detailed as follows.

1. The Thermo Scientific Yeast DNA Extraction Kit (78870) was used to extract Genomic DNA from the EJ2 strain previously grown overnight in YPD with 2% glucose. The complete protocol can be found on the manufacturer's website.
2. Following DNA extraction, the Genomic DNA (gDNA) was quantified, and purity was assessed in NanodropTM 2000c spectrophotometer at 260 nm to measure the 260/280 and 260/230 Abs ratio (pure DNA should have a 260/280 ratio of ~1.8 and 260/230 of 2 to 2.2).
3. The resulting DNA was diluted in sterile DNase free water to achieve an approximate concentration of ~200ng/μL of gDNA which was used for Next

Generation Sequencing (NGS) to observe the genetic modifications of our strain along with some other genes of interest.

4. The Rapid sequencing Kit and sequencing protocol (SQK-RAD004) from Oxford Nanopore Technologies (ONT) were used for sequencing in MinION Flongle flow cells. The protocol is only accessible through the mmanufacturer's website.
5. The MinKnow software from ONT was opened on a computer and the default settings were selected to start the sequencing with a quality score of 9.
6. The results were stored in an external hard drive for further analysis.

The results were analysed using 2 different approaches: referenced-based assembly and *de novo* assembly.

2.13.1 Referenced-based assembly

The reference-based assembly was done by using the tool analysis provided by ONT (MinKnow and Epi2me Agent).

1. In this method the reads were aligned against the *S. cerevisiae* reference genome obtained from https://www.ncbi.nlm.nih.gov/genome/?term=s.+Cerevisiae), this was done by choosing the “align to reference “ option from the menu, it’s important to mention that this alignment is only made for assembly purposes and not for comparing sequences.
2. Basecalling was made using the High accuracy algorithm (Guppy v6.3.8).
3. Once the Fastq files were generated the reads were aligned once more using EPI2ME Desktop Agent and Fastq Custom Alignment protocol which generates BAM files (Binary Alignment Map).
4. Finally, the BAM files were visualised using the Broad Institute's Integrative Genome Viewer (IGV) (https://software.broadinstitute.org/software/igv/).

2.13.2 De novo assembly

The methodology for the *de novo* assembly included the use of Line Command Tools (LCM).

1. The fastq files generated from the sequencing were assembled using *Flye*.
2. Once the alignment was made, polishing was made by using medaka and aligned using mini_align from Pomoxis.
3. The generated BAM sequences were visualised using IGV.

The code used for the *de novo* assembly is included in appendix section 5 B.

As a side note, it is important to mention that the whole NGS protocol was performed on 4 different occasions, however, only the repetition yielding the best results will be discussed. A comparison between the outputs of the 2 best sequencing runs is included in appendix section 5 A just to highlight the main differences between the repetitions.

The summarised methodology is shown in Table 2.3.

Table 2.3 Methodology used for genome assembly. The fastq files obtained from the sequencing run in the MinIon sequencing device were assembled by two different methods summarised in the table below.

Processing steps	*Reference based*	*De novo*
1.- Basecalling	MinKnow (Guppy)	MinKnow (Guppy)
2.- Assembly	MinKnow (Guppy)	Flye
3.- Polishing	----------	Medaka
4.- Alignment	Epi2Me Agent	Minimap2
5.- Visualization	IGV	IGV

2.14 Data Analysis and Software

All the bioreactor parameters (pH, DO and temperature) were measured online using the my-control system software provided by Applikon Biotechnology.

Xcalibur™ Software from Thermo Fisher Scientific was used to analyse GC-MS metabolites and measure peak areas for identification and quantification. For more information see section 2.12.

All images and figures are original creations (unless stated otherwise) using BioRender with a license provided by LeoRios Lab.

All in-silico designs for genomic assemblies were performed using Benchling®: Cloud-based platform for biotech R&D (https://www.benchling.com). The license to use this software was provided by LeoRios Lab.

MEGA11® was used for the alignment of sequences to look for conserved and homology regions.

2.15 Statistic analysis

All the statistical analyses were done by using the Excel Analysis ToolPak.

It is worth mentioning that in statistics usually 2 different hypotheses are formulated. The hypothesis null states that any difference between the means is likely due to experimental error while the alternative hypothesis states that the results are likely explained by the effect of the treatment. For example, the use of a particular carbon source or a plasmid insertion.

Rejection of hypothesis null and acceptance of alternative hypothesis and vice versa depended on the p-value calculated during the test. If this was higher, it means that hypothesis null is accepted and if it is lower, then it is rejected. It is also important to mention that the Alfa value (α) is also known as the level of significance and when this value is higher than the p-value obtained during the different tests, it means that there is a very high probability that the results were not obtained by chance but rather by an effect of the treatment tested. In other words, the results are said to be statistically significant or just significant.

When comparing the difference between the two means of an experiment with two different samples (for example with or without a particular plasmid), A T-test was done. However, an F-test (Fisher's test) was necessary to determine if the variances were equal or unequal. When hypothesis null was rejected, a T-test for unequal variances was done while a T-test for equal variances was done when hypothesis null was accepted.

When the means of 3 experiments were compared an ANOVA test was done instead of the T-test. The results were interpreted in the same way as described before for the other test (p-Value < 0.05 = hypothesis null is rejected and when p-Value > 0.05 hypothesis null is accepted).

It is important that when performing all the tests an Alpha value of 0.5 was chosen.

Finally, it is important to remark that the rationale followed in this section and for all statistical analysis has been applied in experiments of the same nature comparing different treatments and their effect on the production of natural compounds (Ai et al., 2019; Santoyo-Garcia et al., 2022)

CHAPTER 3:

GENOME ENGINEERING APPROACHES TO ENHANCE THE PRODUCTION OF T5αAC AND T10β-OL

Chapter 3: Genome engineering approaches to enhance the production of T5αAc and T10βol

3.1 Introduction

Paclitaxel —commercially known as taxol— is one of the most widely used drugs to treat different types of cancer however, its production using microbial platforms has been challenging due to the complexity of the molecule and the difficulty of expressing functional enzymes (T. Wang et al., 2021). Furthermore, several works have been done in trying to optimize taxane production in yeast using a combination of bioprocessing and bioreactor methodologies (Malcı et al., 2020; Nowrouzi et al., 2020; Santoyo-Garcia et al., 2022; Walls et al., 2021, 2022). These works have obtained good amounts of the precursors with a maximum production of up to 251 ±25 mg/L of taxadiene,135 ±6 mg/L of the oxygenated taxane Taxa-4(20), 1(12) dien-5α-ol (T5α-ol), and 95 ± 4 mg/L of Taxa-4(20), 1(12) dien-5α-yl-acetate (T5αAc), this last one being the next product in the metabolic pathway (Figure 3.1).

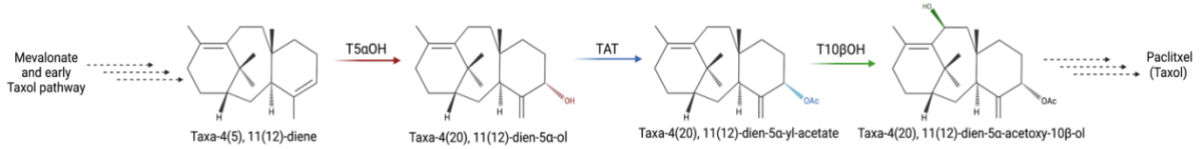


Figure 3. 1. Key enzymes from this study of the Taxol metabolic pathway.

The reduction in the amount of T5αAc when compared to the rest of the precursors can be limited towards the production and detection of the next compound of the metabolic pathway Taxa-4(20), 1(12) dien-5α-yl-acetoxy-10β-ol (T10β-ol). However, it is important to mention that although production of T10β-ol has been achieved using a consortium of *E. coli – S. cerevisiae* (Kang Zhou et al., 2015) it has never been produced and detected in yeast alone. It is also relevant to mention that although the enzyme has been expressed before in yeast the previous studies (Schoendorf et al., 2001) focused only on functional expression while ours focus also in optimization and quantification of the compound.

Another important thing worth mentioning is that for the control of the expression of the desirable compound it is crucial to have an adequate promoter sequence. Among the multiple promoters used for the control of the expression of the gene of interest,

one of the most widely studied in yeast is the GAL promoter (Å tagoj et al., 2005) Hence, promoter GAL1 is of constant use in this work for controlling the expression of paclitaxel precursor enzymes previously mentioned.

Promoter GAL1 has been –along with GAL7 and GAL10— described as one of the strongest inducible promoters in *S. cerevisiae* (Å tagoj et al., 2005). The GAL promoters are controlled by a series of diverse proteins involved in gene regulation known as the GAL regulon (Rajeshkannan et al., 2022). The proteins involved in GAL regulon signalling are GAL3, GAL80 and Gal4. GAL3 is a transcriptional regulator that under the presence of galactose forms a complex with GAL80 to allow the expression of GAL4, (Pego & Smeekens, 2000) while GAL80 is a transcriptional repressor that under the absence of galactose binds to GAL4 presence of galactose represses the transcription from the GAL promoter (Lohr et al., 1995). This mechanism is presented in a diagram in Figure 3.2.

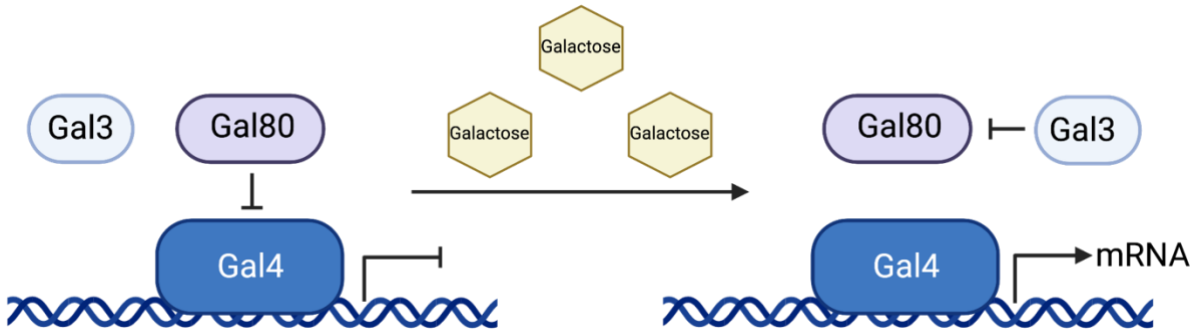


Figure 3.2 GAL regulon in yeast. In the absence of galactose, the GAL3-GAL80 complex is disrupted and GAL80 inhibits the transcription activity of GAL4. When galactose is added to the media, Gal3 inhibits GAL80 and GAL4 can activate transcription.

One of the advantages of the use of GAL promoters is that they enable the expression of a particular gene when galactose is present in the media allowing the control of induction under specific growth phases. However, genes under the control of these promoters remain repressed under the presence of any other carbon source but the “locking mechanism of expression” is even stronger when glucose is present in the media due to the action of the protein Mig1p (Nehlin et al., 1991). The action of Mig1 will be further discussed in the next paragraphs. This galactose dependency on gene expression and the tight repression under the presence of other carbon sources poses one of the biggest disadvantages of pGAL1.

To overcome the null expression of GAL genes under the presence of different carbon sources and to improve taxane production, an easy and fast-to-implement strategy was to delete GAL80. The deletion of this gene has shown to be efficient in increasing the bioproduction of Artemisinin when the responsible genes are controlled by GAL promoter (Ai et al., 2019). It has also been shown that when this gene is deleted the expression can become constitutive (Ai et al., 2019) and furthermore, when grown in glucose, GAL80 deleted strains are expected to show a diauxic-inducible expression (Peng et al., 2018). In the first phase, the microorganism starts to grow, and glucose is fermented to ethanol. In the second phase, once glucose is depleted and ethanol is used as a carbon source, upregulation of GAL promoter activities occurs in strains with this deletion (Peng et al., 2018). It is important to clarify that the upregulation is not precisely caused by the presence of ethanol but by the absence of glucose (Peng et al., 2018). This growth production dynamic is shown in Figure 3.3.

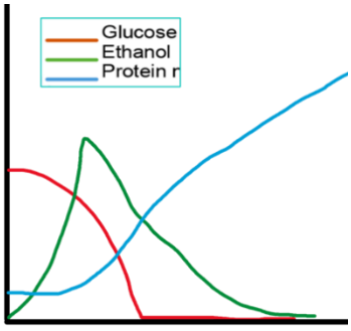


Figure 3.3 Expected growth when the GAL80 gene is deleted, as sugar is depleted and ethanol concentration increases, so do bioproduction (adapted from Peng et al., 2018).

The reason why bioproduction doesn't occur during the first growth phase lies behind a protein called Mig1p which is a zinc finger protein that binds upstream of the GAL promoters (GAL1 and GAL4) to repress its activity under the presence of glucose by a transcriptional cascade (Nehlin et al., 1991; Peng et al., 2018). Therefore, when glucose is present a "double lock" regulatory mechanism prevents GAL-controlled genes to be expressed (Å tagoj et al., 2005; Escalante-Chong et al., 2015). Furthermore, Mig 1p has also been found to inhibit other genes like SUC2 which are also repressed by glucose (Nehlin et al., 1991).

When trying to maximise the heterologous expression of a particular compound, it has been reported that deletion of Mig1p leads to a constitutive expression of GAL genes to be counterproductive as this causes a metabolic burden on the organism and as

consequence reducing growth rates which ultimately translates into a lower amount of product (Hayat et al., 2021). To overcome this issue, some authors have suggested a conditional depletion of Mig1p carried out by auxins (Hayat et al., 2021).

All the different control mechanisms described in the previous paragraphs from the (GAL genes to the Mig1) along with the strategy for overcoming galactose induction are illustrated in Figure 3.4. This figure also shows all the different heterologous genes inserted in our collection of yeast strains for paclitaxel production. The different chromosomes in which the enzymes were cloned are also noted. It can also be observed that all the genes are under the control of the pGAL1 promoter making our strains dependable in galactose for the expression of taxane genes.

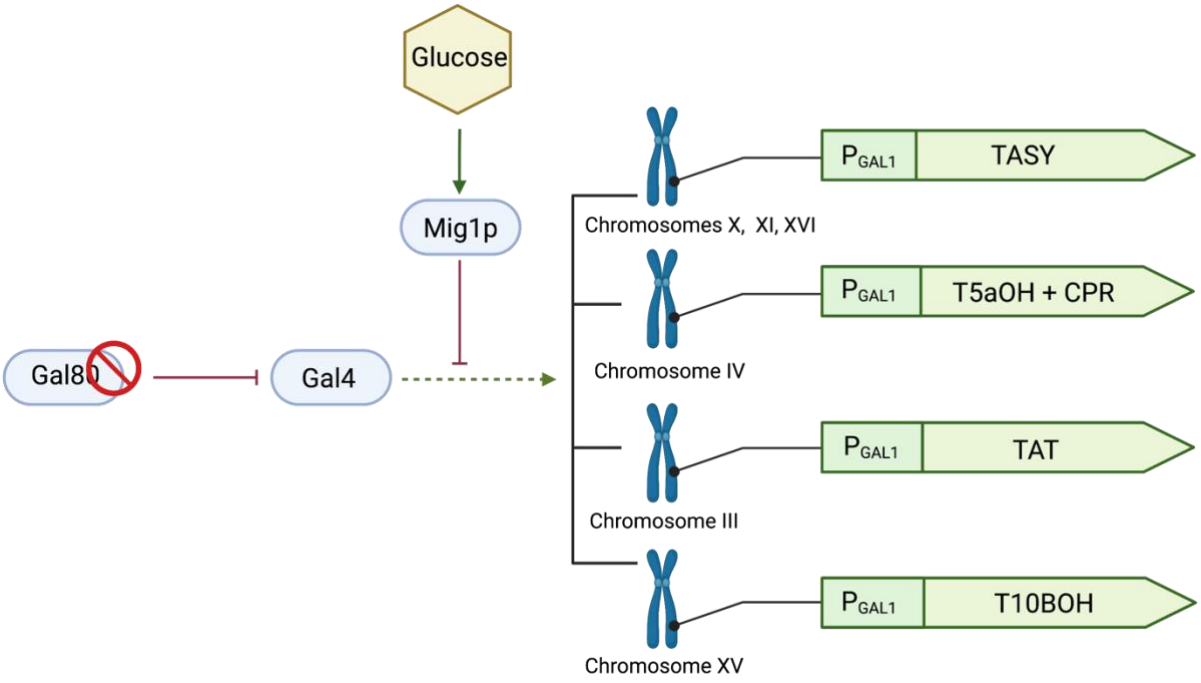


Figure 3.4. GAL80 protein acts as a repressor protein under the presence of any carbon source (other than galactose) By deleting GAL80 expression becomes constitutive except if glucose is used as a carbon source. When glucose is added to the media the activator protein GAL4 is repressed by the action of the protein Mig1p. The figure also illustrates the different genes and the different chromosomes in which they were inserted.

In this work, we aim to use a novel CRISPR/CAS9 assembly method (Malcl et al., 2022) to engineer our latest *S. cerevisiae* strain KM to enhance the production of the precursor T5αAc by deletion of GAL80. The same method was used for chromosomal insertion of the T10βOH enzyme for the conversion of T5αAc to Taxa-4, 11-dien-5a-

acetoxy-10b-ol (T10βol). A diagram containing the new enzyme and the steps enzymatic steps expressed in our yeast platforms is shown in Figure 3.1.

3.2 Methodology

3.2.1 Strains

All the strains in this study were the following and derived from:

- ***S. cerevisiae s10 or LRS4*** *CEN.PK2-1C (EUROSCARF),* (MATa, leu2-3, 112::HIS3MX6-GAL1p-ERG19/GAL10p-ERG8;ura3-52::URA3-GAL1p-MvaSA110G/GAL10p-MvaE [codon optimized]; his3Δ1::hphMX4-GAL1p-ERG12/GAL10p-IDI1; trp1-289::TRP1_GAL1p-CrtE(X.den)/GAL10p-ERG-20;YPRCdelta15::NatMX-GAL1p-CrtE(opt)/GAL10p-CrtE; ARS1014::GAL-282 |1p-TASY-GFP; ARS1622b::GAL1p-MBP-TASY-ERG20; ARS1114a::TDH-3p-MBP-TASY-ERG20),
- ***S. cerevisiae S14 or LRS5:*** *LRS4,* ARS511b::GAL1p-T5αOH/GAL3-CPR.
- ***S. cerevisiae KM:*** *LRS5,* RKC3::GAL1p-TAT, RKC4::GAL1p-TAT.
- ***S. cerevisiae s10.1:*** s10, ΔGAL80.
- ***S. cerevisiae S14.1***: S14, ΔGAL80.
- ***S. cerevisiae EJ1***: KM, ΔGAL80.
- ***S. cerevisiae EJ2****:* EJ1, ARS1531::Galp1-T10β

The parental strains LRS4 and LRS5 were provided by Dr Leonardo Rios (Dr Leonardo Rios, the University of Edinburgh/ University of Newcastle) while KM was provided by Koray Malcı.

3.2.2 Oligonucleotides, Reagents, and Plasmids

All the chemicals were sourced from Sigma-Aldrich unless otherwise stated. All the primer design was done by using Benchling. Designed oligos were ordered as standard DNA oligos from Integrated DNA Tecnologies (IDT). Depending on the use of the parts (PCR, homologous recombination, or CRISPR/CAS assay) the fragment size of the oligos varied in length from 17 bp to 80 bp. All the designs where made following the methodology described in the literature (Malcl et al., 2022). All primers

used in the study are listed in Table A1 in the Appendix section. The primer TM calculation was done by using the ThermoFisher Scientific TM calculator which can be found online at: https://shorturl.at/rtJTY.

The synthetic parts amplification for the CRISPR/CAS9 plasmid assembly was done by PCR by using Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific F548S). For colony PCR, DreamTaq Green PCR Master Mix (Thermo Fisher Scientific EP0711) was used. The PCR protocol was performed as follows:

1. In a PCR Eppendorf tube 10 mL of Master Mix, 2 mL of template DNA, 1 mL of forward primer, 1 mL of reverse primer and 6 mL of distilled water were added. Under some circumstances—such as low DNA concentration or poor amplification—the total volume was 40 mL ad each of the reagents was adjusted accordingly.
2. The PCR tube was placed inside the thermocycler (ProFlex™ from Fisher Scientific) and the thermocycling parameters were adjusted according to the table shown below (Table 3.1).

Table 3.1 Thermocycling conditions used for PCR.

Step	Temperature in C°	Time	Number of cycles
Initial denaturation	95	1 min	1
Denaturation	95	30 s	30
Annealing	Primer specific	30 s	
Extension	72	1 min	
Storage	4	∞	1

3. To verify the DNA amplification 1mL of the PCR product was loaded onto a 1% agarose gel dyed with 0.75% of SYBR-safe DNA gel stain.
4. The gel was run in 1X TAE buffer for 30 minutes at 100 V.
5. The results were visualised by using a UV light transilluminator.

Following PCR in some cases it was necessary to purify the product and, in such cases, QIAquick PCR Purification Kit was used to purify and clean up PCR-generated

fragments while GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) was used for plasmid extraction, on both cases the methodology followed was that indicated by the manufacturer.

After purification, it was necessary to assess not only the DNA quantity but also its quality. This was done as follows.

1. 2μL of the DNA recovered was analysed with a NanodropTM 2000c spectrophotometer.
2. Absorbance was measured at 260 nm and the DNA concentration is calculated automatically by the equipment. 260/280 ratios of ~1.8-2 and 260/230 ratios of ~2-2.2 are expected in pure DNA samples.
3. When necessary, the DNA sample was diluted in distilled sterile water to achieve the desired concentration.

All the CRISPR plasmid parts (*URA3:* uracil auxotrophic selection marker, 2μ ori: yeast origin of replication, *AmpR*: ampicillin resistance gene for plasmid storage in bacteria and bacterial origin of replication and CAS cassette) were provided by my colleague Koray Malcı from my research group and amplification of those parts were needed was done with PrimeSTAR® GXL DNA Polymerase (TaKaRa) following the protocol indicated by the manufacturer. Synthetic gRNA cassettes were designed in Benchling following the methodology from (Malcl et al., 2022) and ordered from Twist Bioscience.

3.2.3 Media

All the chemicals were sourced from Sigma-Aldrich unless otherwise stated. Different media were used for this research and all the ingredients are listed below in Table 3.2. The quantities included are for the preparation of 1L of media.

Table 3.2 Ingredients used for the preparation of different media.

Ingredient	Quantity
Yeast extract	10 g
Peptone	20 g
Carbon source (glucose, galactose, fructose, raffinose, melibiose, sucrose and ethanol)	200 g
Agar (optional)	20 g
Yeast nitrogen base	1.7
Ammonium sulphate	5
CSM-URA	Detailed by the manufacturer

3.2.3.1 YP Media

1. To prepare 1 L of YP media with a 2% carbon source concentration yeast extract and peptone were added in a Duran bottle according to Table 2.2. Agar was added when Petri dishes were needed.
2. 800 mL of distilled water was added to the bottle.
3. The chosen carbon source was added to the media in a different bottle and filled with 200 mL of distilled water.
4. Both bottles were autoclaved for 20 min
5. In a sterile environment the bottle containing the carbon source was added to the bottle of media to achieve a final concentration of 2%
6. OPTIONAL: if Petri dishes (15 mm x 100mm) were prepared, add ~20 mL to each and leave them open until the media is solidified.
7. Store Petri dishes in a cabinet or at 30 °C in an incubator.

3.2.3.2 CMS -URA

For the selection of positive *S. cerevisiae* transformants carrying and expressing the URA3 selection marker synthetic defined medium containing a complete supplement mixture minus uracil (CSM-Ura, MP Biomedicals) agar was used.

1. To prepare 1 L of CSM-URA media with a 2% carbon source concentration yeast nitrogen base (YNB) without amino acid and ammonium sulphate, 2%

(w/v) glucose, and 2% (w/v). yeast extract and peptone were added in a Duran bottle according to Table 2.2. Agar was added when Petri dishes were needed.

2. 800 mL of distilled water were added to the bottle.
3. The chosen carbon source was added to the media in a different bottle and filled with 200 mL of distilled water.
4. Before autoclave sterilisation pH was adjusted to 6 using a few drops of 3M NaOH. This step is important as T5□OH activity has been shown to decrease at lower pH (Hefner et al., 1996).
5. Both bottles were autoclaved for 20 min
6. In a sterile environment the bottle containing the carbon source was added to the bottle of media to achieve a final concentration of 2%
7. OPTIONAL: if Petri dishes (15 mm x 100mm) were prepared, add ~20 mL to each and leave open until the media is solidified.
8. Store Petri dishes in a cabinet or at 30 °C in an incubator.

3.2.4 Shake flask fermentations for preliminary experiments using ethanol as a carbon source.

The strain S14 was tested to evaluate the production of taxanes when ethanol is used as feedstock. This preliminary experiment was performed in 250 mL shake flasks previously autoclaved.

1. Prior to the fermentation an overnight culture was prepared with 4.5 mL of YP medium, and 0.5 mL of the respective carbon sources used (galactose and ethanol) at 20%.
2. The next day, the OD of the overnight culture was measured by diluting 15 μL in 985 μL of distilled water in a cuvette for the spectrophotometer and measuring absorbance at 600 nm***. The result was later multiplied by 1000 and divided by 15 which resulted in the original OD. This number was used to adjust OD to 1 for motives of standardization.
3. The strain S14 (containing mvaS, TASY (3x), T5αH + CPR, TAT) was fermented using a 250 mL shake flask at one-fifth of their capacity using ethanol at 2% and a combination of ethanol + galactose in a 3:1 ratio (both at 2%).
4. The flasks were stored in a 30°C shaking incubator at 250 rpms for 3 days. Up to this point, everything was done under sterile conditions.

5. Finally, after the fermentation in flasks, 50 µL of the dodecane layer was separated from the culture by using a pipette and transferred to GC-MS vials for further analysis.

***This dilution method was used throughout the whole research to quantify the OD of the cultures and adjust to 1.

3.2.5 In-Vivo Assembly and Yeast Plasmid Transformation

The strains S10, S14 and KM were engineered for GAL80 deletion while the resulting strain EJ2 was further engineered for T1BOH insertion and MIG1P deletion (see Table 3.4). The following protocol details all the steps followed for in vivo plasmid assembly (van Leeuwen et al., 2015; Van Leeuwen et al., 2015) and plasmid transformation. With some variations, the protocol followed was the one described by (Gietz & Schiestl, 2007). Everything was done in sterile conditions.

1. Overnight cultures were typically made by adding 4.5 mL of YP media in a 50 mL falcon tube and 0.5 mL of 2% glucose for a night at 30°C in a shaking incubator at 200 rpm.
2. Between 80-100 µL from the overnight culture were added to a 50 mL falcon tube containing fresh YPD media and left inside a shaking incubator at 200 rpm and 30°C for 3 to 4 hours.
3. After this, the cells were pelleted by centrifugation at 3000 g for 5 minutes and the supernatant was discarded to be later resuspended with distilled water by gently pipetting up and down the pellet. This procedure was repeated at least 3 times to wash off any residual media.
4. During the final step, the resuspended pellet was transferred to a 1.5 mL Eppendorf tube where it was once more centrifuged, and the supernatant was removed.
5. To prepare the transformation mix, 240 µL PEG (50%(w/v)), 36 µL 1.0 M lithium acetate (LiAc) and 50 µL single-stranded (previously denatured at 100□ C for 5 min) carrier DNA (2.0 mg/mL) (herring sperm DNA, Promega) were mixed in a 2mL tube added to the pelleted cells.
6. Approximately ~750 fmol – ~1500 fmol from each donor DNA previously quantified in the spectrophotometer were added to the transformation mix.

7. The mixture was briefly vortexed and the solution in the tube was incubated for ~45 minutes at 42ºC using Thermomixer C (this is known as the heat shock method).
8. The cells were centrifuged for 1 minute at 3000 g and the supernatant was discarded.
9. 1mL of sterile water was added to resuspend the pellet and from this, only 200 µL were plated into a petri dish with CSM -Ura plates for 2-3 days at 30ºC.

It is important to mention that the intention is that all plasmids resulting from this research will be available through the Addgene Vector Database.

3.2.6 Microscale fermentations for the production of acetylated and oxygenated taxanes

To evaluate the best strains and conditions taxanes production, strains S10.1, S14.1, EJ1 and EJ2 were tested using Axygen's polypropylene 24 V-shaped well plates in a microplate's Thermomixers C by Eppendorf.

1. Before the fermentation, all strains were grown in overnight cultures for at least 16 hours in 50 mL Falcon tubes containing 5 mL YPD and were incubated at 30ºC inside shaking incubators at 200 rpms.
2. The next day 160 µL (final 2% concentration) of the relevant carbon source was added to the well, OD was measured at 600nm using nanodrop to quantify biomass and an adequate amount of inoculum was set OD to 1 was added to the well.
3. The remaining volume was completed with YP until a volume of 1600 µL was reached.
4. Finally, 400 µL of dodecane were added to the well. To preserve sterility the plates were covered with Thermo Scientific's sterile gas-permeable adhesive films (AB0718).
5. The plates were taken to Eppendorf's Thermo mixers C –with a plate holder adapter— and were agitated at 350 RPMs at 30ºC for 3 days. The preparation of the plates was done in a laminar airflow cabinet to preserve sterility. All samples were done in duplicates or triplicates unless specified.

6. Once the fermentation finished, the media was collected in 2mL Eppendorf tubes to measure the final OD.
7. The dodecane layer was separated from the culture by centrifugation at 1000 rpm for 5 minutes and 50 µL was transferred to GC-MS vials for further analysis.

3.2.7 Up-scale batch fermentations for the production of acetylated and oxygenated taxanes using bioreactors.

The conditions that worked better were replicated at a larger scale with strains EJ1 and EJ2 by using a MiniBio500 bioreactor by Applikon Biotechnology. The steps and conditions for bioreactor fermentation are based on those previously reported by Walls et al., (2021) and are described in the next paragraphs.

1. The first step was to prepare the media (as instructed in 2.3.1) inside the bioreactor vessel and take it to the autoclave for sterilisation with the pH and dissolved oxygen (DO) probes already in place.
2. The bioreactor vessel was closed appropriately and taken to the biohood to add the sterile carbon source (glucose and raffinose) was added to reach a final concentration of 2%.
3. 50 mL of dodecane were added to the vessel to achieve a total volume of 250 mL. The 20% dodecane overlay was added for *in situ* liquid-liquid extraction of the taxane products.
4. For calibration of the DO probe, the bioreactor was left connected to the stirrer at 300 rpm for a whole night.
5. After preparing an overnight culture yeast inoculation was set so that the initial OD was 1 and added to the bioreactor.
6. To prevent foam formation a few droplets were added of polypropylene glycol P2000 (antifoaming agent by Alfa Aesar) previously diluted in water and sterilised by autoclave.
7. The different parameters such as pH, DO, and temperature was controlled using my-control system software provided with the bioreactors. The bioreactors were set for 30ºC with a 70% setpoint dissolved oxygen saturation and pH was set to be within a range of 5 - 6.5. Whenever pH went out of range, it was adjusted by using a 1M NaOH solution. The experiment was run for 4 days.

8. After a specific time interval (~4 hrs), samples were taken by using a 20 mL syringe connected to a sample tube inside the bioreactor.
9. The samples taken were stored at 4°C for offline measurement of OD, ethanol content, sugar consumption and taxane production. From each sample taken the dodecane layer was separated from the culture by centrifugation at 1000 rpm for 5 minutes and 50 μL were stored in GC-MS vials for further analysis.
10. At the end of the experiment a tiny portion of the sample was analysed using a 100X lens in an inverted brightfield microscope (Leica’s model DM IRB) to check if contamination was present in the culture.

It is important to mention that for these experiments it was not possible to perform any replicates so in both bioreactor fermentations the n=1.

3.2.8 Biomass, sugars, and ethanol quantification

3.2.8.1 Biomass quantification

To quantify biomass, dilutions of the cultures were done in cuvettes and the sample was measured at 600 nm using a NanodropTM 2000c spectrophotometer. The result was later multiplied by 1000 and divided by the number of µL used for the dilution to finally obtain the original OD.

3.2.8.2 Sugars quantification

Reducing sugars were measured by using the dinitrosalicylic acid (DNS) method described by Miller, 1959.

7. The first step was to centrifuge the sample and separate 500 µL into a new Eppendorf tube to which the same amount of DNS reagent (3,5–dinitrosalicylic acid, 1 %; sodium sulphite, 0.05 %; sodium hydroxide, 1 %) was added.
8. The solution was mixed by inversion of the tube several times and incubated for 5 minutes at 100 °C.
9. While the solution was still warm, 167 µL of Rochelle salt solution was added (40 % potassium sodium tartrate tetrahydrate).
10. The tubes were allowed to cool, and absorbance was measured in plastic cuvettes at 575nm using the nanodrop previously mentioned in this section.

11. Dilutions were made for samples whose measured values were below or above the range of the spectrophotometer ~0.15-1.2 g/L.
12. The reducing sugar concentration was calculated by using standard curves previously prepared with the sugar of interest.

3.2.8.3 Ethanol quantification

For measuring ethanol in the samples, it was necessary to remove the contaminants by the Carrez clarification method.

5. The samples were centrifuged and 100mL of the supernatant was diluted into a new tube with 700 µL distilled water.
6. Next, 50 μL of Carrez reagent I (potassium hexacyanoferrate, 3.6 %), 50 µL of Carrez reagent II (zinc sulphate heptahydrate, 7.2 %) and 100 μL of 100mM NaOH were added to each sample tube to add up a total of 1000mL.
7. Finally, samples were briefly vortexed and centrifuged at 3000 G for 5 minutes.
8. Once the sample was ready, ethanol quantification was done per instructions of the Ethanol Assay Kit by Megazyme (K-ETOH).

3.2.9 Taxanes detection by Gas Chromatography-Mass Spectrometry (GC-MS)

Taxane identification and quantification were done by GC-MS. Out of the 50µL of dodecane, only 1 µL was injected into the TRACE 1300 Gas Chromatograph (Thermo Fisher Scientific, UK) which was coupled to an ISQ LT single quadrupole mass spectrometer by the same company. The column used was Thermos Scientific TG-5MS with 30m x 0.25mm x 0.25um dimensions. Settings were as follows: Injector temp 250^{o}C, split less injection, split flow 10ml/min, split ration 33.3, carrier flow 2ml/min. The initial GC temperature was 120^{o}C and held for 1 minute, then increased up to 20^{o}C/min until a temperature of 300^{o}C was achieved, and the sample was held for 3 minutes.

To identify and quantify the production of taxanes standard concentrations of taxadiene and GGOH were related to the peak area for each product and concentrations were calculated. To distinguish a peak from the background noise, the mass spectrum for each peak was analysed. Also, only peaks with three times the base high were considered as true compound peaks. The pure standards of taxadiene

and GGOH were previously obtained by a donation from the Baran Lab (The Scripps Research Institute, USA) and Sigma Aldrich, respectively. Xcalibur™ Software from Thermo Fisher Scientific was used to analyse GC-MS metabolites and measure peak areas for identification and quantification.

3.2.10 Statistic analysis

All the statistical analyses were done by using the Excel Analysis ToolPak.

An ANOVA test was done to compare if there was a significant difference in taxadiene production between S14, EJ1 and EJ2 and a t-test comparing the means of S14 vs EJ1 was also done. For both experiments, 2 different hypotheses were formulated. The hypothesis null was that GAL80 deletion does not have an effect in taxadiene production while hypothesis alternative indicated that GAL80 deletion does have an effect in taxadiene production.

Statistical analysis for:

Hypothesis null is that GAL80 deletion will not affect the production of taxadiene.

Anova:
Single
Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
S14	3	235.744212	78.581404	172.96153
EJ1	3	202.229437	67.409812	149.70031
EJ2	3	228.371378	76.123792	248.28021

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	206.7777699	2	103.388885	0.5432541	0.6069554	5.1432528
Within Groups	1141.884114	6	190.314019			
Total	1348.661884	8				

3

Statistical analysis for:

Hypothesis null is that GAL80 deletion will not affect the production of taxadiene.

t-Test: Two-Sample Assuming Equal Variances

	S14	*EJ1*
Mean	78.58140404	67.4098122
Variance	172.9615298	149.700315
Observations	3	3
Pooled Variance	161.3309224	
Hypothesized Mean Difference	0	
df	4	
t Stat	1.077213727	
P(T<=t) one-tail	0.171008398	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.342016796	
t Critical two-tail	2.776445105	

An F-test followed by a T-test analysis was made to compare if melibiose influenced the production of taxanes in EJ1.

Statistical analysis for:

Hypothesis null is that melibiose does not affect the production of taxanes in EJ1.

F-Test Two-Sample for Variances

	Fructose	*Mel+ Fruc*
Mean	138.4281751	141.8189869
Variance	298.6281055	389.4843156
Observations	3	3
df	2	2
F	0.766726909	
P(F<=f) one-tail	0.433981565	
F Critical one-tail	0.052631579	

t-Test: Two-Sample Assuming Unequal Variances

	Fructose	*Mel+ Fruc*
Mean	138.4281751	141.8189869
Variance	298.6281055	389.4843156
Observations	3	3
Hypothesized Mean Difference	0	
df	4	
	-	
t Stat	0.223889955	
P(T<=t) one-tail	0.416906669	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.833813338	
t Critical two-tail	2.776445105	

Finally, an F-test followed by a T-test analysis was made to compare if raffinose influenced the production of taxanes in EJ2.

Statistical analysis for:

Hypothesis null is that melibiose does not affects the production of taxanes in EJ2.

F-Test Two-Sample for Variances

	Fructose	*Mel+ Fruc*
Mean	205.3741036	239.0413876
Variance	328.48648	44.39909952
Observations	2	2
df	1	1
F	7.398494193	
P(F<=f) one-tail	0.224285388	
F Critical one-tail	161.4476388	

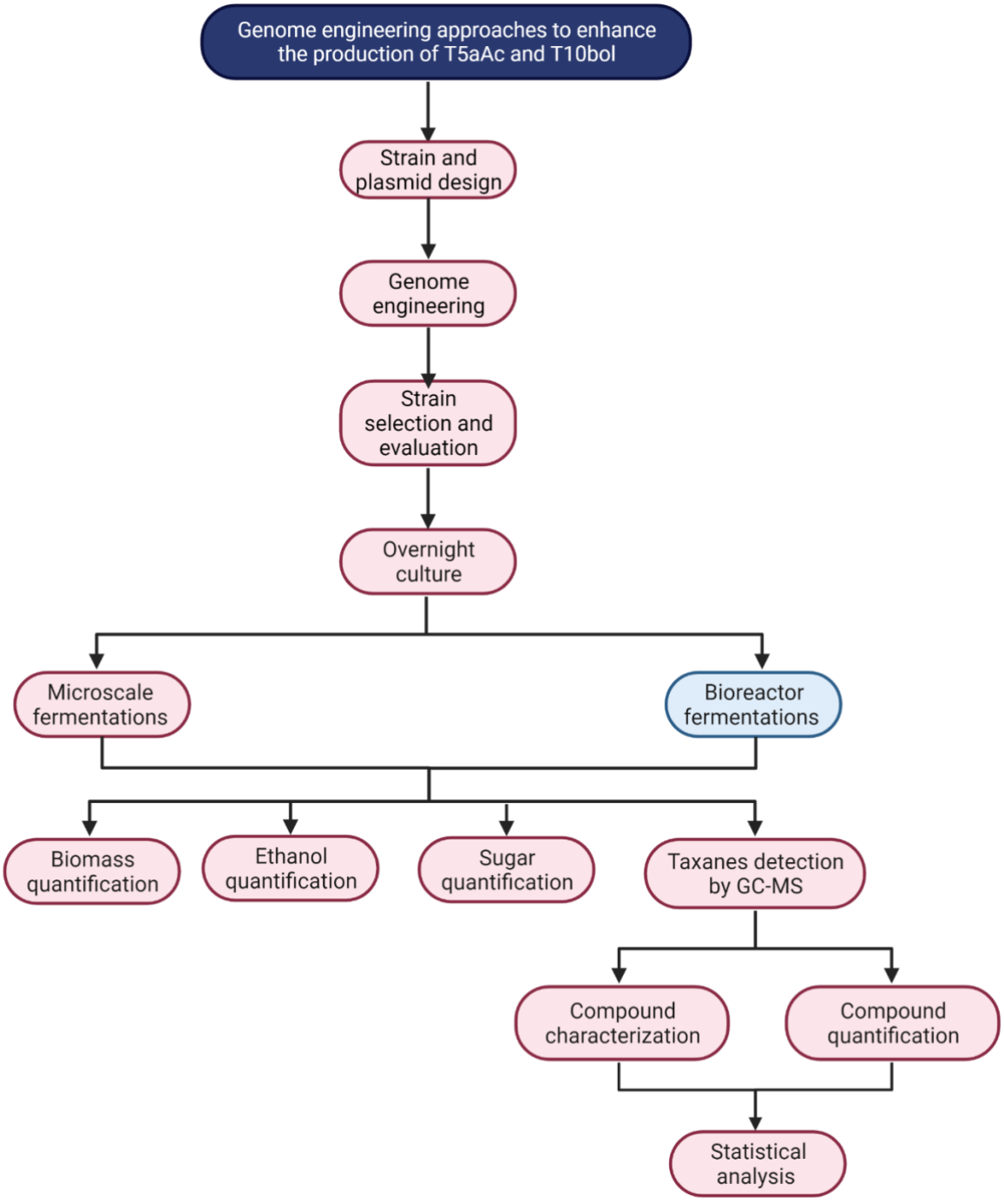
t-Test: Two-Sample Assuming Equal Variances

	Fructose	*Mel+ Fruc*
Mean	205.3741036	239.0413876
Variance	328.48648	44.39909952
Observations	2	2
Pooled Variance	186.4427897	
Hypothesized Mean Difference	0	
df	2	
	-	
t Stat	2.465671902	
P(T<=t) one-tail	0.066277424	
t Critical one-tail	2.91998558	
P(T<=t) two-tail	0.132554848	
t Critical two-tail	4.30265273	

The results are shown and discussed in section 3.3.4. All these experiments were done in triplicates (n=3) except were specified. An Alpha value of 0.5 was chosen for all statistical tests.

Finally, it is important to remark that the rationale followed in this section and for all statistical analysis has been applied in experiments of the same nature comparing different treatments and their effect on the production of natural compounds (Ai et al., 2019; Santoyo-Garcia et al., 2022).

3.2.11 Methodology summary



3.3 Results and discussion

3.3.1 Evaluation of GAL promoter and strain engineering to overcome galactose-dependant expression.

Typically, in yeast fermentations, the consumption of fermentable sugars leads to the production of ethanol. The first thing we wanted to evaluate was whether our yeast strain could produce taxanes by the sole use of ethanol. To do this we carried out fermentations in a 250 mL shake flask at one-fifth of their capacity using ethanol and a combination of ethanol + galactose in a 3:1 ratio. OD was set at 1 using the preculture and 5mL of dodecane was added to avoid air stripping of taxadiene.

The analysis by GC-MS of the preliminary experiments with taxane producer *S. cerevisiae* (strain S14) did not show any production of taxanes. This was completely expected, and it can be explained by the fact that for this strain, taxadiene genes are both controlled by the galactose-induced promoter GAL (Figures 3.2 and 3.3). Therefore, the genes are expected to only be expressed when galactose is present in the media (Peng et al., 2018). This is consistent with the results of the experiment as there was only taxadiene production during the fermentation stage when galactose was used as a carbon source, whereas when only ethanol was used there was no production of taxadiene at all (Figure 3.5).

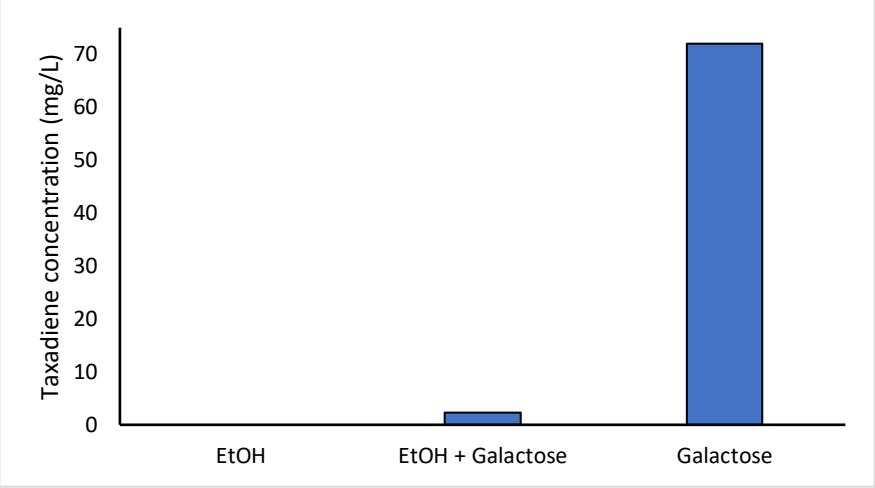


Figure 3.5 Comparison of taxadiene production in S14 when using only ethanol as a carbon source, a mixture of ethanol + galactose (in a 3:1 ratio) and only galactose. Experiments were done in 250 mL shake flasks at 30°C for 3 days at 250 rpms As the n=1 for this experiment any replication may be subject of variations (no statistical analysis was done on this data).

3.3.2 Strain engineering and preliminary evaluation of the new modified strains

Following the evaluation of taxane production on ethanol, it was determined that different strains of *S. cerevisiae* from our collection needed to be engineered to allow for the expression of taxadiene genes under the presence of any carbon source. To accomplish this, we targeted the GAL80 and MIG1p genes in all strains. It was decided to use CRISPR/Cas9 to knock out both genes to achieve a constitutive expression of taxadiene genes without the need of using galactose as an inducer. The gRNA sequence was taken from a previous research (Ai et al., 2019) and the donor DNA (~1000 bp Upstream and Downstream homology regions to GAL80) was amplified from the yeast genome. A list of all genotypes resulting from this chapter can be found in Table 3.3.

Table 3. 3. Strains after and before the transformation. The table shows all the strains engineered in this chapter and its genotype. It should also be noted that the strains are renamed after the modifications

Before transformation		**After transformation**	
Strain	**Genotype**	**Strain**	**Genotype**
S10	*mvaS* ***TASY (3x)***	S10.1	*mvaS* ***TASY (3x)*** ***ΔGAL80***
S14	*mvaS* ***TASY (3x)*** ***T5αH + CPR*** ***TAT***	*S14.1*	*mvaS* ***TASY (3x)*** ***T5αH + CPR*** ***TATΔGAL80***
KM	***S14, TAT (2x)***	EJ1	***S14, TAT (2x),*** ΔGAL80
EJ1	***S14, TAT (2x),*** ΔGAL80	EJ2	***EJ1, T10βH***

The following experiments in this chapter made use of the newly developed strains EJ1, EJ2 and S14. It is important to mention that although S14 was transformed into

S14.1 the former was consistently used across many experiments to be used as a control to compare production when GAL80 deletion is not present.

All genetic modifications were done by using the ACtivE (Malcl et al., 2022) CRISPR/Cas9 transformation method. Deletions and insertions were confirmed by Colony PCR (Figure 3.6). The list of primers for insertion and colony PCR is included in appendix 3A.1. The approach followed was to design primers amplifying a region of ~500 bp inside the GAL80 gene, strains without the deletion will show this band in the gel (bottom lanes in figure 3.6) while strains with the deletion will show no corresponding band. To have a more robust confirmation we also included another set of primers aiming for amplification of a ~1300 bp fragment of the up and down homology regions mentioned before (top lane of figure 3.6)

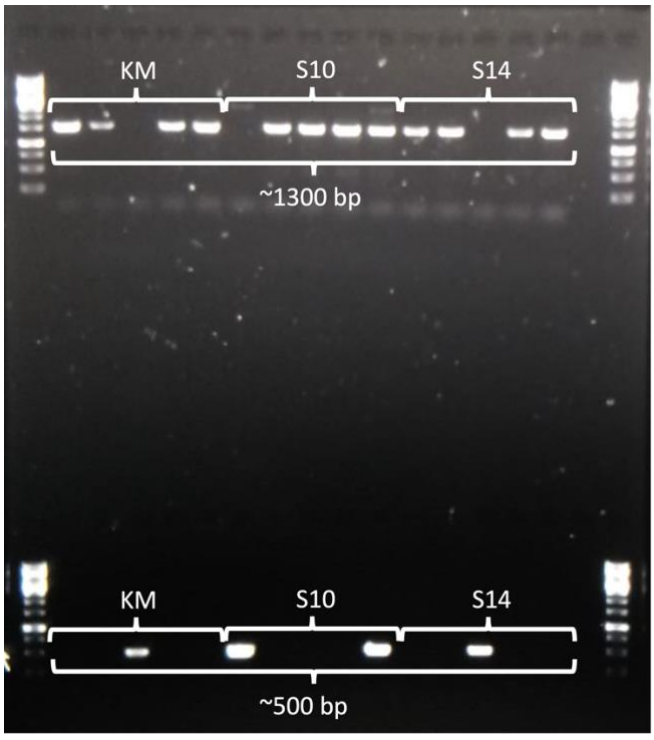


Figure 3.6 Electrophoresis gel confirming the deletion of GAL80. The top lane shows different bands corresponding to different colonies, and the strains positive for GAL80 deletion show a band of ~1300 bp. The bottom lane shows the different strains negative for GAL80 deletion with a positive band of 500 bp.

Two of the colonies of each strain were confirmed again for GAL80. CRISPR plasmid curation was achieved by multiple subcultures.

From the strains that tested positive for GAL80 deletion, two of each strain were grown in glucose and ethanol in 2% and 4% concentrations to test if the deletion was now

enabling the expression of taxane genes. These experiments were done on small scale using 24-well plates at 350 rpm at 30 degrees for 4 days in Thermomixer. GC-MS analysis confirmed that after deletion the strains were producing taxanes.

The results shown in Figure 3.7 demonstrate that the strain was now capable of producing taxanes without the need for induction by galactose.

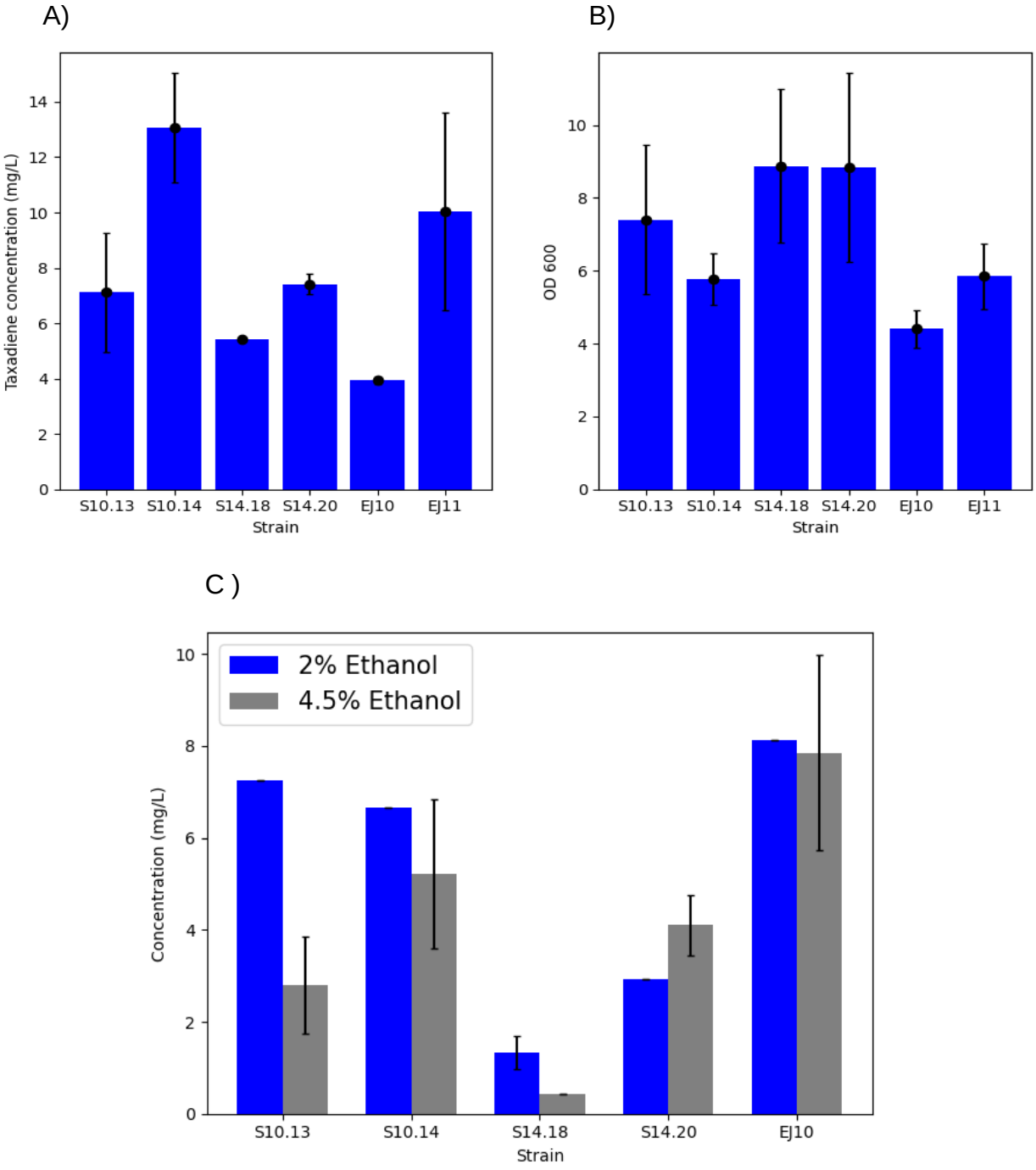


Figure 3.7. Taxadiene production in glucose [4.5%] (A) and Biomass (B). Taxadiene production in ethanol [2% and 4.5%] after GAL80 deletion, biomass was not measured in this experiment. Evaluation of different strains was carried out at 30° C for 3 days in 24 well plates. Values are mean ± standard deviation for triplicates (no statistical analysis was done on this data).

The results for the deletions were satisfactory as all strains in both carbon sources produced taxadiene. The best results from the different fermentations in glucose were as follows: S10-14 led to a production of 13 ± 2 mg/L, followed by strain S14-20 with a production of 10 ± 4 mg/L and lastly by EJ1-11 (previously KM) with 7 ± 6 mg/L. When compared to their specific biomass the best growth was registered by S14 strains followed by S10 and finally EJ strains.

The results in 2% ethanol showed the best strains production of 8 mg/L by strain EJ1-10, 7 mg/L by S10.13 and finally 3 mg/L by S14-20. For the fermentation in 4.5% ethanol, the best results were achieved once again by strains EJ1-10, S10.14 and S14-20 with 8 ±2 mg/l, 5 ± 2 mg/L and 4 ±1 mg/L respectively. The strains showed better production when grown in a lower concentration of ethanol.

Overall, better production results were achieved in glucose however, EJ1-10 and S10 strains showed better production in 2% ethanol. There was very high variability in taxadiene production by the strains so the standard deviations could not be calculated for most of the cases. All the variabilities may be due to experimental errors.

The increased results of the s10 strain in glucose and 2 % ethanol in comparison to S14 and Km are probably explained by the fact that s10 only is capable of only producing taxadiene, while the other 2 strains express the enzymes for the conversion of taxadiene into T5α-ol and T5αAc leading to a decrease quantity of this precursor. However, previous reports using S14 (referred to previously as LRS6) mentioned a 40 mg/L production using microscale instrumentation and a 4% galactose (Walls et al., 2021). This was 67% more than the best results in Figure 3.7. Another interesting thing to notice is that the author reported the presence of other paclitaxel precursors while in these experiments only taxadiene and the non-taxane by-product GGOH were detected (not shown). It was probable that the rest of the precursors were not detected due to the small amounts detected of taxadiene.

The results from these preliminary experiments suggest that despite enabling taxadiene production in other carbon sources the use of galactose still yields better results. However, it is probable that the results in glucose are lower due to the

expression of Mig1p and that ethanol concentrations might have been too high resulting in toxicity to the cells coming from an overnight glucose culture. To explore this hypothesis different experiments were carried out on different carbon sources.

To further investigate the Mig1p effect in our strain, it was decided to delete this gene using the method successfully used for GAL80. The donor DNA consisted of two fragments –located upstream and downstream of Mig1P— named “Up Homology Arm” (UPHA) and “Down Homology Arm” (DHA). DHA had a total length of 1001 bp and was amplified from the yeast genome with the primers marked with “^” in Table 3.4. However, over several attempts and multiple primer designs, we were unsuccessful in amplifying the upstream sequence of Mig1p. The list of primers used can be found in Table 3.4.

Table 3. 4 Different primers used for the amplification of Upstream and Downstream homology arms of Mig1p.

*Lowercase letters indicate the bases overlapping the CAS9 Cassette.

Name	**Sequence 5’ → 3’**
°UpHaMig1-F	TTCTTCCATCAAATTAGGGATG
°N-UpHaMig1-F	CTTCCATCAAATTAGGGATGG
°N2-UpHaMig1-F	GGGATGGAAGTGATGATGATA
°UHA-KM-F	GTGGAATGACGGTTCTGAAG
°N5-UpHaMig1-F	GAAGTACGATGTCTACTCGTCC
°N3-UpHaMig1-F	TCTTTCATACCGTCCACGTCG
°N4-UpHaMig1-F	ACGTTAGTAAGCAGCAGCAG
°UpHaMig1-R	TACGCTGACAAGTTTTTGG
°N-UpHa-R	TACGCTGACAAGTTTTTGGCG
°N2-UpHa-R-	CGCTGACAAGTTTTTGGCG
°UHA-KM-R	CTGGAAACATTACCACCCATAAG
Sg RNA for Mig1p deletion *	gcaacaccttcgggtggcgaatgggactttAGAAAAACCAAATTCGAA ATgttttagagctagaaatagcaagttaaaat
°DownHaMig1-F	GGCTATGGTAGTATGTCGTC
°DownHaMig1-F-HA	TGATTTATCTGCACCGCCAAAAACTTGTCAGCGTA GGCTATGGTAGTATGTCGTC
°DownHaMig1-R	GCTGCCACTAACCTACATTA

It is important to remark that 7 different forward primers and 4 different reverse primers (marked with “°”) were used in different combinations using different TM’s and amplifying different regions of different base pairs ranging from 100 to 600 and even inside the gene itself.

The problems amplifying this region were probably a consequence of the low GC content as the formation of secondary structures that block the polymerase along with mis-annealing and mispriming has been reported in regions with GC contents of less than 30% and more than 70% GC (Guido et al., 2016). In the case of the homology arms designed for this experiment, DHA had a GC content of 43.36% while UHA showed 32.66% a percentage very close to the threshold found in the literature.

The genomic characterization of this gene is made in chapter 5 where further discussion is made.

3.3.3 Scale up in bench scale bioreactor of strain EJ1

To further investigate taxadiene production for the ΔGAL80 strains, KM with GAL80 deletion –renamed EJ1 to avoid further confusion— was selected to be tested in glucose and scaled up in 500 ml mini bioreactors. The reason for this selection was not only that it showed a better production of taxadiene than ethanol but also because it was the most complete strain genetically for the purposes needed up to the point when this experiment was planned (see Table 3.1). The bioreactor experiment was run for 4 days and the strain was grown in a glucose concentration of ~5 g/L using the media formulation previously reported by Walls et al., (2021).

Results showed poor performance of the strain with a maximum taxadiene production of ~27 mg/L (Figure 3.8 A) in comparison to 53 mg/L reported using galactose in previous works (Walls et al., 2021), furthermore, production in other organisms like *E. coli* have shown much more higher productions in glucose with yields of up to 1 g/L in fed-batch bioreactors (Ajikumar et al., 2010). The compound with the highest production was the by-product GGOH with 60 mg/L. In this experiment there was no detection of production of the oxygenated taxane T5α-ol, however, there was some production of other uncharacterised oxygenated taxanes previously reported (Walls et al., 2021) and more importantly, we were capable of detecting the acetylated taxane with a production of 13 mg/L which represents an increase of 238 % compared to ~4 mg/l previously reported for yeast in galactose fermentations (Walls et al., 2021). However, this result falls short when compared to more recent literature in which a reported production of 135 mg/L of T5αAc was achieved by using microscale tools for

fed-batch and semi-continuous (Santoyo-Garcia et al. 2023). Finally, probably the oxygenated T5α-ol was not detected due to the complete conversion of it to the oxygenated acetylated T5αAc.

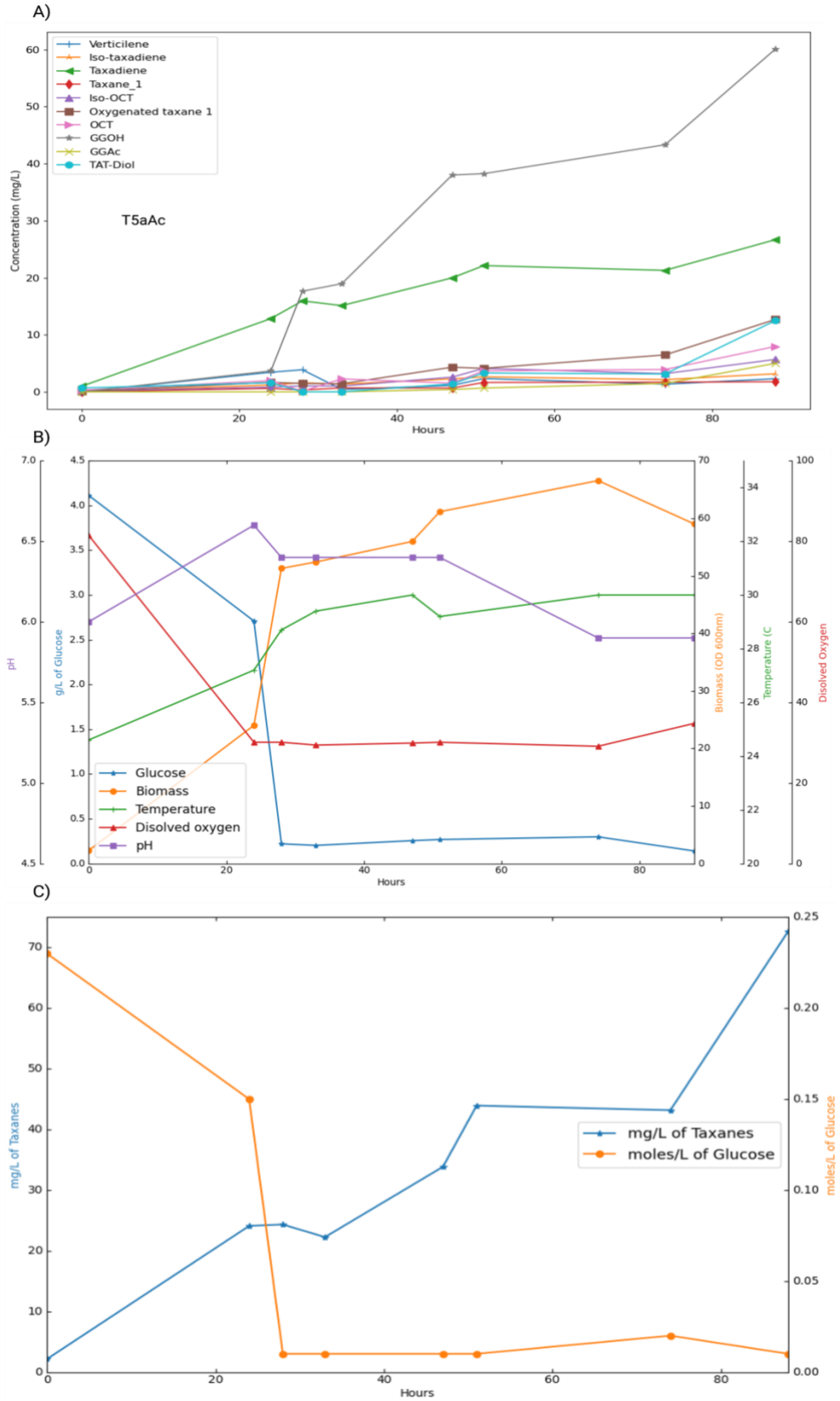


Figure 3.8 Taxanes production of EJ1 in glucose [4.5%]. The bioreactor was run for 3 days with pH controlled at 6 and temperature at 30 ° C. A) Metabolites kinetics produced by EJ1 strain. B) Kinetics of the bioreactor parameters, pH, temperature, dissolved oxygen as well as glucose and biomass. C) mg/L of taxanes Vs moles/L of carbon source. The bioreactor fermentation was done in a single repetition (no statistical analysis was done on this data).

Figure 3.8 B shows the bioprocess measurements where all the glucose was consumed after 24 hours. It is also interesting to notice that after 24 hrs, there was a significant increase in biomass with a corresponding drop in dissolved oxygen levels*** and a slight diauxic growth could be observed after this time probably due to the consumption of ethanol. However, after the sugar depletion, it was expected an increase in the production of taxanes, however, such an increase was only evident for GGOH while for the rest of the taxanes, the production rate decreased. Despite this, for all compounds production increased slightly until the end of the fermentation. Additionally, taxanes showed in Figure 3.8 A (except for GGOH and GGAc as these are not taxanes) were add up and plotted against the glucose in moles/L (Figure 3.8 C) to allow for a better comparison on how production increased as the carbon source was consumed.

***It is very important to remark that the dissolved oxygen probe was failing so despite the oxygen drop near 24hrs aeration was still optimal as the air sparger was left on for the duration of the whole experiment. Furthermore, during the experiment, there was a variation in the bioreactor temperature which may have had a negative effect on the yeast growth leading to errors and variations in the results.

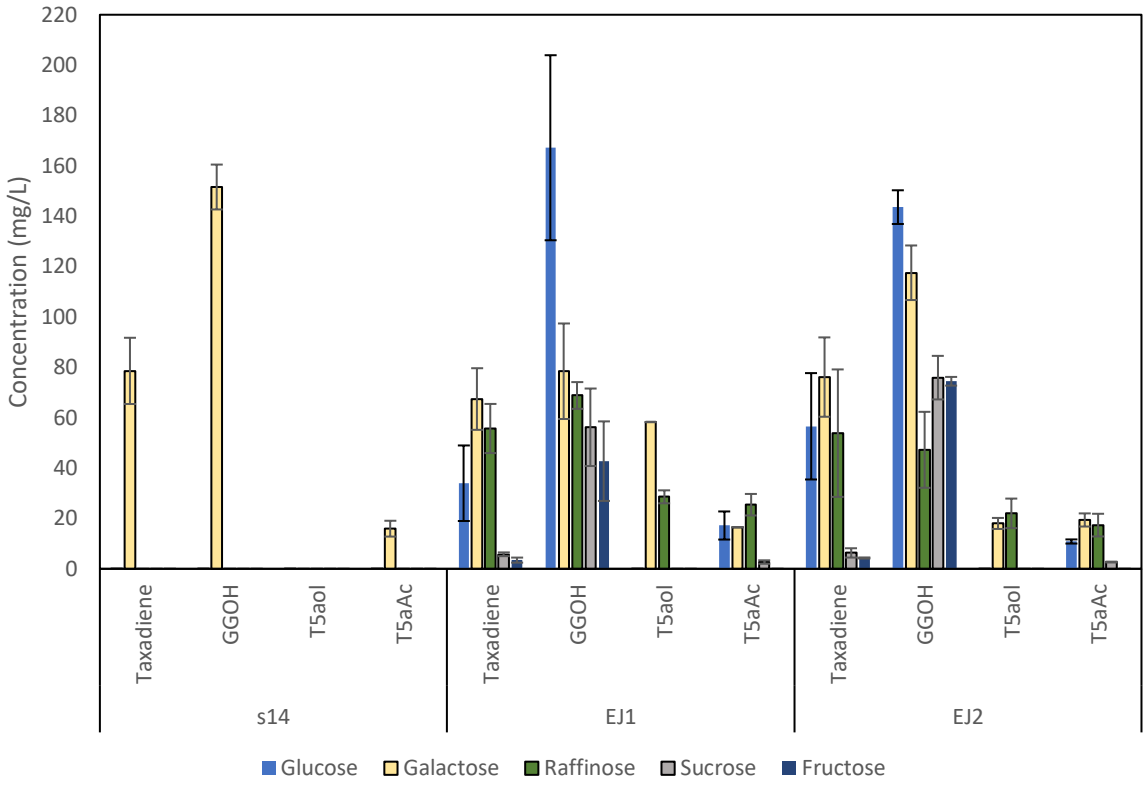
3.3.4 CRISPR knock-in of P450 10Beta hydroxylase.

To further expand paclitaxel's metabolic pathway, T10βOH was inserted in the previously described loci ARS1531 (Malcl et al., 2022). This region was recently characterised in a study previously and evidence has shown that heterologous genes inserted close to autonomous replicating sequences (ARS) tend to show higher expression rates (X.-L. Wu et al., 2019). The PCR-positive strain (named EJ2) was sent to GENEWIZ for sanger sequencing to further validate the insertion (the list of primers for insertion and colony PCR is included in appendix 3A.2). It has been shown that *S. cerevisiae* is capable of growing under different carbon sources that are both fermentable –like glucose, fructose, sucrose, galactose, maltose and raffinose— and non-fermentable— like ethanol, glycerol, acetate, oleic acid, which could have an important impact of the production of the heterologous compounds (Paulo et al., 2015).

Therefore, taxadiene production of this new strain was compared with S14 (control expressing only TASY and T5αOH and containing GAL80 gene), EJ1 (expressing

TASY, T5αOH, 2 copies of TAT and GAL80 deletion) and EJ2 (expressing TASY, T5αOH, 2 copies of TAT, T10βOH and GAL80 deletion) by performing small-scale fermentations in 24-well plates at 350 rpm at 30 degrees for 4 days in Thermomixer. 5 different carbon sources were tested: glucose, galactose, raffinose, sucrose, and fructose. The results of these experiments are shown in Figure 3.9.

A)



B)

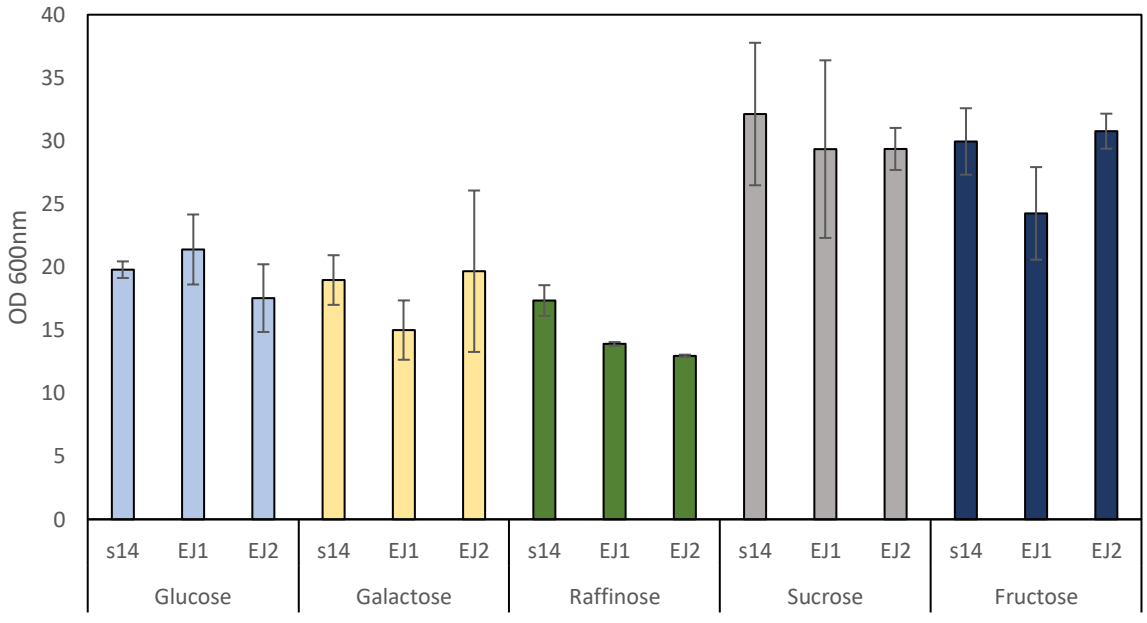


Figure 3.9 A) Production of different taxanes by strains S14, EJ1 and EJ2 in different carbon sources at 2% concentration in CSM media. B) OD at 600nm corresponding to the strains in chart panel A. The experiment was done in 24 well plates for 3 days at 30℃ and 350 rpm. n=3 for all carbon sources except for sucrose where n=6 error bars represent the standard deviation of the replicates. For chart A an ANOVA comparing taxadiene production in galactose by S14, EJ1 and EJ2 showed no statistical significance (p-value 0.6 > 0.05). At the same time, a t-student test showed no significant difference in taxadiene production in galactose between S14 and EJ1 after Gal80 deletion (p-value 0.34 > 0.5 for t-test).

There was no taxane production in S14 in any carbon source except for galactose which shows consistency with the preliminary results obtained previously in Figure 3.5 using shake flasks. This shows a clear difference between the S14 control and the rest of the strains with GAL80 deletion (Figure 3.9).

Considering taxadiene production, the best result was achieved in galactose by S14 with ~79 ±13 mg/ L which represents an increase of 225% compared to the results obtained in the same conditions by Walls et al., (2021). An ANOVA test (for S14, EJ1 and EJ2) and a t-test (S14 vs EJ1) showed no statistical difference (p-value 0.6 > 0.05 for ANOVA and p-value 0.34 > 0.5 for t-test) for taxadiene production after GAL80 deletion. Although this result may be discouraging, it is important to also consider that for strains EJ1 and EJ2 conversion of taxadiene into other precursors was better which is supported by looking at the production of other precursors in EJ1 And EJ2 and comparing it to the null production of T5α-ol and the production of ~16 ±3 mg/L of T5αAc in S14. The conversion of taxadiene into other precursors may have led to a decrease in taxadiene detected in EJ1 and EJ2 giving the impression that production decreased and thus that the effect of GAL80 deletion was non-significant.

As implied in the previous paragraph –when compared to S14— strains EJ1 and EJ2 showed a better (and similar) distribution and production of other precursors with the best results obtained once more in galactose followed by raffinose, glucose, sucrose and fructose. EJ1 achieved a superior T5α-ol production in galactose with 58 mg/L compared to 18 ± 2 mg/L achieved by EJ2, this represented a difference of almost 70%. The production of T5α-ol was good with a total amount of 58 mg/L an increase of 1350 % when compared to previous results of ~4mg/L in the exact conditions by S14 in previously published works (Walls et al., 2021), however, even higher amounts of up to 135 ±6 mg/L have been obtained by using fed-batch semi-continuous fermentation combined with resin beads recovery (Santoyo-Garcia et al. 2023). When comparing the production of T5α-ol in galactose to the production of the same compound in raffinose, it was interesting to notice that the difference in production in EJ1 was 51% higher (a difference of ~30 mg/L) while in EJ2 it was only 22% lower (a difference of only 3mg/L). Finally, the production of T5αAc was better in raffinose with 25 ±4 mg/L compared to 17 mg/L obtained in galactose, this represents of difference of nearly 55%, while in EJ2 this difference was smaller with a similar production in

galactose of 19 ±3 and of 17 mg/L in raffinose. It is important to highlight that the production of T5αAC by EJ1 in raffinose was the highest ever reported up to date by using the dodecane overlay recovery method as previous reports using this method have obtained yields of 1mg/ L (Kang Zhou et al., 2015) and ~4 mg/L (Walls et al., 2021). This represents an increase of 25-fold and 7-fold respectively. However, advanced recovery methods using resin beads combined with fed-batch cultivations in galactose have yielded a production of 95 ±4 mg/L of T5αAc (Santoyo-Garcia et al. 2023) which is still very far from the results achieved.

As mentioned before, the production in glucose was the third best among all carbon sources. The results for EJ1 and EJ2 in this carbon source were very similar to those obtained when EJ1 was scaled in the mini bioreactors and only production of taxadiene, GGOH and T5αAc was detected. This last compound resulted in a production of 17 ±6 mg/L by EJ1 which is like that obtained in galactose for the same strain and also for EJ2.

When comparing growth in the different carbon sources (Figure 3.9 B), the strains reached their highest biomass density when grown in sucrose and fructose. Sucrose is a disaccharide formed by fructose and glucose which—as mentioned before—causes repression of the genes controlled by the GAL promoter. Its hydrolysis into these monomers is carried out by the action of a β-fructosidase encoded by the gene SUC2 (Moreno & Peinado, 2012). When grown in sucrose only negligible amounts of taxadiene were detected in EJ1 and EJ2 with 6 ±1 and 6 ±2 mg/ L while in fructose the amounts were 3 ±1 mg/ L and 4 ±0.22 mg/L respectively. Furthermore, in these experiments, only GGOH was produced in good amounts that reached levels of up to ~76 ±9 mg/ L in EJ2 when grown in sucrose. As taxane production was the lowest among all the sugars (only compared to that of ethanol and even when compared to glucose) it is unclear why, despite GAL80 deletion, the strains failed to produce taxanes when growing in these carbon sources. Furthermore, a study in which the production of limonene (a different terpenoid compound) by *Yarrowia lipolytica* yeast was tested in different carbon sources (glucose, citric acid, fructose, maltose, sucrose, mannose, and galactose) also showed that production of the terpenoid was low in sucrose and fructose, especially when compared to production in other carbon

sources like glycerol or even glucose (Cheng et al., 2019). No further hypothesis of why this occurs is given by the authors.

Raffinose is a trisaccharide composed of galactose, glucose, and fructose (Moreno & Peinado, 2012). This sugar is hydrolyzed extracellularly to fructose and melibiose by the expression of the gene SUC2 (Moreno & Peinado, 2012; Paulo et al., 2015). The remaining melibiose is a disaccharide formed by glucose and galactose linked by an α-1,6-glycosidic bond. When raffinose was chosen as a carbon source for this experiment, the hypothesis was that as the different sugars were released in the media, galactose could enhance the expression of taxadiene genes, while glucose and fructose were going to be assimilated for cell growth. The results shown in Figure 3.9 of raffinose compared to the other carbon sources indicate that this may be true, yet if the enhanced effect on the expression was due to the release of the galactose molecule from raffinose, this enhanced effect should also have been present for S14 without GAL80 deletion which was not the case (Figure 3.9 A). Moreover, we later found out that to hydrolyze the α-1,6-glycosidic bond that links glucose and galactose, it is necessary the action of anα-galactosidase encoded in yeast by the gene MEL1 and other variants of the same gene (Liljeström, 1985; Naumov et al., 1996). However, not all strains of *Saccharomyces* have this gene as it has been observed that its only present in *S. uvarum* and *S. carlsbergensis* also known as (*S. pastorianus)* (Liljeström, 1985; Lodolo et al., 2008). Thus, the absence of the α-galactosidase suggested two things: the first that, likely, galactose was not “free” in the media as initially believed, thus, the enhanced expression in raffinose may not be explained by the galactose, and the second, that the only carbohydrate being used from the raffinose molecule was fructose. However, as shown before, fructose did not yield any production of taxanes.

To try to understand the contradicting results obtained, the fermentations using raffinose and fructose were repeated. Also, despite knowing that melibiose cannot be degraded and consumed by our strains we wanted to evaluate if the galactose molecule present in the melibiose was affecting the regulations of GAL genes, hence, it was decided to test fermentations with a 2% concentration of melibiose and a 1:1 mixture of 2% fructose and melibiose. To have a more robust understanding of how

our strains behave in ethanol, the fermentations on this carbon source were also repeated. The metabolites and growth results are shown in figure 3.10.

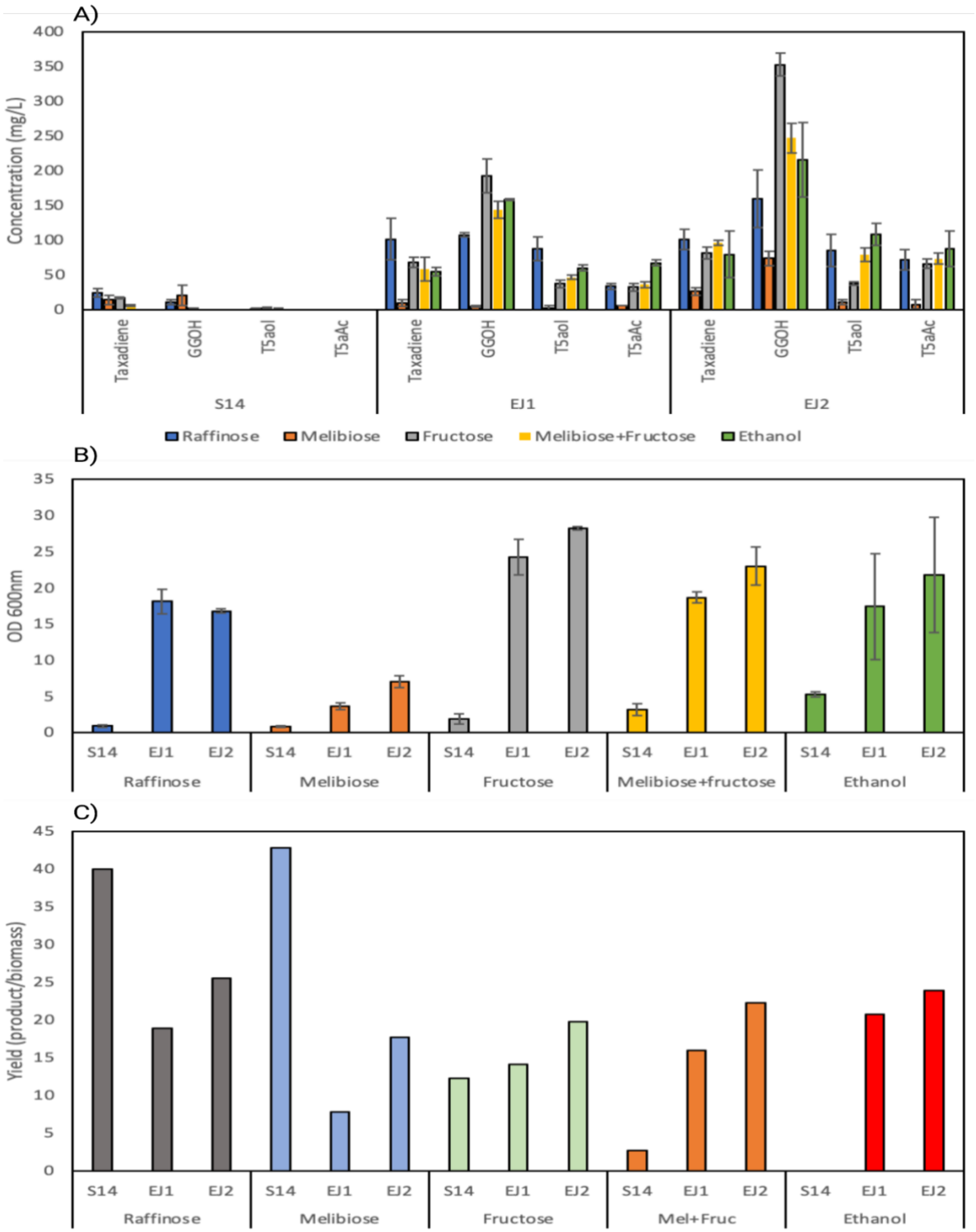


Figure 3.10. A) Production of different taxanes by strains S14, EJ1 and EJ2 in different carbon sources at 2% concentration in CSM media. B) OD at 600nm corresponding to the strains in chart panel A. The experiment was done in 24 well plates for 3 days at 30°C and 350 rpm. n=3 for all carbon sources except for ethanol where n=2 error bars represent the standard deviation of the replicates. C) Yield of product/biomass achieved by the different strains in different carbon sources (no statistical analysis was done on this data).

For S14 the first thing to note was that even though it does not have the GAL80 deletion, maximum production of 24 ±6 mg/L of taxadiene was achieved in raffinose. In the rest of the carbon sources, there was also some production of this compound with a minimum of 6 ± 1 mg/L in the 1:1 mixture of melibiose + fructose. Once again, there was no production of any taxane in ethanol for S14.

When grown in raffinose, taxadiene production in EJ1 and EJ2 strains was almost the same with a production of ~100 mg/L for both strains (Figure 3.10 A). This represented an increase of 100% compared to the previous fermentation using the same carbon source for the same strains (Figure 3.9 A). Moreover, good levels of T5α-ol and T5αAc were also achieved with ~87 mg/L for the oxygenated compound in both strains and 34 ± 4 mg/L for EJ1 and 72 ±14 mg/L in EJ2 for the acetylated one.

To better understand the production of taxanes by the different strains in the different carbon sources a comparison of the total production of taxanes was made in Table 3.5. T5αAc is also included in the table as this compound is of major importance towards the production and detection of T10β-ol.

Table 3. 5 Comparison of the total taxanes and the T5αAc production in the different treatments. A t-Test was done to analyse total taxane production in EJ1 in fructose and Mel+Fruc and the results (.05 <0.833 and 2.77> - 0.22) probed that the difference in taxane production between fructose and Mel+Fruc was not significant. The same test was done for EJ2 showing once again no significant difference (0.05 < 0.143 and 4.3 >-2.46)

CARBON SOURCE	TOTAL TAXANES		TOTAL T5αAC	
	EJ1	EJ2	EJ1	EJ2
MELIBIOSE	19 mg/ L	45.4 mg/L	6.3 mg/L	7.7 mg/L
FRUCTOSE	138 mg/L	186 mg/L	32.4 mg/L	66.3 mg/L
MEL+FRUC (1:1)	141 mg/L	249 mg/L	36.6 mg/L	73.7 mg/L
RAFFINOSE	224 mg/L	258.2 mg/L	33.9 mg/L	71.7 mg/L
ETHANOL	182.6 mg/L	275.7 mg/L	67.4 mg/L	87.5 mg/L

The total production of taxanes in EJ1 was very similar when growing in fructose and Mel+Fruc with a total of ~140 mg/L. Moreover, the production of taxanes was the lowest among the different strains when grown in melibiose. Not only production was the lowest, but also total biomass was consistently the lowest for strains growing in this disaccharide (Figure 3.10 B). This seems to confirm that α-galactosidase was not

present and melibiose was not being hydrolyzed for the release of its monosaccharide's glucose and galactose. These results may be an indication that only fructose was being utilized from raffinose.

An F- test showed that the variances of EJ2 between the samples grown in fructose and in Mel+Fruc were unequal. Following this result a t-Test: Two-Sample Assuming Unequal Variances was done to analyze total taxane production in EJ1 in fructose and Mel+Fruc. The results (.05 <0.833 and 2.77> - 0.22) probed that the difference in taxane production between fructose and Mel+Fruc was not significant. The same tests were done for EJ2 however the F-test showed that for these treatments the variances were equal. Based on this a t-Test: Two-Sample Assuming Equal Variances was done, and the results (0.05 < 0.143 and 4.3 >-2.46) probed once again that the difference in taxane production in both treatments was not significant. The results in both samples probed that adding melibiose did not have any effect on taxane production and that the difference in production in EJ1 and EJ2 in fructose and Mel+Fruc was merely due to experimental error.

When comparing the fructose results with those previously obtained there was a massive improvement. While results in Figure 3.9 A show nearly any production of taxanes, the new results in Figure 3.10 A show a production compared to that obtained in raffinose. The production of T5αAc in this carbon source was 32 ±5 mg/L for EJ1 and 66± 7 mg/L for EJ2.

Despite the good production in fructose, raffinose production was consistently higher in both strains (Figure 3.10 A and Table 3.5). Interestingly, the difference using raffinose is still significant when compared with strains growing in fructose, although it seems to be shortened when melibiose is combined with fructose. Therefore, it appears the higher production achieved by strains in raffinose may be explained by additional factors other than the presence of fructose. However, future experiments are necessary to elucidate this. It is important to remark that the literature on raffinose consumption by yeast is very contradictory as many studies failed to mention that the degradation of this sugar for consumption is only possible by the expression of the α-galactosidase previously mentioned. However, other works have focused on the

optimization of the recombinant production and characterization of α-galactosidase, especially for industrial use (Álvarez-Cao et al., 2019).

Finally, production in ethanol was very impressive as it achieved a total taxane production of 183 mg/L in EJ1 and 275 mg/L in EJ2, these results are very different to those shown in Figure 3.5. The difference in both experiments using ethanol is probably explained by the progressive improvement that our strain showed through this series of experiments. In EJ2 this was the highest production achieved within all the treatments. Furthermore, EJ2 production of T5αAC oxygenated taxanes was 87 ± 25 mg/L which is 22% more than the second-best result achieved by the same strain in raffinose and more impressively 71% more than the result previously reported in EJ1 by raffinose in Figure 3.9.

3.3.5 Scale up in bench scale bioreactor of strain EJ2

Finally, following the assessment of different carbon sources at the microscale, and the good results obtained by EJ2 in raffinose, it was decided to scale up this experiment using the 500 ml bioreactor to test if the product of *T10βH* could be detected. The results are shown in Figure 3.11.

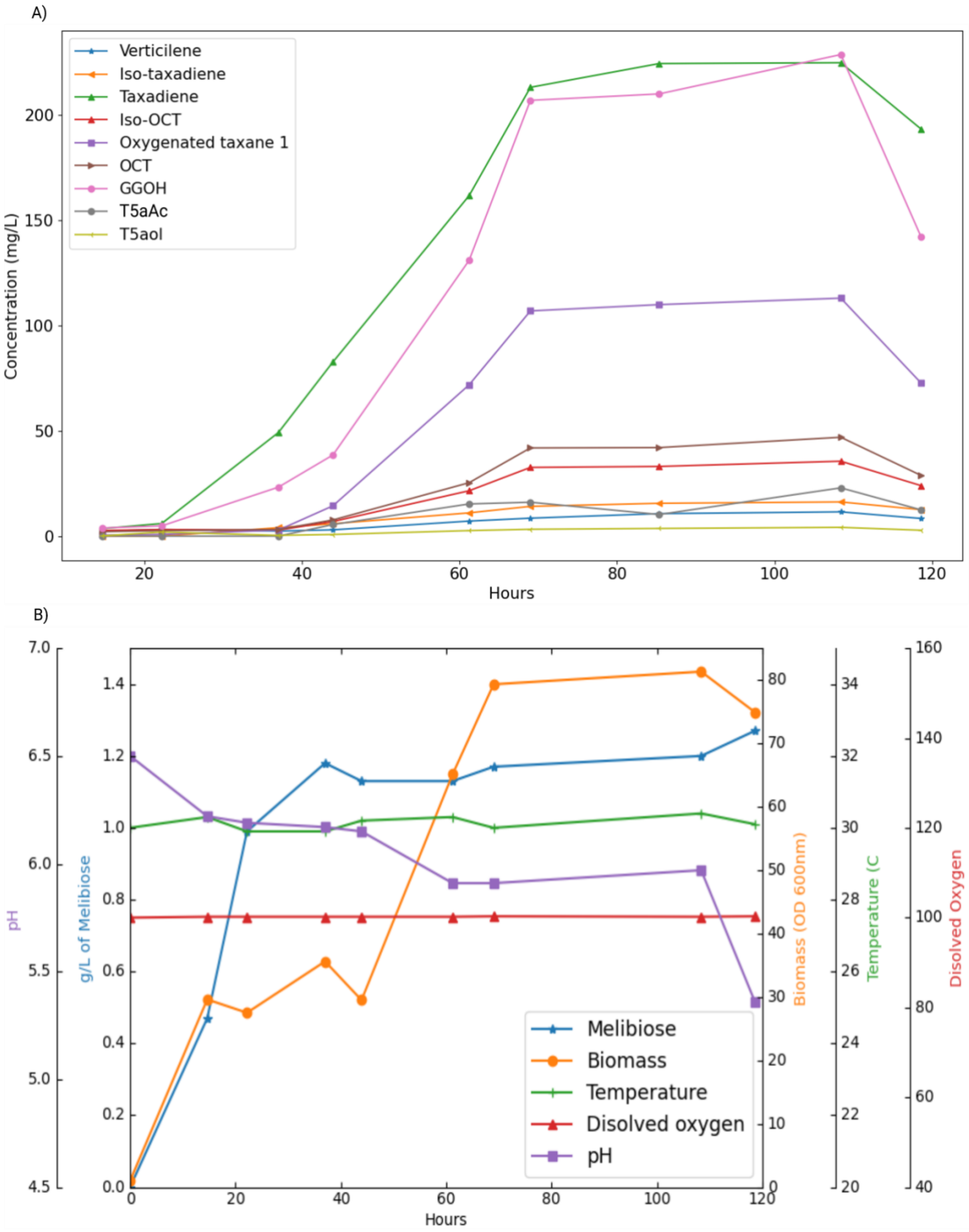


Figure 3.11. Bioreactor kinetics for Taxanes production of EJ2 in raffinose [4.5%]. The bioreactor was run for 3 days with pH controlled at 6 and temperature at 30 °C. A) Metabolites produced by EJ2 strain. B) pH, temperature, and dissolved oxygen were measured online by using the bioreactor software while glucose and biomass were measured offline using DNS assay and spectrophotometry respectively. It is important to clarify that the bioreactor fermentation was done in a single repetition (no statistical analysis was done on this data).

As can be seen in Figure 3.11 A that the first 24 hrs there is no production of taxanes. However, it is after this that taxadiene and the by-product GGOH start to be produced while the rest of the precursors start to show after 40 hrs. This may be due to the way raffinose is metabolised by yeast. As can be seen in Figure 11 B, biomass and sugar consumption correlate perfectly as the latter seems to increase over the consumption of the former. It is important to remind that raffinose is not a reducing sugar, therefore it couldn't be quantified directly by DNS assay as it was with glucose. Hence its consumption was quantified indirectly by measuring melibiose production (a reducing sugar). After 24 hrs all raffinose was degraded into melibiose and the levels of this disaccharide remain nearly the same through all the fermentation, providing even further evidence that it was not being hydrolysed for consumption.

Taxadiene production in this experiment peaks around the 70 hours mark at 225 mg/L. The production achieved represents an increase of 742% when compared to 27 mg/L obtained by EJ1 in the same bioreactor in glucose (Figure 3.8 A). Moreover, the amount obtained is 64% more than the previously obtained in strain s10 without GAL80 deletion using galactose (Walls et al., 2021) and it gets close to the 251 ±25 mg/L achieved using fed-bath and semi-continuous fermentation methods (Santoyo-Garcia et al. 2023).

Despite the good results of taxadiene achieved, conversion of these products to the desired precursor T5αAc was 3 times lower than using the microscale bioreactors (Figure 3.10). It is interesting to mention that the reduced production of oxygenated taxanes in bioreactors has been reported before (Walls et al., 2021). Also, it is important to remark that once again the product of T10βOH was not detected. Despite this, T5αAc production in this experiment was ~23 mg/L, almost 6 times more than the previous report of ~4 mg/L in the same conditions with a 4% galactose concentration (Walls et al., 2021).

Lastly, there were no major variations in the set values for the fermentation parameters. The temperature (30°C) remained steady throughout the entire experiment. However, some problems were detected with the dissolved oxygen probe at the beginning of the fermentation. To prevent oxygen shortages the air sparger was

left active during the duration of the experiment and so even though the set value was left at 70% the dissolved oxygen value was always the maximum of 100%. Finally, the medium's pH remained steady until the end of the fermentation, when it experienced a sharp drop of close to 5.4. It's important to clarify that the set value for pH was 6, however, the upper and lower threshold was ±0.5 which explains why the acidity was not correct by the end of the fermentation.

3.3.6 Limitations

For the first preliminary experiment (Figure 3.5) there were no replicates. Due to this, it was not possible to perform any statistical analysis in this experiment. However, as the results expected were of qualitative nature (presence or absence of taxadiene) there was no further need for statistical analysis although these are still encouraged for future experiments.

One of the first limitations was that due to the complexity of the experiments with the bioreactors, it was only possible to do two fermentations with each carbon source (glucose and raffinose). This made it impossible to perform a statistical analysis of the results. Repeating this experiment may lead to different results until the "n" is high enough to finally conduct statistical tests to standardise the experiment. Furthermore, at some points, we had some issues with the Dissolve Oxygen probe which appeared to be damaged. However, this was solved by turning the air supply on to prevent oxygen depletion in the culture. At the same time, for an unknown reason other than a failure in the system, the temperature in the reactor was below the setpoint of 30° C during the first 24 hrs (Figure 3.8 B) which may have had a negative effect on yeast growth and subsequently also in production.

3.4 Conclusions

The main goal of this project was to enhance the production of taxanes to facilitate the detection of new precursor compounds. To achieve this, the approach followed was to engineer our yeast strains to express taxanes in galactose and other carbon sources. By using a novel CRISPR/CAS9 method we successfully constructed strains with GAL80 deletion that could produce an almost identical amount of taxadiene as in galactose but using other carbon sources like raffinose, fructose and ethanol.

By performing fermentations on a microscale, the total taxane production in EJ1 was 183 mg/L while for EJ2 it was 275 mg/L. Furthermore, the conversion of taxadiene into T5αAc yielded a production of 67 mg/L and 88 mg/L in EJ1 and EJ2 respectively. This latter result was the best production achieved in any carbon source for this product. The strains with GAL80 deletion will greatly benefit from this as many of the sugars fermented by yeast are metabolized to ethanol.

The deletion of GAL80 proved to be successful as we managed to produce in raffinose fermentations as much products as in galactose. When compared to strain S14 without the knockout and in galactose, GAL80 strains EJ1 and EJ2 in raffinose managed an increase in T5αAc production of 2-fold and 4.5-fold respectively (Figure 3.9 A). Initially, it was believed that the results in raffinose were likely due to the presence of galactose in the raffinose molecule, however, fermentations without GAL80 deletion strains in this carbon source did not show any level of expression. Further experiments showed that taxane production in raffinose fermentations was due to the fructose monomer. Also, by growing the strains in melibiose and noticing no growth or production it was confirmed that our strains lack MEL1 gene encoding for α-galactosidase for melibiose hydrolysation and release of glucose and galactose monomers.

The results in glucose and sucrose were not as expected. Initially, glucose was expected to show the best results, but these were not better than those achieved in galactose or raffinose and in the second microscale fermentation even fructose did better. For the case of glucose and sucrose (glucose + fructose), it has been shown that the Mig1 protein has a repressive effect that can potentially affect the expression (therefore production) of genes controlled by GAL promoters. This repressor effect was most evident when doing fermentations in mini bioreactors. The maximum production of taxadiene in this experiment was only 27 mg/L a yield 21% inferior when compared to 34 mg/L achieved in microscale for the same strain. To enhance the production of glucose we aimed to delete gene Mig1p, however, our attempts were unsuccessful.

By deletion of GAL80, the constitutive production of taxanes was enabled for fermentations on certain carbon sources, especially on those that did not contain the glucose monomer as part of its molecule. The production of the acetylated product

T5αAc was the highest ever reported in batch mode, despite this, there was no detection of Taxa-4, 11-dien-5a-acetoxy-10b-ol (T10β-ol) which more likely is due to issues with the functional expression of T10βOH enzyme rather than with T5αAc supply. Future scaling up with fructose should be considered due to the lower cost compared to raffinose. Further work needs to be carried out, on the expression of T10βOH.

CHAPTER 4:

MICROBIAL CONSORTIA FOR TAXANES PRODUCTION

Chapter 4: Microbial consortia for taxanes production

4.1 Introduction

The bioproduction of heterologous compounds has been hindered by the complexity of some of these molecules. Added to this, the cell performs numerous activities such as carbon source consumption, substrate assimilation and transport, cell wall repair and maintenance, cell growth and cell division, gene expression and protein synthesis. All these activities impose a metabolic burden on the cell that can be worsened by the insertion of heterologous genes and metabolic pathways (Qian et al., 2019) (Figure 4.1). Although recent advances have been successful at engineering heterologous metabolic routes in yeast (J. Zhang et al., 2022), the reality is that these editions can be time-consuming and require a great number of resources.

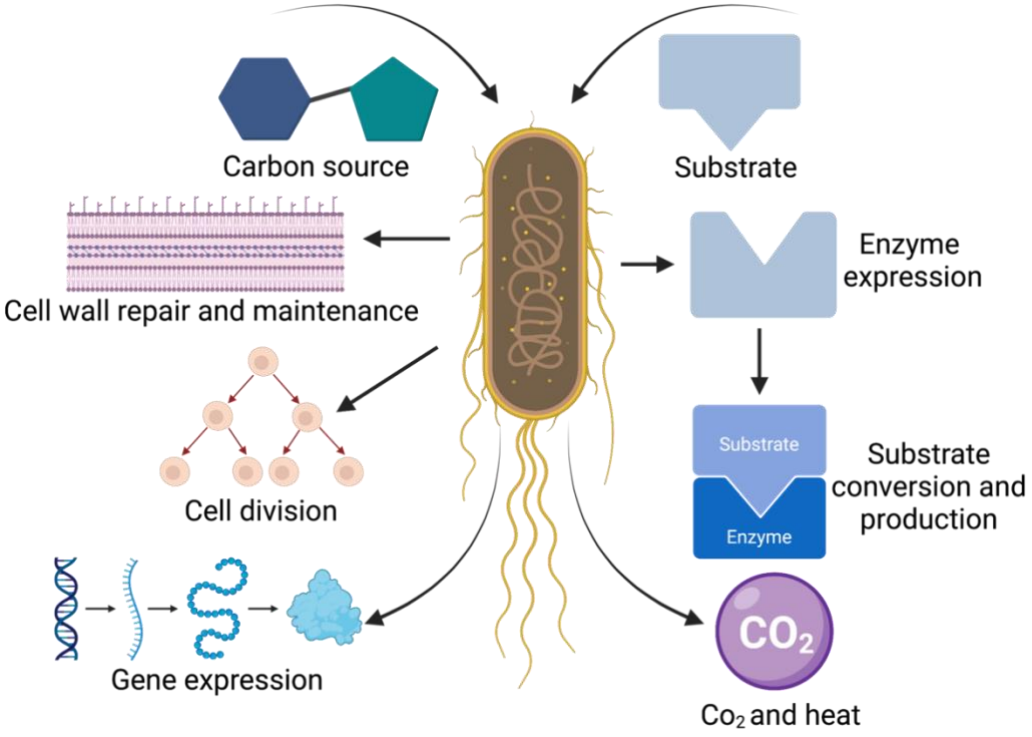


Figure 4.1. Metabolic burden inside the cell. The cell carries out various functions, including consumption of carbon sources, assimilation and transportation of substrates, maintenance and repair of the cell wall, growth and division of the cell, expression of genes, and synthesis of proteins.

An approach taken to reduce this metabolic burden is the use of microbial consortia by concepts such as strain specialization, division of labour and modularity (Tsoi et al., 2019). The multi-functionality of the whole system can be achieved by the specialization of the strains, hence allowing an improved rational organization of the

pathway, meaning that the process can be compartmentalized. Specialist strains can focus on one step of the pathway which not only reduces the metabolic burden but also proves to be useful when two processes cannot co-exist inside one cell (Tsoi et al., 2019). By creating modules (each one consisting of a microbial strain), the manipulation and engineering of the system can be facilitated, allowing the replacement and modification of the modules towards the optimisation of the consortia (Flores et al., 2019a; X. Li et al., 2022).

A challenge often mentioned is the complexity of controlling microbial consortia populations (Duncker et al., 2021; Roell et al., 2019; Shong et al., 2012). To control populations, many strategies have been proposed. Among these, we can mention the use of optogenetics (Lalwani et al., 2021) and peptide signalling (Kong et al., 2018). Interestingly even machine learning-based methods have been proposed as potential tools for controlling and predicting microbial populations (Treloar et al., 2020). Although these methods offer versatile tuning of the microbial population, their availability and adaptability in other microbial platforms hinder the possibility of their use for many researchers. As a consequence, adjusting inoculation ratios is still the preferred method for balancing microbial populations (X. Li et al., 2022) (Figure 4.2). Interestingly, recent research has proven that initial inoculation ratios not only have a crucial impact on the results of the co-culture, but also they can regulate the metabolic capacity of the co-culture by generating different patterns of species interactions (Gao et al., 2021).

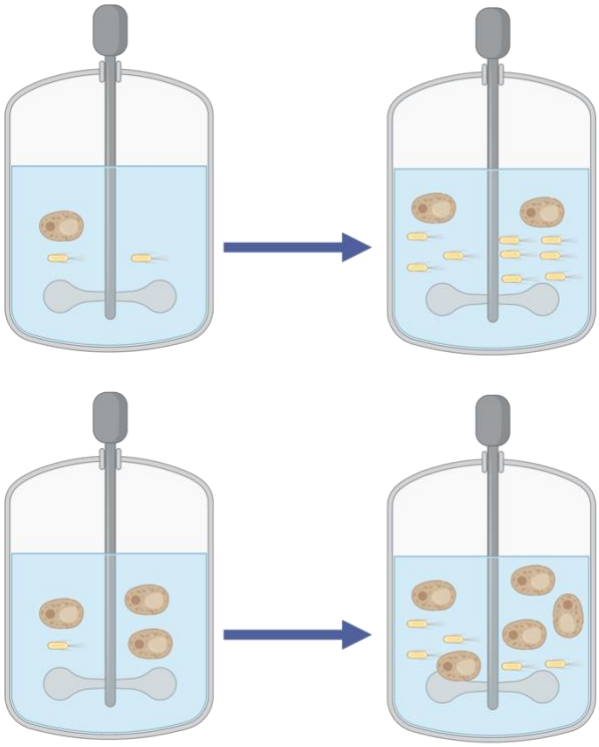


Figure 4.2. Controlling the inoculation ratio is a simple way to balance a consortium's microbial populations. The idea is that after a specific period, all populations reach a particular desired proportion of each microorganism which potentially enhances bioproduction.

There are numerous examples of microbial consortia for bioproduction. However, most of the commercially available products rely on consortia from a specific ecological niche, meaning that in many cases the consortia are not fully characterised. Examples of this are the consortia used for crop production and yield enhancement (M. Mishra et al., 2021) or the use of consortia using chicken litter as a substrate for biogas production (Marchioro et al., 2018). On the other hand, synthetic microbial communities have struggled to be scaled into commercial processes, greatly due to the big challenge of maintaining microbial consortia stability (Lalwani et al., 2021). Nevertheless, there are many examples of microbial consortia being engineered in the lab for the optimization and bioproduction of several compounds. For example, using a co-culture of *E. coli-S. cerevisiae showed* an improvement in naringenin production close to eight-fold higher when compared to yeast mono-culture achieving a total amount of 21 ± 0.4 mg/L (W. Zhang et al., 2017). In this strategy, *S. cerevisiae* was transformed with 4 genes capable of transforming L-tyrosine –from an *E. coli* strain— into naringenin. Both strains were balanced by a mutualistic approach in which the

acetate produced by *E. coli*, served as a carbon source for yeast. Other variables like fermented medium, and inoculation size and ratio were also considered.

Furthermore, another example is the consortia of *E. coli* strains to produce cadaverine, a precursor to the synthesis of diverse polymers (J. Wang et al., 2018). This work consisted of one L-lysine-producing strain that was fed with glucose and another strain capable of metabolising glycerol and converting L-lysine into cadaverine. After optimisation, the authors managed a production of 28.5 g/L. Lastly, a notable study concerning microbial consortia for taxane production is probably the one published in 2015 by Stephanopoulos's research group (Kang Zhou et al., 2015). In this work, they used an *E. coli* strain capable of producing taxadiene. This compound was later oxygenated, acetylated, and oxygenated by an *S. cerevisiae* strain expressing, TAT, T5αOH and T10βOH enzymes (the last two fused with their respective reductases). By following this approach, they managed to produce 33 mg/L of oxygenated taxanes including the monoacetylated dioxygenated taxane currently known as Taxa-4, 11-dien-5a-acetoxy-10b-ol (T10β-ol) in small concentrations of up to ~1.1 mg/L. Interestingly, the population balance was accomplished by adjusting inoculation ratios and by following a mutualistic approach in which the xylose fed to *E. coli* was metabolised into acetate to feed *S. cerevisiae* (a graphical representation of this experiment can be found in Figure 3 in the methodology section).

In this work, different combinations of *E. coli- S. cerevisiae* and *S. cerevisiae- S. cerevisiae* consortia were engineered and tested. By controlling inoculation ratios, we managed to establish a microbial consortium capable of producing different acetylated and oxygenated taxanes. By establishing an *S. cerevisiae- S. cerevisiae* consortium, we achieved a total production of up to 2235 mg/L of taxanes while production of T10β-ol was close to 26.5 mg/L, the highest ever reported up to date for this compound.

4.2 Methodology

4.2.1 Strains

All the strains in this study were the following and derived from:

- ***S. cerevisiae s10.1:*** s10, ΔGAL80.
- ***S. cerevisiae S14.1***: S14, ΔGAL80.
- ***S. cerevisiae EJ2****:* EJ1, ARS1531::Galp1-T10β
- ***S. cerevisiae EJ2-TAT****:* EJ2 but carrying a plasmid with the TAT gene.
- ***S. cerevisiae EJ2-T10β*** : EJ2 but carrying a plasmid with the T10β gene.
- ***E. coli*** LYglc1: frdBC::(Zm frg celY$_{Ec}$); ΔldhA::(Zm frg casAB$_{Ko}$); adhE::(Zm frg estZ$_{Pp}$ FRT); ΔackA::FRT; rrlE::(pdc adhA adhB FRT); ΔmgsA::FRT; LY180 ΔxylR; LYglc lacZ::cat-sacB (CmR).
- ***E. coli*** LYxyl3: frdBC::(Zm frg celY$_{Ec}$); ΔldhA::(Zm frg casAB$_{Ko}$); adhE::(Zm frg estZ$_{Pp}$ FRT); ΔackA::FRT; rrlE::(pdc adhA adhB FRT); ΔmgsA::FRT; ΔptsI; ΔptsG; ΔgalP; glk::kanR (KanR); xylR::xylR* (KanR).

E. Coli strains LYglc1 and LYxyl3 were kindly provided by Professor Xuang Wang from Arizona State. More information about these strains can be found in the corresponding paper (Flores et al., 2019a).

4.2.2 Oligonucleotides, Reagents, and Plasmids

All the chemicals were sourced from Sigma-Aldrich unless otherwise stated. All the primer design was done by using Benchling. Designed oligos were ordered as standard DNA oligos from Integrated DNA Tecnologies (IDT). Depending on the use of the parts (PCR or Plasmid assembly) the fragment size of the oligos varied in length from 17 bp to 60 bp. All primers used in the study are listed in the Appendix section from 4A to 4F. The primer TM calculation was done by using the ThermoFisher Scientific TM calculator which can be found online at: https://shorturl.at/rtJTY.

The synthetic parts amplification for the CRISPR/CAS9 plasmid assembly was done by PCR by using Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific F548S). For colony PCR, DreamTaq Green PCR Master Mix (Thermo Fisher Scientific EP0711) was used. The PCR protocol was performed as follows:

1. In a PCR Eppendorf tube 10 mL of Master Mix, 2 mL of template DNA, 1 mL of forward primer, 1 mL of reverse primer and 6 mL of distilled water were added. Under some circumstances—such as low DNA concentration or poor amplification—the total volume was 40 mL ad each of the reagents was adjusted accordingly.
2. The PCR tube was placed inside the thermocycler (ProFlex™ from Fisher Scientific) and the thermocycling parameters were adjusted according to the table shown below (Table 4.1).

Table 4.1 Thermocycling conditions used for PCR.

Step	Temperature in C°	Time	Number of cycles
Initial denaturation	95	1 min	1
Denaturation	95	30 s	30
Annealing	Primer specific	30 s	
Extension	72	1 min	
Storage	4	∞	1

3. To verify the DNA amplification 1mL of the PCR product was loaded onto a 1% agarose gel dyed with 0.75% of SYBR-safe DNA gel stain.
4. The gel was run in 1X TAE buffer for 30 minutes at 100 V.
5. The results were visualised by using a UV light transilluminator.

Following PCR in some cases it was necessary to purify the product and, in such cases, QIAquick PCR Purification Kit was used to purify and clean up PCR-generated fragments while GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) was used for plasmid extraction, on both cases the methodology followed was that indicated by the manufacturer.

After purification, it was necessary to assess not only the DNA quantity but also its quality. This was done as follows.

1. 2 µL of the DNA recovered was analysed with a NanodropTM 2000c spectrophotometer.

2. Absorbance was measured at 260 nm and the DNA concentration is calculated automatically by the equipment. 260/280 ratios of ~1.8-2 and 260/230 ratios of ~2-2.2 are expected in pure DNA samples.
3. When necessary, the DNA sample was diluted in distilled sterile water to achieve the desired concentration.

To look for conserved and homology regions in T5αOH and T10βOH, the sequences were downloaded in .fasta format from the LeoRios Group Benchling repository and a pair-wise alignment was performed by using MEGA11®.

4.2.3 Media

All the chemicals were sourced from Sigma-Aldrich unless otherwise stated. Different media were used for this research and all the ingredients are listed below in Table 4.2. The quantities included are for the preparation of 1L of media.

Table 4.2 Ingredients used for the preparation of different media.

Ingredient	Quantity
Yeast extract	10 g
Peptone	20 g
Carbon source (glucose, galactose, fructose, raffinose, melibiose, sucrose and ethanol)	200 g
Agar (optional)	20 g
Yeast nitrogen base	1.7 g
Ammonium sulphate	5 g
CSM-URA	Detailed by the manufacturer
LB media	Detailed by the manufacturer
Ampicillin	100 ug/mL (prepare from a 100 mg/ mL stock solution)

1. To prepare 1 L of YP media with a 2% carbon source concentration yeast extract and peptone were added in a Duran bottle according to Table 2.2. Agar was added when Petri dishes were needed.

2. 800 mL of distilled water was added to the bottle.
3. The chosen carbon source was added to the media in a different bottle and filled with 200 mL of distilled water.
4. Both bottles were autoclaved for 20 min
5. In a sterile environment the bottle containing the carbon source was added to the bottle of media to achieve a final concentration of 2%
6. OPTIONAL: if Petri dishes (15 mm x 100mm) were prepared, add ~20 mL to each and leave them open until the media is solidified. For LB plates it is necessary to add ampicillin at the specified concentration indicated in Table 4.2 is achieved.
7. Store Petri dishes in a cabinet or at 30 °C in an incubator. Petri dishes with antibiotic need to be stored at 4°C for no longer than 2 weeks.

4.2.4 In-Vivo Assembly and Yeast Plasmid Transformation and Selection

The strain EJ2 was transformed with a plasmid carrying the TAT gene and the T10BOH gene. The following protocol details all the steps followed for in vivo plasmid assembly (van Leeuwen et al., 2015; Van Leeuwen et al., 2015) and plasmid transformation. With some variations, the protocol followed was the one described by (Gietz & Schiestl, 2007). Everything was done in sterile conditions.

1. Overnight cultures were typically made by adding 4.5 mL of YP media in a 50 mL falcon tube and 0.5 mL of 2% glucose for a night at 30°C in a shaking incubator at 200 rpm.
2. Between 80-100 µL from the overnight culture were added to a 50 mL falcon tube containing fresh YPD media and left inside a shaking incubator at 200 rpm and 30°C for 3 to 4 hours.
3. After this, the cells were pelleted by centrifugation at 3000 g for 5 minutes and the supernatant was discarded to be later resuspended with distilled water by gently pipetting up and down the pellet. This procedure was repeated at least 3 times to wash off any residual media.
4. During the final step, the resuspended pellet was transferred to a 1.5 mL Eppendorf tube where it was once more centrifuged, and the supernatant was removed.

5. To prepare the transformation mix, 240 µL PEG (50%(w/v)), 36 µL 1.0 M lithium acetate (LiAc) and 50 µL single-stranded (previously denatured at 100 °C for 5 min) carrier DNA (2.0 mg/mL) (herring sperm DNA, Promega) were mixed in a 2mL tube added to the pelleted cells.
6. Approximately ~750 fmol – ~1500 fmol from each donor DNA previously quantified in the spectrophotometer were added to the transformation mix.
7. The mixture was briefly vortexed and the solution in the tube was incubated for ~45 minutes at 42ºC using Thermomixer C (this is known as the heat shock method).
8. The cells were centrifuged for 1 minute at 3000 g and the supernatant was discarded.
9. 1mL of sterile water was added to resuspend the pellet and from this, only 200 µL were plated into a petri dish with CSM -Ura plates for 2-3 days at 30 ºC.
10. The growing colonies were screened by using the PCR methodology described in section 4.2.2.

It is important to mention that the intention is that all plasmids resulting from this research will be available through the Addgene Vector Database.

4.2.5 Preparation of competent *E. coli* cells

For *E. coli* overnight cultures 4.5 mL of LB media (prepared as per instructions of the fabricant) from Sigma-Aldrich was used in combination with 0.5 mL of a 2% glucose or xylose solution all mixed in a 50 mL falcon tube incubated all night at 37°C in a shaking incubator at 200 rpm.

1. 250 mL of fresh LB media was inoculated with 250 µL of the overnight culture and incubated for 3 hours to reach an OD of approximately 0.4.
2. The cells were put in ice for 10 minutes and swirled to cool evenly all the media.
3. The cells were pelleted by a 3500 g centrifugation cycle and the supernatant was discarded.
4. The cells were resuspended in 10 mL cold 100 mM $CaCl_2$ solution to then incubate on ice for 20 minutes.

5. The cells were once again centrifuged at 3500 g for 5 minutes and after discarding the supernatant, 5 mL of cold 85mM $CaCl_2$/ 15% (v/v) glycerol solution was added to resuspend the cells.
6. Aliquots of the cells in 2 mL Eppendorf tubes were stored at -80°C.

All the procedure was done in sterility conditions.

4.2.6 Plasmid Assembly, Bacteria Transformation and Colony Selection

Both ***E. coli*** LYglc1 and ***E. coli*** LYxyl3 were transformed with plasmids carrying the codon-optimized genes for expressing T5αOH, TAT and T10BOH genes. Plasmid assembly for *E. coli* was done in vitro using NEBuilder® HiFi DNA Assembly by New England Biolabs® and following the instructions detailed by the manufacturer. Plasmid transformation was done by the heat shock method.

1. The first step was to defrost an aliquot of the competent cells.
2. 50 µL of the cells were separated into a different 1.5 mL Eppendorf tube and mixed with 2 - 5 µL of the assembly and incubated on ice for 30 min.
3. Then the cells were incubated for 45 seconds at 42°C in a water batch (step can be done in Thermomixer C) and immediately after this, the tube was placed on ice for 3 minutes.
4. To heal the cells from the heat shock and reconstitute the cell wall, 500 µL of LB media –previously warmed at 37°C—were added to the tube and incubated at 37ºC, 300 rpm for 1 hour in the Thermomixer C.
5. Finally, 50-100 µL of the cells were spread onto LB agar plates with ampicillin (100 µg/mL final concentration) and incubated at 37°C overnight.

All the procedure was done in sterile conditions. The colonies growing in the plate with antibiotic were selected as positive colonies. It is important to mention that the intention is that all plasmids resulting from this research will be available through the Addgene Vector Database.

4.2.7 Microscale fermentations for the production of acetylated and oxygenated taxanes

To evaluate the best consortia and conditions to produce taxanes, *S. cerevisiae* strains s10.1, S14, EJ2, EJ2-TAT and EJ2-T10B and *E. coli* strains LYglc1 and LYxyl3 were tested in different co-cultures in Axygen's polypropylene 24 V-shaped well plates in a microplate's Thermomixers C by Eppendorf.

1. Before the fermentation all the strains mentioned were grown in overnight cultures for at least 16 hours in 50 mL Falcon tubes containing 5 mL YPD for yeast and 4.5 of LB + 0.5 of 20% glucose/xylose/raffinose for *E. coli* and were incubated at 30ºC and 37ºC for yeast and *E. coli* respectively inside shaking incubators at 200 rpms.
2. The next day 160 µL (final 2% concentration) of galactose/ glucose/xylose/raffinose was added to the well of a sterile Axygen's polypropylene 24 V-shaped well plate.
3. For this experiment, different inoculation ratios and initial ODs were tested for the different consortia. To set up the different inoculation ratios the OD was measured at 600nm using nanodrop to quantify biomass and based on this an appropriate volume of inoculum was added to the well. In this manner we also could achieve different combinations and proportions between the organisms integrating the consortia.
4. The remaining volume was completed with YP until a volume of 1600 µL was reached. It is also important to remark that for the *S. cerevisiae- E. coli* consortia YP media was preferred to favoured yeast growth over bacteria.
5. Finally, 400 µL of dodecane were added to the well. To preserve sterility the plates were covered with Thermo Scientific's sterile gas-permeable adhesive films (AB0718).
6. The plates were taken to Eppendorf's Thermo mixers C –with a plate holder adapter— and were agitated at 350 RPMs at 30ºC for 3 days. The preparation of the plates was done in a laminar airflow cabinet to preserve sterility. All samples were done triplicates unless specified.
7. Once the fermentation finished, the media was collected in 2mL Eppendorf tubes to measure the final OD.

8. The dodecane layer was separated from the culture by centrifugation at 1000 rpm for 5 minutes and 50 µL was transferred to GC-MS vials for further analysis.

4.2.8 Biomass, sugars, and ethanol quantification

4.2.8.1 Biomass quantification

To quantify biomass, dilutions of the cultures were done in cuvettes and the sample was measured at 600 nm using a NanodropTM 2000c spectrophotometer. The result was later multiplied by 1000 and divided by the number of µL used for the dilution to finally obtain the original OD.

Alternatively, for microbial consortia, photographs of the samples using a bright field microscope (Leica's model DM IRB) were taken to visualise and count the yeast and bacteria cells in the community. The method for counting consisted in dividing the photograph of the sample into 9 equal parts and counting the number of bacterial and yeast cells in just one of the squares to later multiply the result by 9. However, the implementation of this method was only to have a rough estimation of the consortia population dynamics and as such it needs statical validation to prove its reproducibility.

4.2.8.2 Sugars quantification

Reducing sugars were measured by using the dinitrosalicylic acid (DNS) method described by Miller, 1959.

1. The first step was to centrifuge the sample and separate 500 μL into a new Eppendorf tube to which the same amount of DNS reagent (3,5–dinitrosalicylic acid, 1 %; sodium sulphite, 0.05 %; sodium hydroxide, 1 %) was added.
2. The solution was mixed by inversion of the tube several times and incubated for 5 minutes at 100 $^{\circ}$C.
3. While the solution was still warm, 167 μL of Rochelle salt solution was added (40 % potassium sodium tartrate tetrahydrate).
4. The tubes were allowed to cool, and absorbance was measured at 575nm using the nanodrop previously mentioned in this section.
5. Dilutions were made for samples whose measured values were below or above the range of the spectrophotometer ~0.15-1.2 g/L.

6. The reducing sugar concentration was calculated by using standard curves previously prepared with the sugar of interest.

4.2.8.3 Ethanol quantification

For measuring ethanol in the samples, it was necessary to remove the contaminants by the Carrez clarification method.

1. The samples were centrifuged and 100mL of the supernatant was diluted into a new tube with 700 µL distilled water.
2. Next, 50 μL of Carrez reagent I (potassium hexacyanoferrate, 3.6 %), 50 μL of Carrez reagent II (zinc sulphate heptahydrate, 7.2 %) and 100 μL of 100mM NaOH were added to each sample tube to add up a total of 1000mL.
3. Finally, samples were briefly vortexed and centrifuged at 3000 G for 5 minutes.
4. Once the sample was ready, ethanol quantification was done per instructions of the Ethanol Assay Kit by Megazyme (K-ETOH).

4.2.9 Taxanes detection by Gas Chromatography-Mass Spectrometry (GC-MS)

Taxane identification and quantification were done by GC-MS. Out of the 50µL of dodecane, only 1µL was injected into the TRACE 1300 Gas Chromatograph (Thermo Fisher Scientific, UK) which was coupled to an ISQ LT single quadrupole mass spectrometer by the same company. The column used was Thermos Scientific TG-5MS with 30m x 0.25mm x 0.25um dimensions. Settings were as follows: Injector temp 250^{o}C, split less injection, split flow 10ml/min, split ration 33.3, carrier flow 2ml/min. The initial GC temperature was 120^{o}C and held for 1 minute, then increased up to 20^{o}C/min until a temperature of 300^{o}C was achieved, and the sample was held for 3 minutes.

To identify and quantify the production of taxanes standard concentrations of taxadiene and GGOH were related to the peak area for each product and concentrations were calculated. To distinguish a peak from the background noise, the mass spectrum for each peak was analysed. Also, only peaks with three times the base high were considered as true compound peaks. The pure standards of taxadiene and GGOH were previously obtained by a donation from the Baran Lab (The Scripps

Research Institute, USA) and Sigma Aldrich, respectively. Xcalibur™ Software from Thermo Fisher Scientific was used to analyse GC-MS metabolites and measure peak areas for identification and quantification.

4.2.10 Statistic analysis

All the statistical analyses were done by using the Excel Analysis ToolPak.

Two separate ANOVAs were done to evaluate if the production of total acetylated taxanes between the monoculture and the consortia (of EJ2 – *E. coli* T10βOH and TAT) was statistically significant or not (Figure 4.11). Both ANOVAs are shown below.

Statistical analysis for:
Hypothesis null is that there is no significant increase in T5αAc production between the consortia and EJ2 with LYxyl3 strains in raffinose.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Total Acetylated control	2	50.28280 28	25.14140 14	2.316924 96
EJ2 Raff xyl	2	56.29299 63	28.14649 81	33.04090 57

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	9.030606 39	1	9.030606 39	0.510812 24	0.54895 12	18.51282 05
Within Groups	35.35783 07	2	17.67891 53			
Total	44.38843 7	3				

Statistical analysis for:

Hypothesis null is that there is no significant increase in T5αAc production between the consortia and EJ2 with LYglc1 strains in raffinose.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Total Acetylated control	2	50.2828028	25.1414014	2.31692496
EJ2 Raff gluc	2	71.1266898	35.5633449	21.1378497

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	108.616906	1	108.616906	9.26181618	0.09313245	18.5128205
Within Groups	23.4547747	2	11.7273873			
Total	132.071678	3				

Another ANOVA was performed to confirm or deny if changing inoculation ratios influenced the production of T10β-ol. A one-way ANOVA of the triplicates confirmed the rejection of the null hypothesis highlighting that the inoculation ratio of both strains did significantly affect T10β-ol production (Figure 4.13).

Statistical analysis for:
Hypothesis null is that the ratio does not affect production levels of T10β-ol.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
1:1	4	56.8380386	14.2095097	31.1884071
3:1	4	105.646416	26.4116039	22.6974981

ANOVA

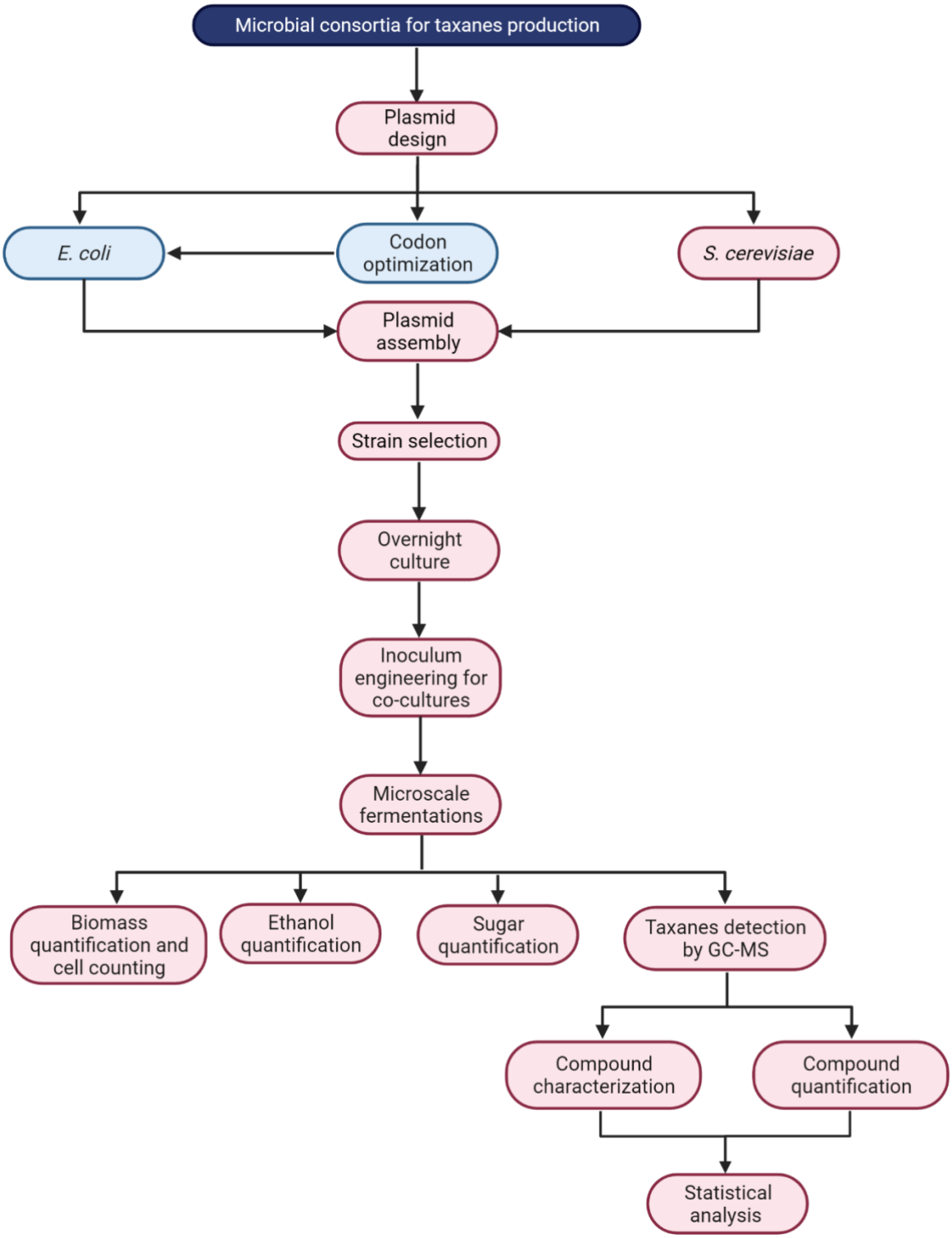
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	297.782207	1	297.782207	11.0523227	0.01591501	5.98737761
Within Groups	161.657715	6	26.9429526			
Total	459.439923	7				

Hypothesis null is rejected while alternative hypothesis is accepted: there is significant difference between treatments.

All the experiments were done in triplicates (n=3) except were specified and an Alpha value of 0.5 was chosen for all statistical tests.

Finally, it is important to remark that the rationale followed in this section and for all statistical analysis has been applied in experiments of the same nature comparing different treatments and their effect on the production of natural compounds (Ai et al., 2019; Santoyo-Garcia et al., 2022).

4.2.11 Methodology summary



4.3 Results and discussion

4.3.1 *E. coli- S. cerevisiae* consortia based in ethanol production.

The initial idea was to create a consortium using similar interactions as the ones described by, Zhou et al., (2015). In this consortium, *E. coli* would produce taxadiene to be oxygenated by *S. cerevisiae* expressing the corresponding P450. To control populations both would grow on glucose while ethanol production would inhibit *E. coli's* growth (Figure 4.3 A). In different consortium the same division of labour for the production of oxygenated taxanes was applied, however, instead of using glucose to feed both strains xylose was fed to *E. coli*. This sugar was latter converted to acetate by the bacteria and was used as carbon source by *S. cerevisiae* (Figure 4.3 B). In these systems, the population was done by controlled by adjusting the inoculation ratios and by the use of different carbon sources that may benefit the organisms differently and may favour the production of different metabolites (such as ethanol) that can have an inhibitory effect on selected microorganisms.

This idea was expanded in this work by using 2 different LY *E. coli* strains capable of metabolizing –orthogonally— glucose and xylose and producing enhanced amounts of ethanol (Flores et al., 2019a). A new co-culture system was engineered in which the ethanologenic strains would produce ethanol which would be used by yeast as a carbon source while inhibiting E. *coli's* growth. A diagram of this approach can be found in Figures 4.3 A and B.

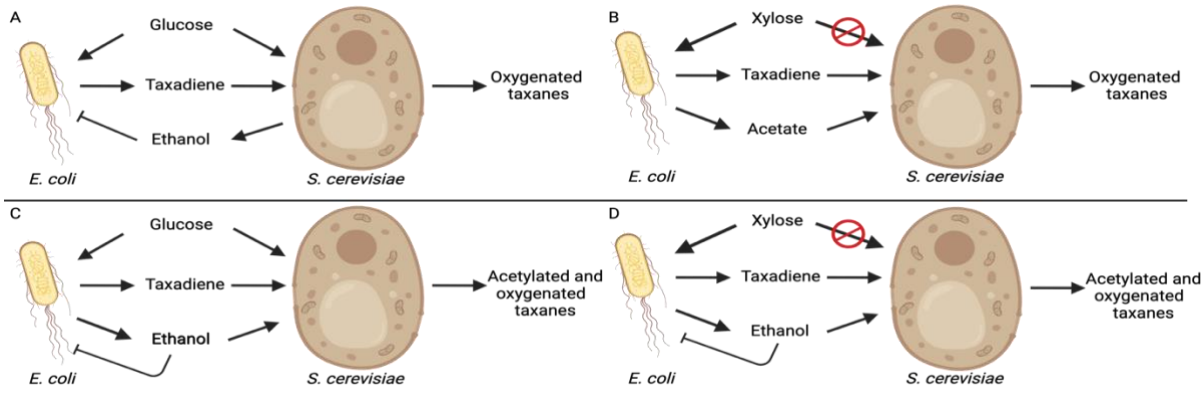


Figure 4.3. Different microbial consortia of E. coli- S. cerevisiae. A) *Microbial consortia suggested by* (Kang Zhou et al., 2015). In the first image, both organisms grow on glucose, however, *E. coli* is inhibited by ethanol. Taxadiene is produced by *E. coli* while the product is oxygenated by yeast, B) A similar approach for taxanes production, however, xylose (not fermentable by yeast) was consumed by *E. coli* and produced acetate that was used as a carbon source by yeast. C and D) Ethanologenic strains produced ethanol that served as a carbon source for yeast while causing inhibition to *E. coli*. The metabolic pathway was expanded with T5αH, TAT and T10βH enzymes.

The first preliminary experiments involved the evaluation of ethanol production of ethanologenic *E. coli LYxyl3, E. coli* LYglc1 *and P. putida KT2440* (this strain grows in the presence of oxygen, while its metabolic activity is under anoxic conditions). The fermentation for *P. putida KT2440* was done according to the methodology described by (Nikel & de Lorenzo, 2014). Both *E. coli* strains were also tested for ethanol production as suggested in the literature (Flores et al., 2019a). The results of the monocultures are shown in Figure 4.4.

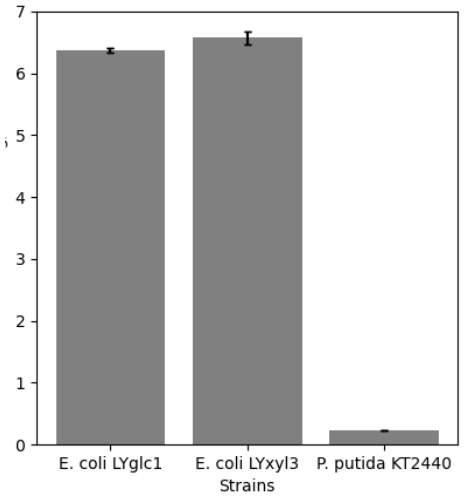


Figure 4.4. Ethanol production by 3 different ethanol-producing strains (*E. coli* LYglc1, *E. coli LY*xyl3 and *P. putida* KT2440). All strains were grown for 48 hours at 37° C using LB media. For *E. coli* LYglc1 and *P. putida* KT2440, 2% glucose was used as feedstock while for *E. coli* LYxyl3 2% xylose was used. Error bars represent the standard deviation of the mean for the triplicates (no statistical analysis was done on this data).

Results showed to be inconsistent with those previously reported reaching ethanol concentrations up to 24 mM (1.1 g/L) (Nikel & de Lorenzo, 2014). Results in our experiments showed an average concentration of 0.024 g/L, which –compared to the literature (Nikel & de Lorenzo, 2014)— was notably lower than that expected. These results may be explained by a discrepancy in our experimental methodologies such as the use of a different media and vessel configuration. Furthermore, *P. putida KT2440* only produced ethanol in anoxic conditions which is undesirable in a consortium with *S. cerevisiae* as the ethanol consumption in this strain happens in the presence of oxygen (Gasmi et al., 2014). Also, for simplicity, LB media was used instead of M9 minimum media. It is important to mention that, as ethanol will be used as the carbon source, sufficient concentrations are needed to prevent this from being a growth-limiting factor for *S. cerevisiae.* Also, ethanol was necessary to control the population of *E. coli* as it is toxic at high concentrations. Due to the lower ethanol

production using *P. putida KT2440* as shown in Figure 4.4, it was decided to work with *E. coli* LYxyl3 and *E. coli* LYgluc1 for the consortium.

After assessing ethanol production, *E. coli* LYxyl3 was grown with *S. cerevisiae* S14 to evaluate the production of taxanes and measure microbial consortia balance. This experiment was done in a 250 mL shake flask using YP and 2% xylose as carbon sources. This experiment was performed with a yeast strain dependent on galactose for the expression of the mevalonate pathway and taxane genes (the GAL80 deleted strain described in the previous chapter was not yet developed). The results are shown below in Figure 4.5.

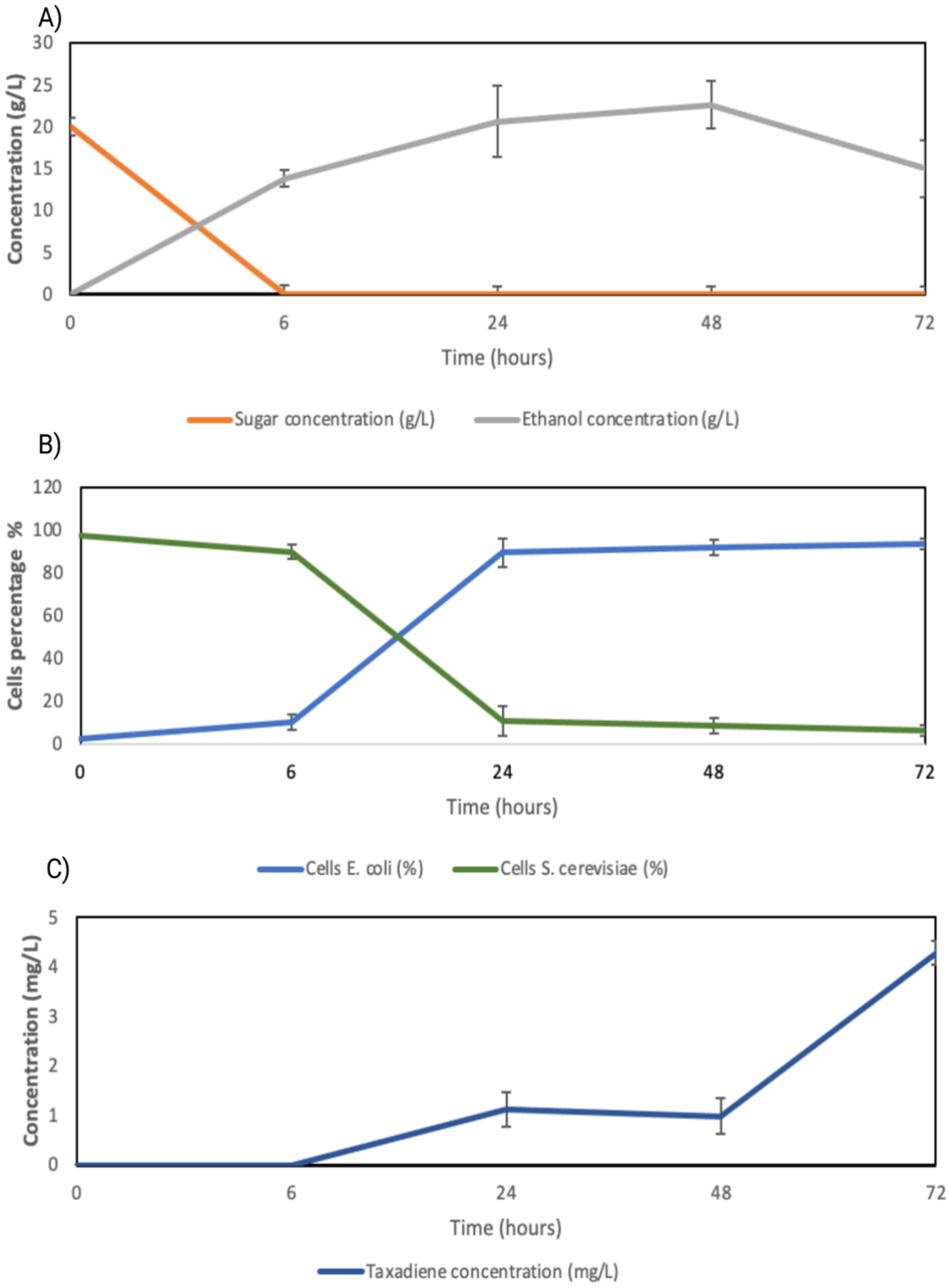


Figure 4.5. Kinetics of a consortium of E. coli– S. cerevisiae showing A) carbon source consumption, ethanol production/consumption, B) population balance and C) taxadiene production. Fermentations were done at 30°C in a 250 mL shake flask using YP and 2% xylose. The experiment was performed in triplicates, and it was run for 3 days, samples were taken every 6, 24, 48 and 72. Error bars represent the standard deviation of the replicates (no statistical analysis was done on this data).

In Figure 4.5 A, the kinetics of the sugar consumption and ethanol production of the consortium can be observed. The initial sugar concentration was 20g/L, which after 6 hrs was completely consumed. This correlates perfectly with ethanol production. The highest ethanol concentration was achieved after 48 hrs reaching ~22.5 g/L. At 72 hours, no sugars were being consumed which led to a decrease in ethanol concentration, possibly by the utilization of this carbon source by yeast. Figure 4.5 B shows how the balance population fluctuated over a period of 72 hours. After approximately 6 hours, the population balance was still the same as at the beginning of the fermentation but after 48 hours *E. coli* had completely outcompeted *S. cerevisiae.* This trend continued until the end of the experiment. It is relevant to mention that the starting point for inoculating the fermentation plates was 40:1 in favour of yeast following the ratios reported in the literature for similar consortia (Kang Zhou et al., 2015). For subsequent experiments, this ratio was further increased based on the results obtained in Figure 4.5 B. Finally, in Figure 4.5 C, we can appreciate the progressive taxadiene production in the co-culture. The highest concentration achieved after 72 hours was 4.3 ±0.24 mg/L. It is important to remark that as mentioned before, the GAL80 deleted yeast strain was not yet developed so to induce expression of taxadiene genes, 1/3 of the carbon source added was galactose, which may have contributed to the limited taxadiene production compared to the titres of 129 ±15 mg/L obtained in previous reports in shake flask was only galactose was used (Nowrouzi et al., 2020).

4.3.2 *E. coli- S. cerevisiae* microbial consortia based on different carbon sources

Following a similar approach to the one described in Figure 4.3, a similar microbial consortia approach was designed based on using raffinose as a carbon source as can be seen in Figure 4.6.

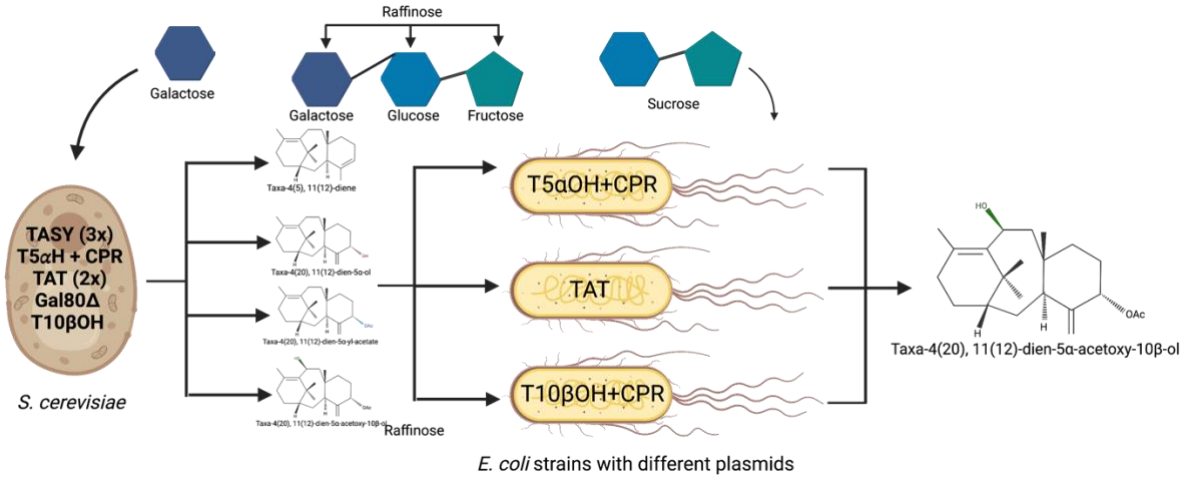
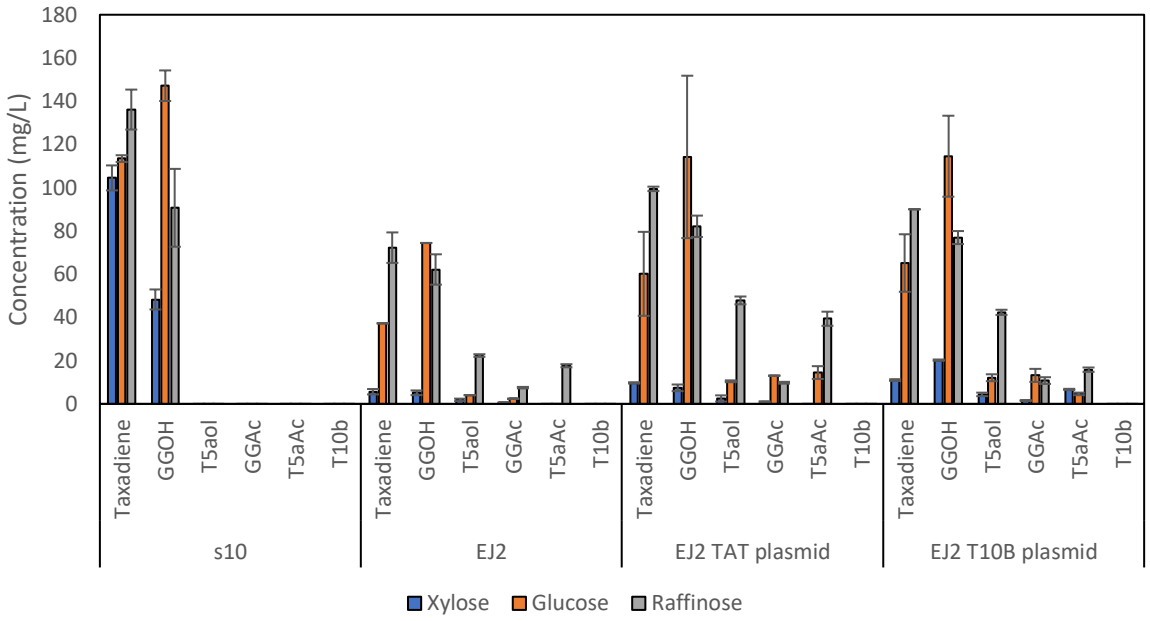


Figure 4.6. Co-culture of S. cerevisiae and E. coli engineered to produce taxanes and expected division of labour. The co-culture is designed to be fed with raffinose. Yeast would metabolize galactose making the disaccharide sucrose (glucose and fructose) available for E. coli.

The first step in this new approach involved focusing first on *S. cerevisiae* individually using raffinose. For this experiment, two different plasmids carrying TAT and T10βOH enzymes were designed for expression in the EJ2 strain (plasmid maps are shown in Appendix 4). An OD of 4 was set for the selected S. *cerevisiae* strains S10.1, EJ2, EJ2-TAT and EJ2-T10βOH controls to replicate the same amount of yeast cells as previously used in the 40:1 inoculation ratio employed for *S. cerevisiae- E. coli* consortia (Figure 4.5). The growth and taxane production of these strains is shown in Figures 4.7 A and B.

A)



B)

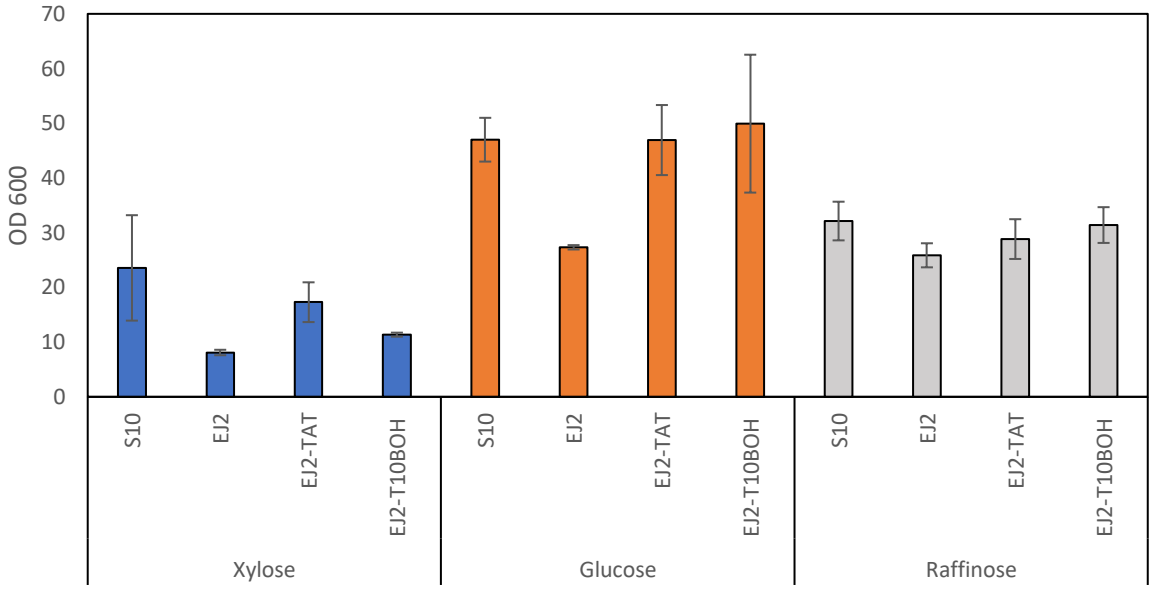


Figure 4.7. Controls fermentations of EJ2, EJ2-TAT, EJ2-10BOH and S10.1. A) Production of different taxanes by different strains in different carbon sources at 2% concentration in YP media B) shows the growth of the same strains. The experiments were done in duplicates in 24 well plates for 3 days at 30° C using. Error bars represent the standard deviations of the duplicates (no statistical analysis was done on this data).

Despite the expression of the T10βOH plasmid in the EJ2 strains the T10β-ol precursor was not detected. However, T5αAc was detected in EJ2-TAT, EJ2-T10βOH and in EJ2 only when grown in raffinose. The maximum production of this compound was accomplished in EJ2-TAT in raffinose with 39 ±3 mg/L while the same strain

produced nearly half of this product when grown in glucose with 15 ±3 mg/L, no production was detected in xylose. The EJ2-T10βOH strain yielded a production of 16 ±1 mg/L in raffinose while in xylose and glucose, this was only 7 ± 0.3 mg/L and 5 ±0.6 mg/L respectively. Finally, the T5αAc precursor in the strains expressing no plasmids was detected only in EJ2 in raffinose with a production of 18 ±0.7 mg/L. These results suggest that the production of T5αAc is doubled by the presence of the TAT plasmid. The production of this compound was not detected in S10.1 which is expected as this strain is only capable of producing taxadiene.

Furthermore, the production of the T5αAc precursor, T5α-ol, reached its maximum by strain EJ2-TAT with 48 ±2 mg/L followed by EJ2-T10βOH with 42 ±1 mg/L and finally EJ2 with 22 ±0.6 mg/L all of these in raffinose. As with the previous compound production in xylose and glucose probed to be significantly lower with a maximum production of 12 ±2 mg/L accomplished by EJ2-T10βOH in glucose which represents a reduction of ~46% when compared to the worst result for the same compound in raffinose.

Finally, strain S10.1 reported the highest taxadiene production over all treatments. When grown in raffinose S10.1 showed a taxadiene production of 136.13 ± 9.2 mg/L which is 27% more compared to the second highest which is EJ2-TAT in raffinose and 96% more when compared with the lowest taxadiene production which was EJ2 in xylose. The explanation behind this is that as mentioned before, the genotype of EJ2 includes T5αOH, TAT and T10βOH enzymes which convert taxadiene into acetylated and oxygenated taxanes while S10.1 is fully focused on taxadiene production. These findings are consistent with what has been reported in the literature (Walls et al., 2021).

When comparing the carbon sources, the lowest production overall was obtained in xylose fermentation while the highest was obtained with raffinose. Once again there is a correlation between biomass and taxane production as shown previously in the last chapter. However, the exception for this is once again when using glucose, probably due to the repressing action of Mig1 as discussed in Chapter 3.

It is also very interesting to mention that growth rates achieved in xylose were initially unusual as it has been reported that native *S. cerevisiae* can't metabolize xylose (Kang Zhou et al., 2015) and this yeast is usually reported as non-xylose consuming (Jin et al., 2004). However, since 1986 there have been reports of xylose utilization in *S. cerevisiae,* especially in the presence of a mixture of substrates (Batt et al., 1986; Toivari et al., 2004; van Zyl et al., 1993; Van Zyl et al., 1989). Based on these reports the xylose growth shown in figure 4.8 B can be justified.

As a side note, it is important to mention that the by-product GGOH included in Figure 4.8 A is naturally produced by yeast by the action of phosphatases, however, its overproduction responds to the fact that mevalonate pathway is overexpressed for GGPP production (Nowrouzi et al., 2020) (Figure 1.2). This compound is also the substrate of the TAT enzyme (Walls et al., 2021) which converts it to GGAc (also shown in Figure 4.7 A). The inclusion of both compounds was just to show the functional expression of this enzyme in the cases in which T5αAc is not detected.

With the results obtained from these controls, it was possible now to compare whether our different consortia were effective in enhancing the production of different paclitaxel precursors. So the next step was to construct different *E. coli* strains transformed with plasmids carrying codon-optimized versions of TAT and the P450's enzymes T5αOH, and T10βOH. However, the expression of plants P450's in *E. coli* has been challenging due to the presence of the lipophilic N-terminal amino acids (Kitaoka et al., 2015). It has been reported that these lipophilic ends direct membrane localization in plants (Biggs et al., 2016). However, in *E. coli* they cause a decrease in solubility that can lead to inclusion bodies and loss of functionality (Biggs et al., 2016). Considering this, 24 amino acids were removed from the N-terminal as previous studies suggested an increased activity with such modifications (Biggs et al., 2016). Furthermore, the deleted amino acids were replaced with an eight amino acid peptide (MALLLAVF), which was previously shown to be effective (Biggs et al., 2016). The newly engineered P450 was then linked to the CPR with an "operon sequence" 5′-GGATCCAAGGAGATATACC-3′ taken from the same study.

Due to the lack of previous reports for protein engineering and optimisation for expression of T10βOH in *E. coli,* an alignment was made comparing the protein

sequences of codon-optimised T5αOH, the truncated version of T5αOH previously reported and the codon-optimised version of T10βOH (Figure 4.8). Based on the results from the alignment it was decided to also delete the first 24 N-terminal amino acids, and add the same eight amino acid peptides and the same operon sequence s for T5αOH. Finally, based on the results achieved by their CPR, the sequence for the reductase was taken from Biggs et al. (2016). The version of this enzyme contains a 74 amino acid deletion at the N-terminal end that is once again replaced by the MALLLAVF peptide. The DNA sequences for all modified proteins are included in the appendix section.

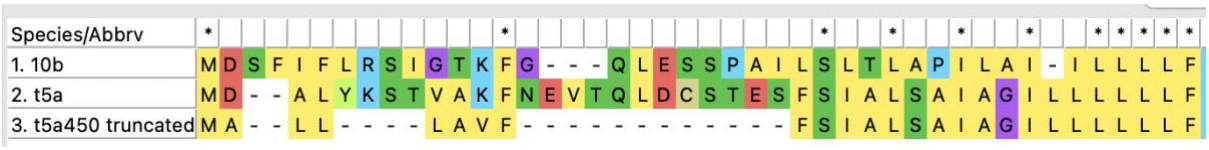


*Figure 4.8. Alignment comparing codon-optimised protein sequences of tT5αOH, T5αOH and T10βOH. The conserved amino acids are indicated by *.*

Once the design was completed, T5αOH, T10βOH and CPR (Cytochrome P450 reductase) enzymes were ordered to be synthesised from Twist Bioscience. The gene corresponding to the TAT enzyme was ordered in an expression plasmid. The same backbone was amplified for assembling the plasmids carrying T5αOH-o-CPR and T10βOH-o-CPR. Furthermore, two different plasmids carrying TAT and T10βOH enzymes were designed for expression in the EJ2 strain. The full sequences, plasmid maps and primers used for the cloning and assembly can be found in Appendix section 4A-4F.

Once taxanes production in yeast was standardized, and all bacteria transformations were confirmed by cell growth in ampicillin media, the 6 different ethanologenic *E. coli* strains transformed with T5aOH, TAT and T10βOH plasmids (three LYglc1 *E. coli* strains and three LYxyl3 *E. coli* strains, each with a different plasmid) and 4 different *S. cerevisiae* strains (S10.1, EJ2, EJ2-TAT and EJ2-T10βOH) were assessed in different microbial consortium *of E. coli- S. cerevisiae* to find the optimal conditions, combinations, and ratios to maximise the production of acetylated and oxygenated taxanes. The starting point for inoculating the fermentation microplates was 40:1 in favour of yeast following the ratios reported in the literature for similar consortia (Kang

Zhou et al., 2015). It is important to clarify that unless specified all bacterial strains with the 3 different plasmids were grown together with the specified yeast strain. For example, consortia labelled as LYglc1 strains/S10.1 would be composed of LYglc1 T5αOH, LYglc1 TAT, LYglc1 T10βOH and yeast strain S10.1. The results are shown below (Figure 4.9).

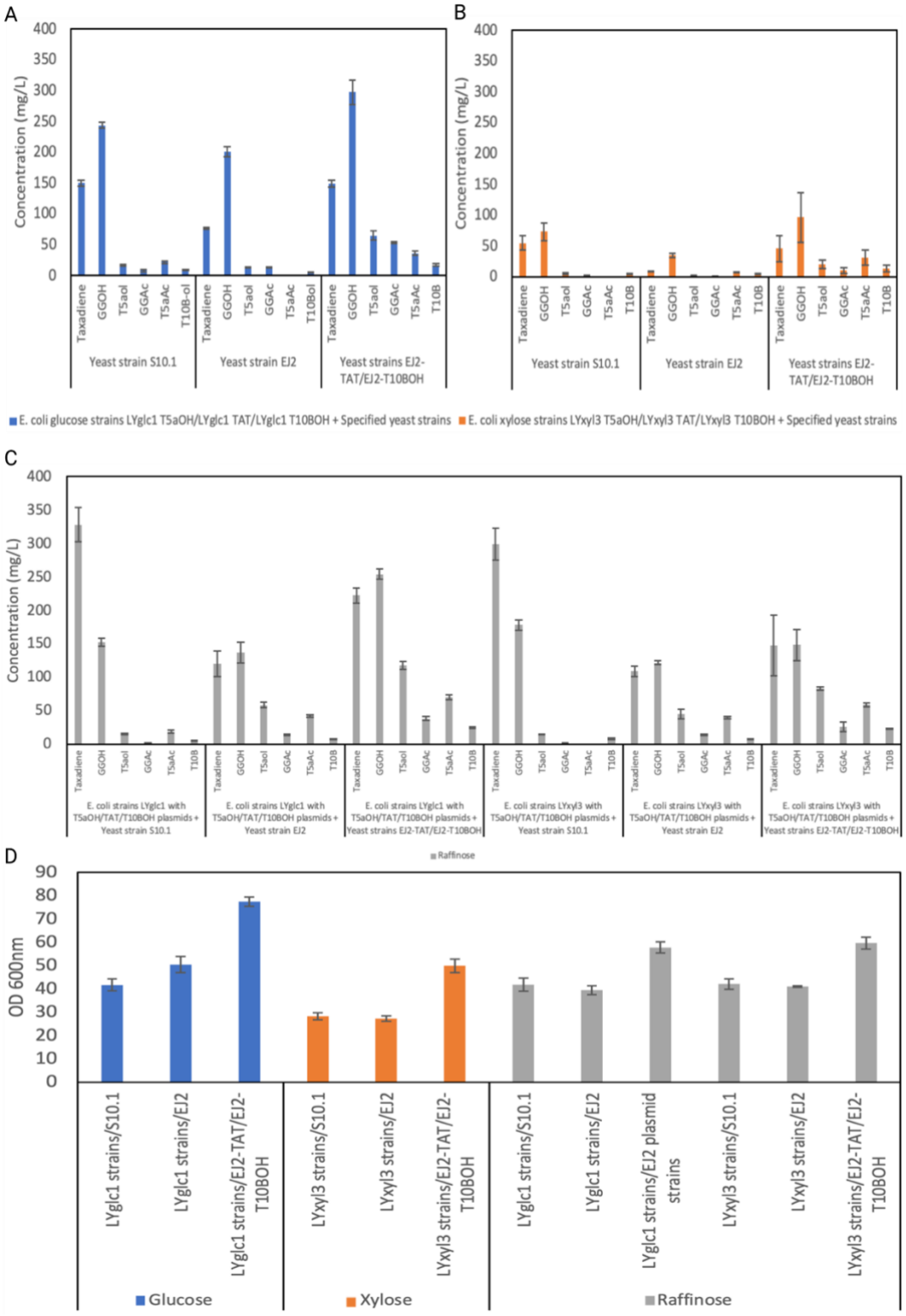


Figure 4.9. LYglc1 and Lyxyl3 E. coli strains (glucose and xylose consuming) containing plasmids with T5aOH+reductase, TAT and T10B+reductase were grown with different yeast strains to quantify the production of acetylated and oxygenated taxanes. The inoculation ratio was 40:1 (4 and 0.1 OD600) in favour of yeast. Unless specified all bacterial strains carrying T5αOH, TAT and T10βOH were used combined with the specified *S. cerevisiae* strain (S10.1, EJ2, EJ2-TAT and/or EJ2 T10βOH). The results for the co-cultures in glucose are shown in A, for xylose in B and raffinose in C. The results showed that oxygenated and acetylated compounds were found in consortia even when yeast did not carry the enzymes to produce these products (S10.1 strain), meaning that there was synergy within the microbial community. D) Final OD600 for the selected consortia using different carbon sources. The experiment was done in 24 well plates for 3 days at 30 °C degrees at 350 rpm. n=3 (no statistical analysis was done on this data).

When compared with the results shown in Figure 4.7 A using individual yeast, the production of taxadiene in S10.1 was 32 % higher for the consortia in glucose, 49% less for the consortia with xylose strains, 140% more in the consortia with glucose-metabolising bacteria strains grown in raffinose and 120% more in the consortia with xylose-metabolising bacteria strains grown in raffinose.

For the consortium of *E. coli* LYgluc1 – *S. cerevisiae* S10.1 grown in glucose or raffinose, the total production of acetylated taxanes was 29.14 and 20.69 mg/L respectively, while for the rest of the carbon sources, the amounts were negligible. However, a pleasant surprise was the detection of T10β-ol for all the carbon sources for these consortia. The highest T10β-ol amounts were detected in the *E. coli* LYgluc1– *S. cerevisiae S10.1* grown in glucose with a total of 9 ± .30 mg/L. To add more perspective to the importance of this finding, it is important to bear in mind that strain S10.1 did not have the enzymes needed for the conversion of taxadiene into the rest of the compounds, so the detection of these in the co-culture is a sign of success in distributing the metabolic pathway in different microorganisms.

The results of the other consortia are interesting to analyse too. Previous mono-culture fermentations of EJ2 towards the detection of T10β-ol were not successful (Figures 3.9, 3.10, 3.11 and 4.7). However, the consortia grown in raffinose incorporating this strain and the LYgluc1 and Lyxyl3 *E. coli* strains could produce up to 25 ±1and 3 ±1 mg/L of the dioxygenated product (T10β-ol). Interestingly, production of the precursors T5αAc in these consortia was 42±2 mg/L (LYgluc*1*) and 40 ±2 mg/L (Lyxyl3) which were 58% and 55% more compared to 18 ±1 mg/L of the EJ2 monoculture.

Lastly, when the last consortia (consisting of both EJ2 strains with plasmids TAT and T10βOH and *E. coli* strains) was grown in raffinose, the production of T10β-ol, reached a total of 25 ±1 mg/L for LYglc1 strains/EJ2-TAT/EJ2-T10BOH and 23 ± 1 mg/L for LYxyl3 strains/EJ2-TAT/EJ2-T10BOH. It is important to mention that this consortium was the most ambitious of all as it involved 2 different yeast strains plus 3 different bacteria strains, which to the best of our knowledge is one of the most successful and complex consortia including five microorganisms dividing a synthetic pathway and achieving true synergy to divide the metabolic burden among the consortia.

Surprisingly the results of the strains in xylose were not as expected. Upon further investigation, it is probably that *E. coli* was producing ethanol from xylose and rapidly outnumbering yeast, which did not achieve a fast enough ethanol consumption and as a consequence not to reaching good levels of taxadiene to be used as a precursor by the enzymes carried by *E .coli*. Also unlike in the original design by Zhou et al., (2015) our yeast strain did not feed on acetate as the production of this compound and other organic acids were inactivated in *E. coli* LY180 strains (Flores et al., 2019a). Pictures corresponding to these consortia are shown in appendix 4G.1, however further analysis is needed to determine the exact ratio at the end of the fermentation.

The high taxadiene accumulation and little amount of other products indicated that there was still potential for optimization of the consortium. Hence the experiment was repeated with different inoculation ratios this time by increasing the initial bacteria population to promote the conversion of the accumulated taxadiene into the other precursors. Three new different consortia were designed as follows. The first consisted of the *S. cerevisiae* S10.1 strain grown with all the LYgluc1 and LYxyl3 strains. The second was EJ2-TAT with the bacteria only expressing T10βOH enzyme and the third was EJ2-T10β once again with the bacteria only expressing T10βOH enzyme. All of these consortia were grown in the different previous carbon sources (glucose, xylose and raffinose). Inoculation rates were adjusted by increasing the OD600 of *E. coli* (total OD600 of 0.6) and maintaining the OD600 of *S. cerevisiae* (OD600 4). The results of these new consortia are shown in Figure 4.10.

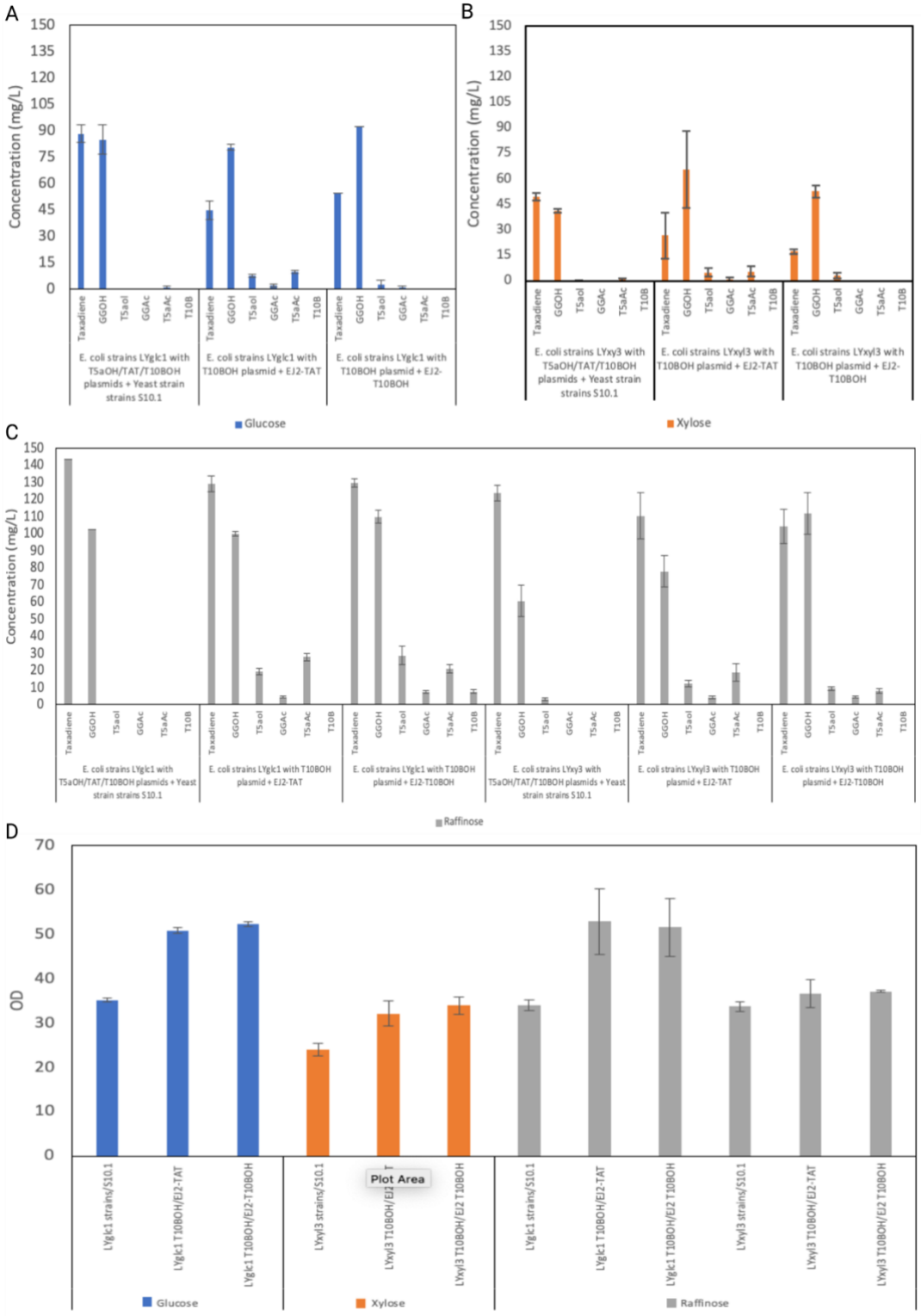


Figure 4.10. LYglc1 and Lyxyl3 E. coli strains (glucose and xylose consuming) containing plasmids with were grown with different yeast strains to quantify the production of acetylated and oxygenated taxanes. The inoculation ratio was ~6.5:1 (4 and 0.6 OD600) in favour of yeast. Unless specified all bacterial strains carrying T5αOH, TAT and T10βOH were used combined with the specified S. cerevisiae strain (S10.1, EJ2, EJ2-TAT and/or EJ2 T10βOH). The results for the co-cultures in glucose are shown in A, for xylose in B and raffinose in C. D) Final OD600 for the selected consortia using different carbon sources. The experiment was done in 24 well plates for 3 days at 30 °C degrees at 350 rpm. n=3 (no statistical analysis was done on this data).

The first thing to note in Figure 4.10 A is that disappointingly, there was no production of the taxadiene derivates in any of the carbon sources for the *S. cerevisiae* S10.1- *E. coli* consortia except for a negligible amount of T5αAc. Compared to the results in Figure 4.8 using mono culture, there was a taxadiene increase of 32% in glucose, a decrease of 48% in xylose and finally another huge increase of 140% in raffinose with glucose strains and 119% in raffinose with xylose strains. Furthermore, compared to the results obtained in the previous consortia, the total production of this precursor decreased by 41% in the consortia grown in glucose, 77.5% in raffinose with glucose-growing strains and 88% also in raffinose with xylose-growing strains. Only in xylose, there was a slight increase of only 9% followed by a negligible amount of T5α-ol.

For the *S. cerevisiae* EJ2 with TAT plasmid- *E. coli* T10βOH consortia, there was no detection of T10β-ol, which in other words meant that there was no conversion of T5αAc into T10β-ol by the bacterial strains. The best production of the T5αAc precursor was achieved by the consortia grown in raffinose. Compared to the production with the EJ2-TAT strain monoculture in raffinose (Figure 4.7 A), the production of T5αAc was 29% and 51% less in the consortia with glucose and xylose-consuming strains respectively. However, if the comparison is made with the previous experiment in the same conditions, there was a reduction of 33% and 52% less for the consortia with glucose and xylose-consuming strains respectively.

Lastly, for the *S. cerevisiae* EJ2 with T10βOH plasmid- *E. coli* T10βOH consortia, there was a production of 7.86 ±1 mg/L of T10β-ol in the treatment with glucose fermenting strains in raffinose. This is interesting when compared to the monoculture (Figure 4.7 A) where there was no production detected. On the other hand, there was no production detected in the treatment with xylose-fermenting strains in raffinose (unlike in the previous experiment where production was 22.9 ±0.85 mg/L). Some pictures of these consortia were also included in appendix 4G .2

Overall, it seems that in this experiment, there was not enough synergy between the strains for a correct distribution of labour and the optimization trial did not have the expected results. Hence one last consortium was designed. To simplify DOL *S. cerevisiae* S10.1 was co-cultured with LY E. *coli* strains with T5αOH plasmids. The idea was that yeast would produce taxadiene and *E. coli* would carry out oxygenation

of this compound. Additionally, another consortium of *S. cerevisiae* EJ2/ LY E. *coli* strains with T10β and TAT plasmids was tested. The hypothesis was that the production of precursors by EJ2 would act in synergy with the enzymes expressed in *E. coli* to enhance the production of aT5αAc and T10β-ol. As in the previous experiments, each consortium was tested with different carbon sources. The total OD600 for *S. cerevisiae* was maintained at 4, and a new OD600 of 2 was set for *E. coli.* The inoculation ratio was 20 :1, this was chosen because it represents a middle point between the first consortia tried in Figure 4.9 (inoculation ratio of 40:1) and the second in Figure 4.10 (inoculation ratio of ~6.5:1). The results for these experiments are shown in Figure 4.11.

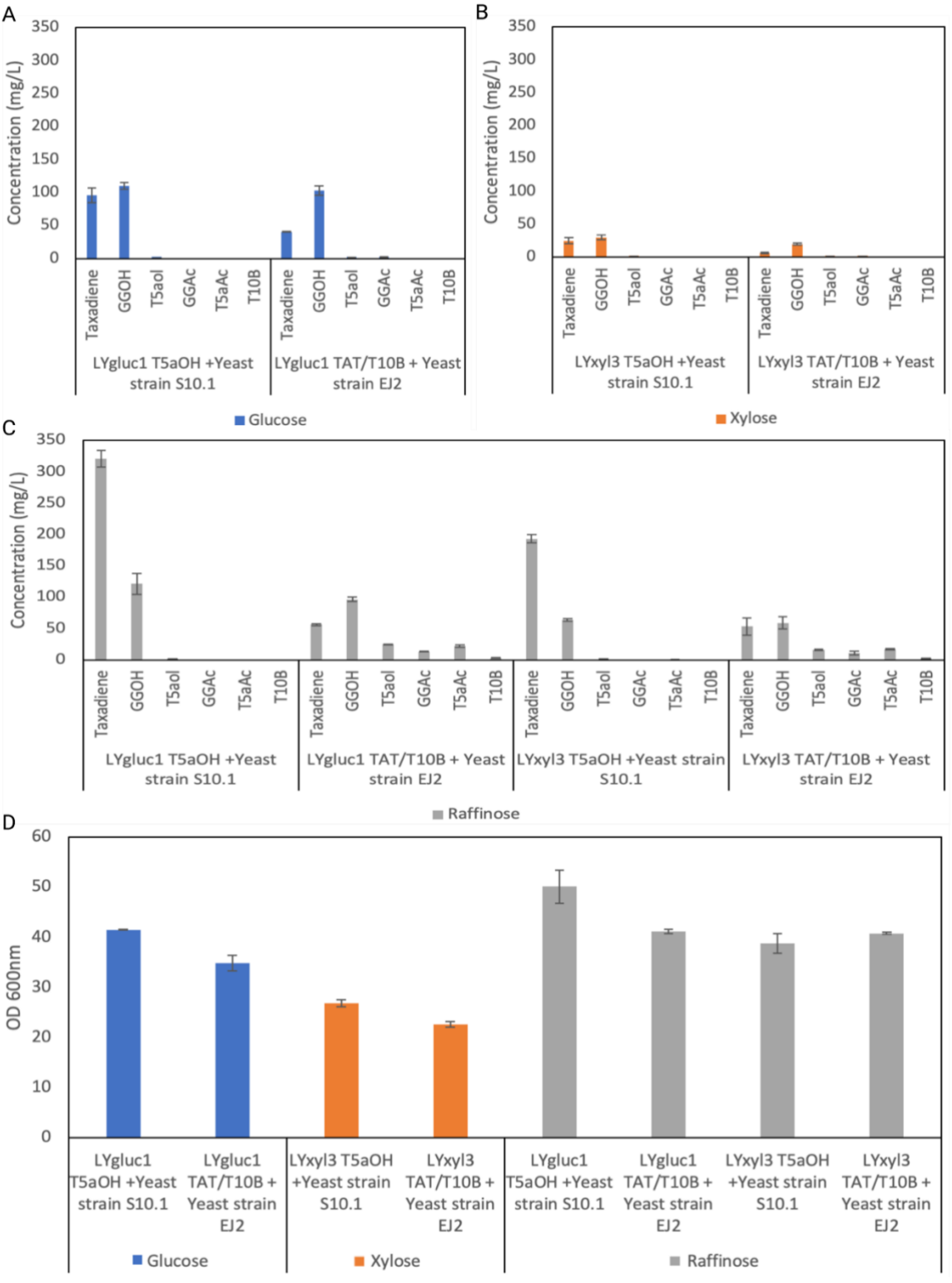


Figure 4.11. LYglc1 and Lyxyl3 E. coli strains (glucose and xylose consuming) containing plasmids with were grown with different yeast strains to quantify the production of acetylated and oxygenated taxanes. The inoculation ratio was 20:1 (4 and 0.2 OD600) in favour of yeast. Unless specified all bacterial strains carrying T5αOH, TAT and T10βOH were used combined with the specified S. cerevisiae strain (S10.1, EJ2, EJ2-TAT and/or EJ2 T10βOH). The results for the co-cultures in glucose are shown in A, for xylose in B and raffinose in C. D) Final OD600 for the selected consortia using different carbon sources. The experiment was done in 24 well plates for 3 days at 30 °C degrees at 350 rpm. n=3. Two different ANOVAs comparing the production of total acetylated taxanes in monocultures vs raffinose consortia with glucose strains and monocultures vs the consortia in raffinose with xylose strains showed no statistical difference in production (p-value 0.09 <0.05 p-value 0.5 <0.05 respectively).

The different consortia of *S. cerevisiae* S10.1- *E. coli* T5αOH did not show any relevant production of T5α-ol. This once more indicated that there was no division of labour between yeast and bacteria. Production of taxadiene was overall less in all the conditions except for the consortium in raffinose with glucose strains. This treatment reached a production of taxadiene of 320 ±13.4 mg/L which was 135% higher in comparison to the S10.1 monoculture (Figure 4.7 A).

Finally, the production of compounds in the consortia consisting of EJ2 – *E. coli* T10βOH and TAT only produced negligible amounts of T10βOH. Taxadiene and total acetylated taxane yields were very similar. Despite this, two separate ANOVAs proved that the difference in the production of total acetylated taxanes between the monoculture and the consortia was not significant (p-value 0.09 <0.05 for the consortia in raffinose with glucose strains and p-value 0.5 <0.05 for the consortia in raffinose with xylose strains).

Once more, some microscope pictures of the different treatments with S10.1 and the bacteria after 72 hours were included in appendix Figure 4G. 3.

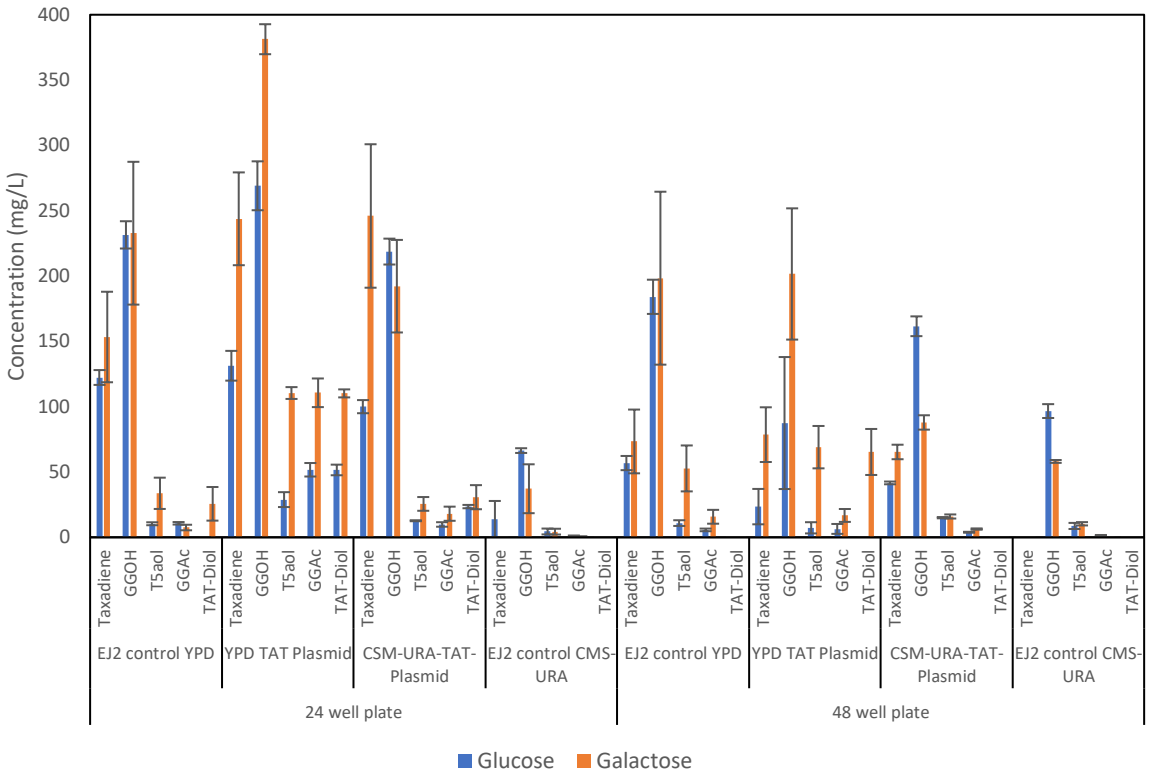
4.3.3 *S. cerevisiae-S. cerevisiae* microbial consortia for the production of T10βOH

Following the previous results, the main objective was to expand the production of the precursor Taxa-4, 11-dien-5a-ylacetate (T5αAc) and Taxa-4, 11-dien-5a-acetoxy-10b-ol (T10βOH). Therefore, it was decided to engineer an *S. cerevisiae-S. cerevisiae* consortia using the different yeast strains described earlier in the methodology section and focus on boosting the expression of TAT and T10βH using high copy number plasmids (EJ2-TAT and EJ2 T10β). However, as mentioned before, the addition of heterologous components in a strain can “kidnap” the molecular machinery of an organism and put an excessive metabolic burden on it. Such a burden can reduce yields in the production of a component and in the worst case, it can even prevent the host from growing (G. Wu et al., 2016). That is why to avoid a metabolic imbalance caused by the overproduction of enzymes, we decided to distribute the labour and transform the EJ2 strain with either a plasmid containing the TAT cassette or the T10βH *one.* Plasmid construct was done by PCR amplification of all the required

fragments and by adding overlapping regions on a subsequent PCR. Assembly was done *in vivo* in yeast by homologous recombination following the methodology established by (Malcl et al., 2022). Transformed colonies were selected using CMS-URA media and Colony PCR was performed to confirm the presence of the interest gene. The appendix section includes all the plasmid maps for yeast (Figures 4E-4F).

As with the previous consortia, strains EJ2 and EJ2-TAT plasmid were analysed in monocultures first. Also, additional fermentations were done with the new strains in CSM-URA and YPD and YP galactose. In the previous experiments, OD600 for yeast was 4 to balance the consortia with *E. coli.* However, since in this case, the consortia will be exclusively yeast strains the OD600 will be adjusted to 1. The same experiment was done in 48-well and 24-well plates to test variations in production and growth. The results are shown in Figure 4.12. As can be seen, taxane production was overall better in 24-well plates hence the discussion will be focused on these results.

A)



B)

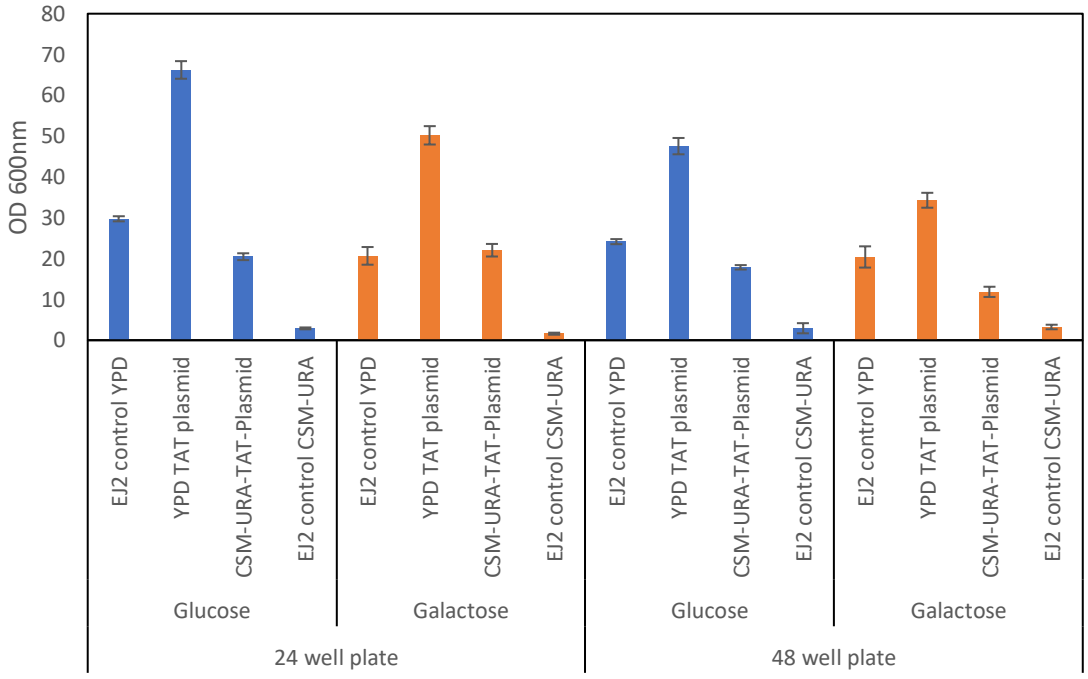


Figure 4.12. The strains EJ2 and its version EJ2-TAT carrying a copy with the TAT plasmid were grown in YP and CSM-URA with glucose and galactose. A) Production of different taxanes by different strains in different carbon sources while B) shows the growth of the same strains. The experiments were done in duplicates in 24 and 48 well plates for 3 days at 30 °C using a 2% sugar concentration. Error bars represent the standard deviations (no statistical analysis was done on this data).

The results in Figure 4.12 showed very encouraging results in terms of TAT production. The strains carrying the plasmid with TAT had better production levels for every compound in comparison with strain EJ2 in all conditions. The poor performance of EJ2 in CSM-Ura was completely expected as this strain was not capable of producing uracil for growth. The production of some taxanes by this strain was carried out by the cells that were added at the inoculation and not by new cells generated during the fermentation. Even though initially the fermentations were done in CMS-URA to avoid plasmid loss, interestingly significantly better results were obtained in YP. It is important to clarify that GGAc is an undesired by-product resulting from the conversion of GGOH by the unspecific activity of TAT (Walls et al., 2021). However, T5αAc is the substrate for T10βOH and precursor of T10β-ol hence obtaining good amounts of this product at this stage was important for the detection of the next product in the paclitaxel metabolic pathway. The production of T5αAc for EJ2 in YP was 38.5 ± 1.5 mg/L, followed by similar values by EJ2-TAT plasmid in CSM-URA with 39.4 ± 7.9 mg/L and finally EJ2-TATplasmid in YP with 110 ±3 mg/L which represents an almost 3-fold increase compared with the other treatments and it's the highest amount ever reported for this compound up to date.

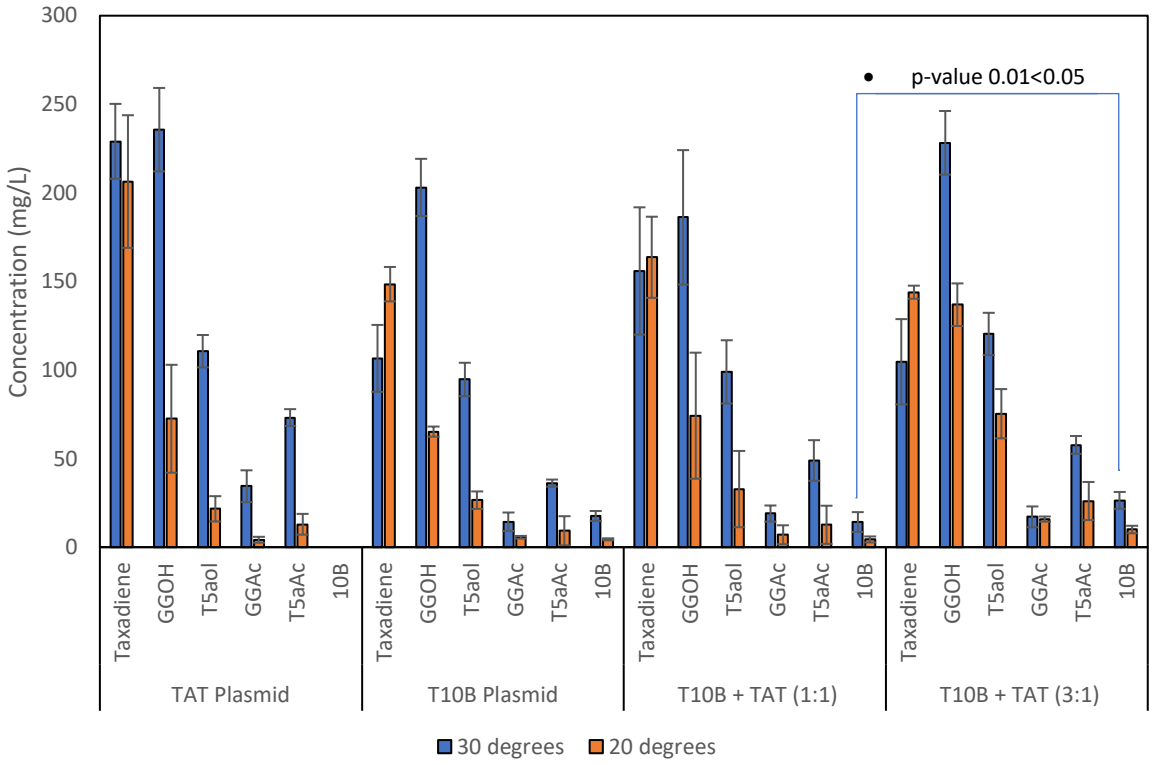
The total production of acetylated compounds was better achieved by EJ2-TAT plasmid in YP with 220.8 mg/L followed by 62.8 mg/L by EJ2-TAT plasmid in CSM-URA and 46 mg/L by EJ2 in YP. The difference in production between the strains growing in YP with and without the TAT plasmid represented a difference of 82%.

The higher production of the acetylated compounds GGAc and T5αAc can be explained not only by the fact that the strain already contained two copies of the enzyme TAT, but also by the high copy number of the plasmid containing the 2μ sequences (with a copy of 40-60 per cell) (Y.-T. Liu et al., 2014) maximizing the expression of the bottleneck enzyme. In addition to this, because of GAL80 deletion, yeast does not stop growing and producing even after galactose is depleted.

Despite the good T5αAc results achieved in this experiment, the product T10βOH was still undetectable, hence a co-culture of both strains with plasmid insertions was engineered. EJ2 strains containing T10βOH and TAT plasmids were grown together in different ratios. Controls of strains having the TAT or T10βOH plasmids only were

also included. The initial ratio concentrations of both strains with plasmid were 1:1 or 3:1 (T10βOH: TAT). As shown in the previous Figure 4.15 A, the production proved to be more efficient in YP media with 2% galactose, therefore the decision was taken to keep using this formulation for this experiment. Also, both temperatures of 20^{o}C and 30^{o}C were tested. The results of these fermentations and their corresponding OD600s are shown in Figure 4.13.

A)



B)

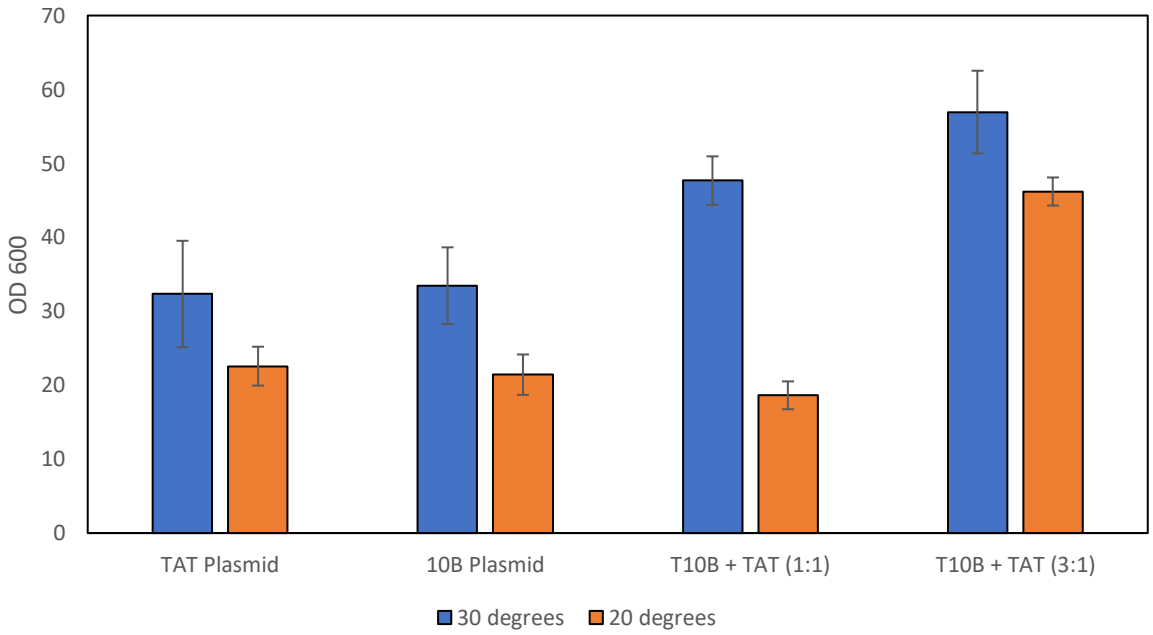


Figure 4.13. Production of A) different taxanes and B) final OD600 of single and S. cerevisiae/ s. cerevisiae consortia strains with TAT or T10BH plasmids in different carbon sources and temperatures. The strains EJ2 and EJ2-TAT were grown in YP and CSM-URA with glucose and galactose. The experiments were done in duplicates in 24 well plates for 3 days at 20 °C and 30 ° C using a 2% sugar concentration. Error bars represent the standard deviations. The results compared are signalled with a line. The asterisk indicates a significant difference between both results (p-value 0.01<0.05) based on Student's t-test.

Taxadiene production at 20°C was up to 40% higher than at 30°C. This showed consistency with the literature, as it has been reported before that the activity of the enzyme metabolizing conversion of GGPP to taxadiene (TASY) was enhanced at lower temperatures (Nowrouzi et al., 2020). In comparison, it's interesting to see that despite the good taxadiene yields, the production of the other compounds was severely reduced at the lower temperature. This aligns with preliminary results showing a decrease in the T5αOH activity in *S. cerevisiae* at 20°C (Walls et al., 2021). Interestingly, the same has been observed to happen at lower temperatures when the same enzyme was expressed in *E. coli* (Biggs et al., 2016). Additionally, as can be seen in Figure 4.7, a reduction in temperature affected yeast growth, which is consistent with previous literature (Nowrouzi et al., 2020).

The strain EJ2 with theTAT plasmid (far left in Figure 4.13 A) showed very similar Taxadiene and T5α-ol production levels to those achieved by the control of EJ2 without plasmid (Figure 4.12 A) with 229± 21.15 mg/L and 110.6 ± 9.1 and 208.7 ± 13.2 mg/L and 110.4 ±7.8 mg/L respectively. The T5αAc production was higher than GGAc production compared with the control, however, the total acetylated taxanes production was doubled in the latter.

The results with the strain containing the T10βOH plasmid showed a 53% decrease in taxadiene production when compared to the strain with the TAT plasmid. Despite this, the results were satisfactory as for the first time, Taxa-4, 11-dien-5a-acetoxy-10b-ol (T10β-ol), was detected with a calculated amount of 17.6 ±2.8 mg/L. In this strain, T5αAc was half of what was obtained with EJ2T-TAT however this was probably due to the conversion of the acetylated taxane into the dioxygenated one T10β-ol.

Next, the results of two consortiums composed of EJ2-TAT and EJ2 T10βOH with different inoculation rates of 1:1 and 3:1 (T10βOH: TAT) are discussed. The logic behind these ratios is that enough levels of T5αAc were being produced so the lack in the detection of T10β-ol must have been due to a deficiency in the production or expression of this enzyme. The consortia with a 3:1 proportion in comparison with the 1:1 proportion showed 33% less production of taxadiene, 21% more of T5α-ol, 10% less of GGAc, 18% more of T5αAc and an impressive 86% increase of T10β-ol. A one-way ANOVA of the triplicates confirmed the rejection of the null hypothesis (p-

value 0.01<0.05) highlighting that the inoculation ratio of both strains did significantly affect T10β-ol production. These results seemed to be explained by an additive effect of the accumulation of the precursor produced by the strains transformed with TAT plasmid and further conversion of the acetylated compound by strains with the T10βOH plasmid.

To visualise the results, in Figure 4.14 a comparison of different Gas-Chromatograms is shown. In the image the different product profiles of A) EJ2 TAT B) EJ2 T10βOH C) EJ2 TAT -EJ2 T10βOH in an equal ratio and D) EJ2 TAT -EJ2 T10βOH in a 1:3 ratio.

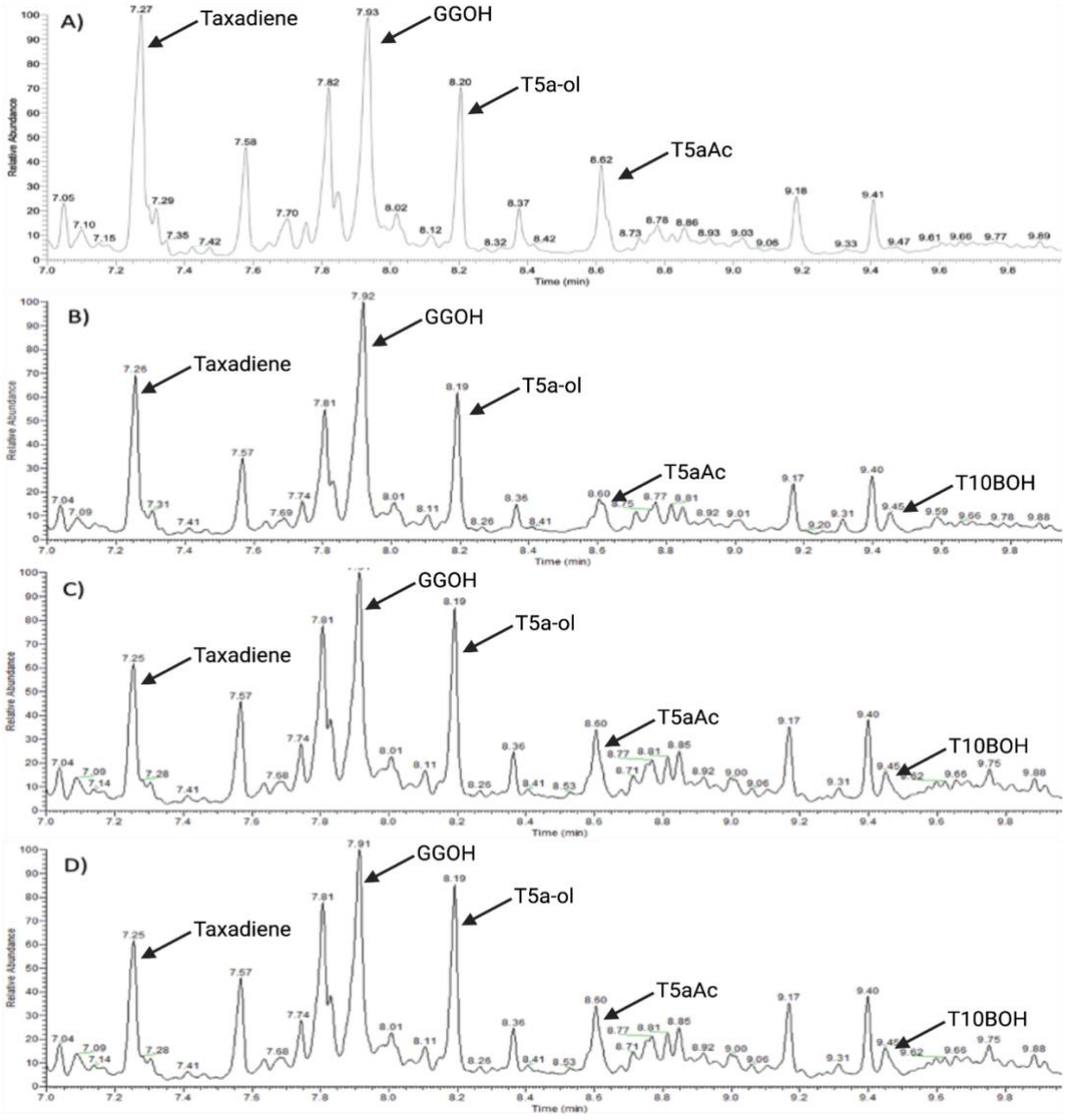


Figure 4.14. Gas chromatograms of the different experiments from Figure 4.13 A) show the product profile of the strain EJ2 with the single TAT plasmid B) shows the product profile of strain EJ2 with a single T10bOH plasmid C) Shows the product profile of a culture with an equal proportion of strains with TAT and 10bOH plasmids D) Shows products of a culture with a proportion of 1:3 10bOH plasmids. TAT is shown at 8.62 min while 10βOH product is shown at 9.45 min.

4.3.4 Limitations

Despite the efforts in standardizing all experiments in this chapter, there still are some limitations that need to be mentioned as these can affect the reproducibility of the experiments or can be the source or errors.

In the experimental results corresponding to Figure 4.5, no viability staining was conducted which means that it is impossible to determine:

A) if decrease in the number of yeast cells after 24 hrs was due to the relative number of yeast cells diminishing in comparison to the increase in *E. coli* cells or
B) if the decrease in the number of yeast cells was a response to the toxicity in the media due to the high ethanol concentrations achieved.

A viability study performed by staining the cells with methylene blue (Matsumoto et al., 2022) could have been enough to determine the real reason behind the decrease in yeast population allowing to develop an strategy to counteract this.

The inoculation was done by using calibrated pipettes to assure that the volumes taken were accurate as any change in the volume may have influenced the final OD in each well and consequently affect the population balance. However, manual methods for liquid handling always carry an error which in this case may potentially affect the reproducibility of the experiments.

The biggest limitation in these experiments was the lack of a method to quantify with accuracy the final populations of each microbial specie as these methods combined with a statistical design of the experiment could have potentially improved the results for many of the consortia tested.

4.4 Conclusions

The modular strategy taken combined with the inoculum engineering approach proved functionality and allowed for different consortia combinations to be tested in an easy and fast manner. Furthermore, this approach enables the detection of T10βOH.

Combining these methods with statistical analysis of the design of experiments can further be used for the optimization of the consortia not only for production optimization but also to test the balance of the consortia given the combinations of different microorganisms and carbon sources.

With the use of plasmids in *S. cerevisiae EJ2*, it was possible to increase the production of the acetylated compounds GGAc and T5αAc. The production of these compounds was the highest ever reported including the results shown in chapter 3.

Taxadiene concentration was the highest when the strain containing TAT enzyme was grown as a monoculture, however, production of the di-oxygenated product T10β-ol was still not observed. On the contrary, the strain containing T10βOH plasmids did show some production of the compound. A co-culture of both strains where the initial concentration was varied from a 1:1 ratio to a 3:1 (TAT: T10βOH) showed the highest T5αAc and T10β-ol concentrations ever reported in yeast with productions of 110 ±3 mg/L and 26 ±5 mg/L respectively. Even though no selection media was used, there was no evidence of plasmid loss. This is also the first time that the production of complex products is reported to be achieved in a yeast co-culture by a combination of genomic modifications (featured in Chapter 3) along with plasmid transformation. The high yields of the T10β product seem to be explained by the additive effect of the accumulation of the precursor produced by the strains transformed with TAT plasmid which also highlights the impact of the inoculum engineering.

The photographs taken of the consortia are useful for illustrating the population balance. However, this method is biased as the area of visualization does not account for the whole sample. The use of better methods like cell sorting using flow cytometry might be better for quantifying the different cell types.

Chapter 5:

MICROBIAL CONSORTIA ENGINEERING FOR THE DISCOVERY OF A NEW TAXOL TAXA-4, 11-DIEN-5 α -ACETOXY-1α,10β-DIOL

Chapter 5: Microbial consortia engineering for the discovery of a new Taxol intermediate Taxa-4, 11-dien-5a-acetoxy-10b-ol

5.1 Introduction

Diterpenes are molecules that naturally have been isolated mostly from higher plants, fungi and bacteria (Toyomasu & Sassa, 2010). These compounds of complex molecular structures and oxidation patterns have had a great impact on different research fields such as organic chemistry and drug discovery (Dibrell et al., 2021). Furthermore, once the carbon skeletons are formed –through the cyclization of GGPP—many different diterpenes are formed by the modification of this skeleton through a big number of reactions that include oxidation, reduction, acetylation, methylation, and glycosylation (Toyomasu & Sassa, 2010). Among diterpenes, paclitaxel has been found to work as a powerful inhibitor of cell growth and as consequence it's been used as a very effective anti-cancerogenic drug (Kusari et al., 2014). Because of this, special emphasis has been put on the synthesis of this compound. However, its complexity has hindered its production. Moreover, with an estimate of 20 reactions (T. Wang et al., 2021), the metabolic pathway is not fully elucidated and different *in silico* approaches and bioinformatic tools have been used to propose new candidate enzymes in the biosynthetic pathway (Ramírez-Estrada et al., 2016; Sanchez-Muñoz et al., 2020). To provide an example, one of the most powerful tools is Next Generation Sequencing (NGS) which has been used for transcriptomic studies of *Taxus* tissues and cell suspension cultures (Mutanda et al., 2021). These studies coupled with the appropriate bioinformatic tools for data analysis can provide new insights including not only new genes but also unravelling the inner interactions of the pathway (Mutanda et al., 2021). A study made in 2019 (Kuang et al., 2019) used third-generation and next-generation sequencing to sequence several *Taxus cuspidate* transcriptomes and reveal 7 potential candidates genes for BAHD ACT, an enzyme involved in the formation of the 2-debenzoyltaxane. Furthermore, many studies analysing transcriptomics have been useful to compare taxol synthesis of different *Taxus* species like *T. mairei* (Hao et al., 2011) and *T.* cuspidata (Kuang et al., 2019) allowing researchers to find correlations between taxol production and expression levels of different taxol pathway genes.

A final and relevant example of the use of transcriptomics is the work done by Ramírez-Estrada et al., (2016). In this study, the authors coupled transcriptomics with DNA-amplified fragment length polymorphism (cDNA-AFLP) to detect 15 putative genes encoding for 5 steps involved in taxane biosynthesis (Figure 5.1).

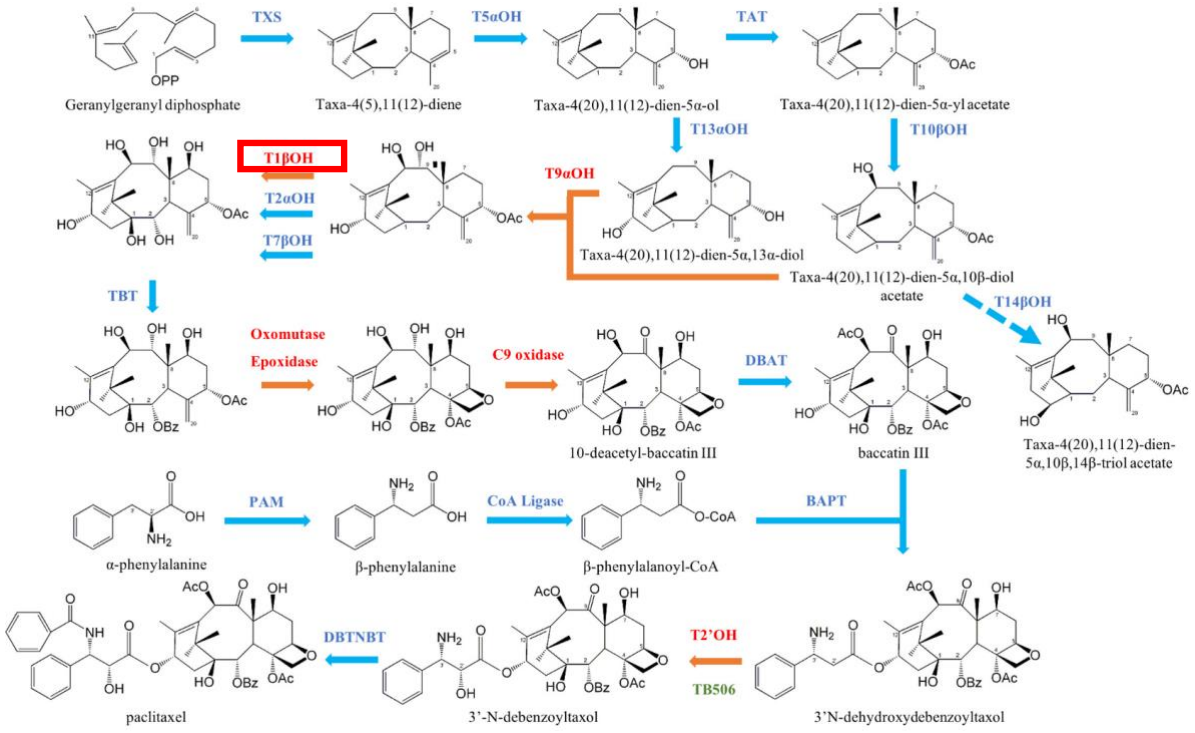


Figure 5.1 Taxane biosynthesis summary. The characterised enzymes are shown in blue while the unknown are shown in red. The enzymes contemplated in our system are taxadiene synthase (TXS), taxane 5α-hydroxylase (T5αOH), taxadien-5α-ol-O-acetyltransferase (TAT), taxane 10β-hydroxylase (T10βOH), and the putative taxane 1β-hydroxylase also known as T1βOH (squared). Taken from (Sanchez-Muñoz et al., 2020).

From the genes discovered in the study previously mentioned, it is important to remark on the discovery of two new cytochromes P450's tagged as TB574 and TB331 (Sanchez-Muñoz et al., 2020). The new candidate genes were later postulated as possible taxane hydroxylases, and both were noted to have very high homology with other characterised taxane hydroxylases. It also was inferred that the new genes would act on carbons 1β and 9α. Finally, in a recent work (Escrich-Montana, 2022) in silico docking analysis was done with both putative genes TB574 and TB331***. The results confirmed the hydroxylase activity for TB331 but discarded its involvement in the paclitaxel biosynthetic pathway. Furthermore, a similar analysis with 37 ligands

revealed the potential of TB574 as the T1βOH gene and its activity as hydroxylase of carbon 1β of the taxadiene core.

In this work, we used the successful microbial consortia strategy described in the previous chapter to produce and characterise the product of the new putative gene T1βOH and expand the paclitaxel metabolic pathway. Furthermore, we use NGS tools to genomically characterised strain EJ2.

*** At the moment of writing this work the research cited had a 12-month embargo so most of the details have been omitted.

5.2 Methodology

5.2.1 Strains

All the strains in this study were the following and derived from:

- ***S. cerevisiae EJ2*** *CEN.PK2-1C (EUROSCARF),* (MATa, leu2-3, 112::HIS3MX6-GAL1p-ERG19/GAL10p-ERG8;ura3-52::URA3-GAL1p-MvaSA110G/GAL10p-MvaE [codon optimized]; his3Δ1::hphMX4-GAL1p-ERG12/GAL10p-IDI1; trp1-289::TRP1_GAL1p-CrtE(X.den)/GAL10p-ERG-20;YPRCdelta15::NatMX-GAL1p-CrtE(opt)/GAL10p-CrtE; ARS1014::GAL-282 |1p-TASY-GFP; ARS1622b::GAL1p-MBP-TASY-ERG20; ARS1114a::TDH-3p-MBP-TASY-ERG20), ARS511b::GAL1p-T5αOH/GAL3-CPR, RKC3::GAL1p-TAT, RKC4::GAL1p-TAT, ΔGAL80, ARS1531::Galp1-T10β
- ***S. cerevisiae EJ2-TAT****:* EJ2 but carrying a plasmid with the TAT gene.
- ***S. cerevisiae EJ2-T10β*** : EJ2 but carrying a plasmid with the T10β gene.
- ***S. cerevisiae* AE1** *KM* but carrying a plasmid with the T10β gene.

AE1 strain was kindly donated by Ainoa Escrich from Universidad Pompeu Fabra. The details of the plasmid construction can be found in (Escrich-Montana, 2022), however, at the moment of writing this dissertation the work cited had a 12-month embargo.

5.2.2 Media preparation

All the chemicals were sourced from Sigma-Aldrich unless otherwise stated. Different media were used for this research and all the ingredients are listed below in Table 5.1. The quantities included are for the preparation of 1L of media.

Table 5.1 Ingredients used for the preparation of different media.

Ingredient	Quantity
Yeast extract	10 g
Peptone	20 g
Carbon source (glucose and/ or galactose)	200 g
Agar (optional)	20 g

5.2.1.1 YP Media

1. To prepare 1 L of YP media with a 2% carbon source concentration yeast extract and peptone were added in a Duran bottle according to Table 2.2. Agar was added when Petri dishes were needed.
2. 800 mL of distilled water was added to the bottle.
3. The chosen carbon source was added to the media in a different bottle and filled with 200 mL of distilled water.
4. Both bottles were autoclaved for 20 min
5. In a sterile environment, the bottle containing the carbon source was added to the bottle of media to achieve a final concentration of 2%
6. OPTIONAL: if Petri dishes (15 mm x 100mm) were prepared, add ~20 mL to each and leave them open until the media is solidified.
7. Store Petri dishes in a cabinet or at 30 °C in an incubator.

5.2.3 Genomic characterization of *S. cerevisiae* EJ2

The necessary steps for the genomic characterization of strain EJ2 are detailed as follows.

1. The Thermo Scientific Yeast DNA Extraction Kit (78870) was used to extract Genomic DNA from the EJ2 strain previously grown overnight in YPD with 2% glucose. The complete protocol can be found on the manufacturer's website.
2. Following DNA extraction, the Genomic DNA (gDNA) was quantified, and purity was assessed in NanodropTM 2000c spectrophotometer at 260 nm to measure the 260/280 and 260/230 Abs ratio (pure DNA should have a 260/280 ratio of ~1.8 and 260/230 of 2 to 2.2).
3. The resulting DNA was diluted in sterile DNase free water to achieve an approximate concentration of ~200ng/μL of gDNA which was used for Next Generation Sequencing (NGS) to observe the genetic modifications of our strain along with some other genes of interest.
4. The Rapid sequencing Kit and sequencing protocol (SQK-RAD004) from Oxford Nanopore Technologies (ONT) were used for sequencing in MinION Flongle flow cells. The protocol is only accessible through the manufacturer website.
5. The MinKnow software from ONT was opened on a computer and the default settings were selected to start the sequencing with a quality score of 9.
6. The results were stored in an external hard drive for further analysis.

The results were analysed using 2 different approaches: referenced-based assembly and *de novo* assembly.

5.2.3.1 Referenced-based assembly

The reference-based assembly was done by using the tool analysis provided by ONT (MinKnow and Epi2me Agent).

5. In this method the reads were aligned against the *S. cerevisiae* reference genome obtained from https://www.ncbi.nlm.nih.gov/genome/?term=s.+Cerevisiae), this was done by choosing the “align to reference “ option from the menu, it’s important to

mention that this alignment is only made for assembly purposes and not for comparing sequences.

6. Basecalling was made using the High accuracy algorithm (Guppy v6.3.8).
7. Once the Fastq files were generated the reads were aligned once more using EPI2ME Desktop Agent and Fastq Custom Alignment protocol which generates BAM files (Binary Alignment Map).
8. Finally, the BAM files were visualised using the Broad Institute's Integrative Genome Viewer (IGV) (https://software.broadinstitute.org/software/igv/).

5.2.3.2 De novo assembly

The methodology for the *de novo* assembly included the use of Line Command Tools (LCM).

4. The fastq files generated from the sequencing were assembled using *Flye*.
5. Once the alignment was made, polishing was made by using medaka and aligned using mini_align from Pomoxis.
6. The generated BAM sequences were visualised using IGV.

The code used for the *de novo* assembly is included in appendix section 5 B.

As a side note, it is important to mention that the whole NGS protocol was performed on 4 different occasions, however, only the repetition yielding the best results will be discussed. A comparison between the outputs of the 2 best sequencing runs is included in appendix section 5 A just to highlight the main differences between the repetitions.

The summarised methodology is shown in Table 5.2.

Table 5.2 Methodology used for genome assembly. The fastq files obtained from the sequencing run in the MinIon sequencing device were assembled by two different methods summarised in the table below.

Processing steps	*Reference based*	*De novo*
1.- Basecalling	MinKnow (Guppy)	MinKnow (Guppy)
2.- Assembly	MinKnow (Guppy)	Flye
3.- Polishing	----------	Medaka
4.- Alignment	Epi2Me Agent	Minimap2
5.- Visualization	IGV	IGV

5.2.4 Microscale fermentations for the production of T1βOH

To evaluate the production of the novel compound T1βOH different consortia were tested using Axygen's polypropylene 24 V-shaped well plates in a microplate's Thermomixers C by Eppendorf. The protocol is detailed as follows:

1. Before the fermentation, the strains *S. cerevisiae EJ2-TAT, S. cerevisiae EJ2-T10b and S. cerevisiae* AE1 were grown in overnight cultures for at least 16 hours in 50 mL Falcon tubes containing 5 mL YPD and were incubated at 30ºC inside shaking incubators at 200 rpms.
2. The next day 160 µL (final 2% concentration) of galactose was added to the well of a sterile Axygen's polypropylene 24 V-shaped well plate.
3. For this experiment, two different inoculation ratios and initial ODs were tested for EJ2-TAT/EJ2-T10□OH/AE1 consortia. The first inoculation ratio was a 1:1:2 (initial ODs were 1,1,2) and the second was 1:1:6 (ODs of 0.125, 0.125 and 0.75). The OD was measured at 600nm using nanodrop to quantify biomass and based on this an appropriate volume of inoculum was added to the well.
4. The remaining volume was completed with YP until a volume of 1600 µL was reached.
5. Finally, 400 µL of dodecane were added to the well. To preserve sterility the plates were covered with Thermo Scientific's sterile gas-permeable adhesive films (AB0718).

6. The plates were taken to Eppendorf's Thermo mixers C –with a plate holder adapter— and were agitated at 350 RPMs at 30ºC for 3 days. The preparation of the plates was done in a laminar airflow cabinet to preserve sterility. All samples were done triplicates unless specified.
7. Once the fermentation finished, the media was collected in 2mL Eppendorf tubes to measure the final OD.
8. The dodecane layer was separated from the culture by centrifugation at 1000 rpm for 5 minutes and 50 µL was transferred to GC-MS vials for further analysis.

5.2.5 Biomass quantification

To quantify biomass, dilutions of the cultures were done in cuvettes and the sample was measured at 600 nm using a NanodropTM 2000c spectrophotometer. The result was later multiplied by 1000 and divided by the number of µL used for the dilution to finally obtain the original OD.

5.2.6 Taxanes detection by Gas Chromatography-Mass Spectrometry (GC-MS)

Taxane identification and quantification were done by GC-MS. Out of the 50µL of dodecane, only 1µL was injected into the TRACE 1300 Gas Chromatograph (Thermo Fisher Scientific, UK) which was coupled to an ISQ LT single quadrupole mass spectrometer by the same company. The column used was Thermos Scientific TG-5MS with 30m x 0.25mm x 0.25um dimensions. Settings were as follows: Injector temp 250ºC, split less injection, split flow 10ml/min, split ration 33.3, carrier flow 2ml/min. The initial GC temperature was 120ºC and held for 1 minute, then increased up to 20ºC/min until a temperature of 300ºC was achieved, and the sample was held for 3 minutes.

To identify and quantify the production of taxanes standard concentrations of taxadiene and GGOH were related to the peak area for each product and concentrations were calculated. To distinguish a peak from the background noise, the mass spectrum for each peak was analysed. Also, only peaks with three times the base high were considered as true compound peaks. The pure standards of taxadiene and GGOH were previously obtained by a donation from the Baran Lab (The Scripps Research Institute, USA) and Sigma Aldrich, respectively. Xcalibur™ Software from

Thermo Fisher Scientific was used to analyse GC-MS metabolites and measure peak areas for identification and quantification.

5.2.7 Statistic analysis

All the statistical analyses were done by using the Excel Analysis ToolPak. All experiments were made in triplicates (n=3) unless specified.

2 different hypotheses were formulated. The hypothesis null was that changing the inoculation rate did not have a significant change in production of T1βOH while hypothesis alternative indicated that changing the inoculation ratio had a significant change in T1βOH production.

First, a F-test (Fisher's test) was necessary to determine if the variances were equal or unequal.

F-Test Two-Sample for Variances

	1:1:2	*1:1:6*
Mean	3.69	44.9967521
Variance	1.3579	18.1507297
Observations	3	4
df	2	3
F	0.074812419	
P(F<=f) one-tail	0.070405294	
F Critical one-tail	0.052180378	

Following this, a T-test assuming unequal variances was done to verify if there was a significant difference in the production of the new compound T1βOH when inoculation ratios changed from 1:1:2 to 1:1:6.

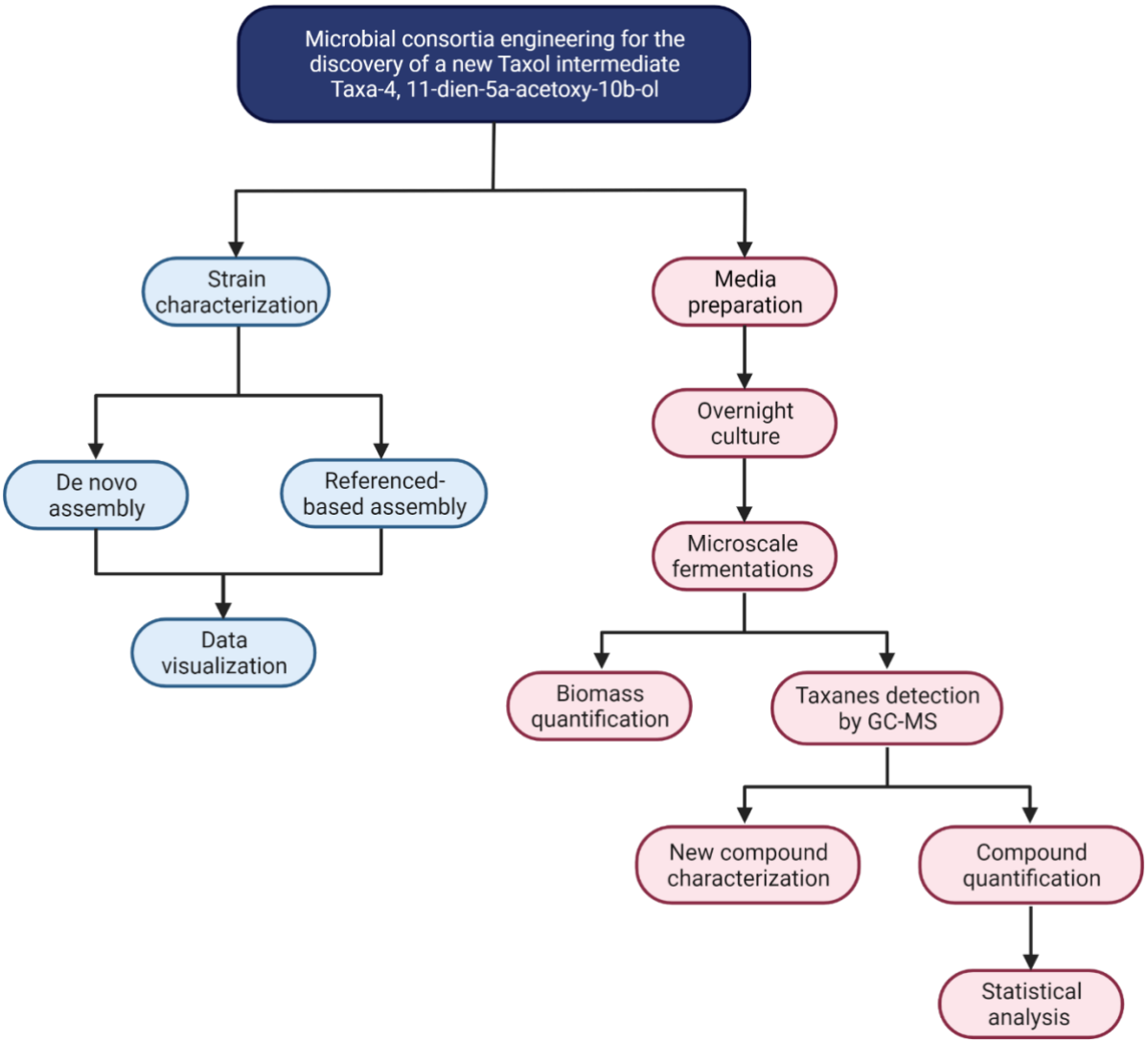
t-Test: Two-Sample Assuming Unequal Variances

	1:1:2	*1:1:6*
Mean	3.69	44.9967521
Variance	1.3579	18.1507297
Observations	3	4
Hypothesized Mean Difference	0	
df	4	
	-	
t Stat	18.49085683	
P(T<=t) one-tail	2.51694E-05	
t Critical one-tail	2.131846786	
P(T<=t) two-tail		
t Critical two-tail	2.776445105	

An Alpha value of 0.5 was used for all statistical tests.

Finally, it is important to remark that the rationale followed in this section and for all statistical analysis has been applied in experiments of the same nature comparing different treatments and their effect on the production of natural compounds (Ai et al., 2019; Santoyo-Garcia et al., 2022)

5.2.8 Methodology summary



5.3 Results and discussion

5.3.1 Genomic characterisation of EJ2 and data analysis.

To understand better the behaviour of our parental EJ2 strain, its genome was sequenced and analysed. It is important to mention that all these genetic modifications were previously confirmed by PCR and sanger sequencing. However, using Next Generation Sequencing (NGS) tools would allow us to have a much higher understanding of the complete genomic background of our strain.

Apart from the insertions of paclitaxel genes (TASY, T5αOH, TAT and T10βOH) and engineered mevalonate pathway, we also aimed to verify the correct deletion of GAL80 and to try to clarify the unsuccessful attempt of Mig1p deletion.

According to the reports in the literature, 3 different TASY cassettes were inserted in the parental strain of EJ2 (referred to as LRS5) (Figure 5.2) (Nowrouzi et al., 2020).

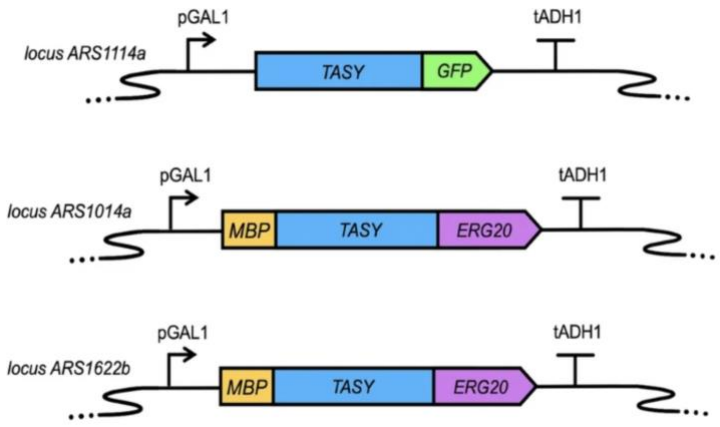
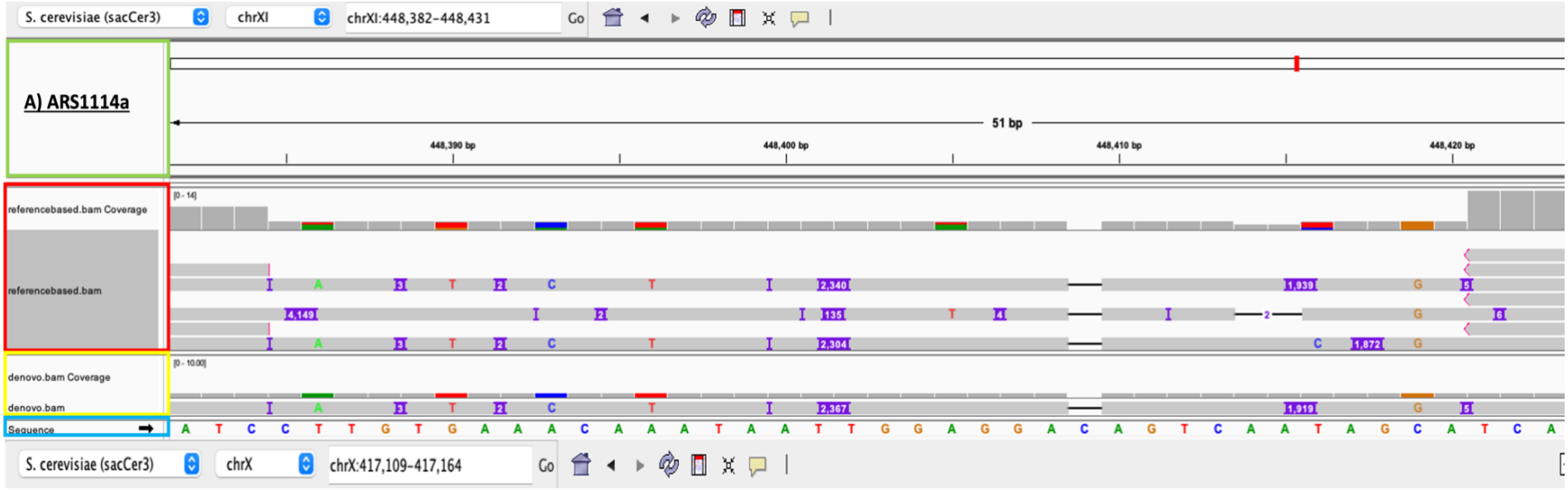


Figure 5.2 TASY cassettes and its locis. 3 different cassettes were integrated into loci ARS1114a, ARS1014q and ARS1622b. The different promoters and terminators and the different tags for each cassette are detailed in the image. The image is taken from (Nowrouzi et al., 2020)

Based on this, an approximate total base pair (bp) was calculated for the inserts giving a total of 3970 bp for TASY ARS1114a, and 5285 bp for TASYs ARS1014a and ARS1622a (Table 5.2). Once the alignment files were visualised in IGV (Figure 5.7), it was possible to see the fragments inserted (purple blocks signalling the number of

bases inserted). The number of bases signalled inside blocks was added with the rest of the blocks within the same alignment (horizontal lines) and the total was compared to that specified in the literature (Table 5.1). The different TASY insertions are shown in Figure 5.3. A colour code is used to help describe each of the elements in the figures. The code is detailed as follows. Green: gene name. Red: coverage of the alignment and alignments by the different contigs generated by the reference-based assembly. Yellow: sequenced generated by de novo assembly and alignment. Blue: reference sequence.


S. cerevisiae (sacCer3)
chrXI
chrXI:448,382-448,431
Go
A) ARS1114a
51 bp
448,390 bp
448,400 bp
448,410 bp
448,420 bp
referencebased.bam Coverage
referencebased.bam
denovo.bam Coverage
denovo.bam
Sequence
S. cerevisiae (sacCer3)
chrX
chrX:417,109-417,164
Go

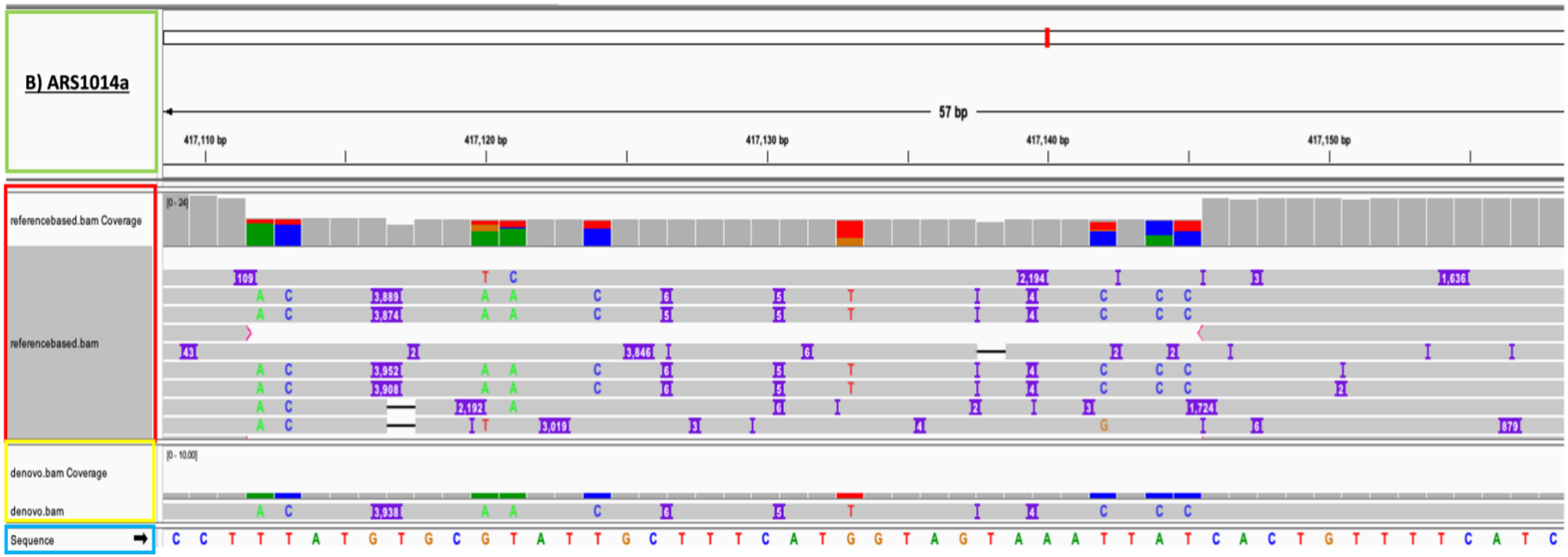

B) ARS1014a
57 bp
417,110 bp
417,120 bp
417,130 bp
417,140 bp
417,150 bp
referencebased.bam Coverage
referencebased.bam
denovo.bam Coverage
denovo.bam
Sequence
C C T T T A T G T G C G T A T T G C T T T C A T G G T A G T A A A T T A T C A C T G T T T T C A T C

Figure 5.3. Comparison of the different TASY genes in the EJ2 strain. Squares of different colours have been used to indicate the different elements in the figure. Green: gene name. Red: coverage of the alignment and alignments by the different contigs generated by the reference-based assembly. Yellow: sequenced generated by de novo assembly and alignment. Blue: reference sequence. A) Shows TASY inserted in ARS1114a while B) shows TASY insertion in ARS1014a and C) shows TASY insertion in ARS1622b.

By subtracting the total bases (reported in the literature) from the total detected (by NGS) some differences were noted in the total length of the inserts. The most obvious differences were for TASY inserted in ARS1114a (Figure 5.3 A) and ARS1014a (Figure 5.3 B) with a total difference of 320 and 1335 bases. When trying to explain this difference, we found some discrepancies in the literature about the structure of the constructs where in one report the insert ARS1114a was described as pGAL1-TASY-GFP-tADH1 and ARS1014a as pGAL1-MBP-TASY-ERG20-tADH1 (Nowrouzi et al., 2020) and in another ARS1114a was described as TDH3p-MBP-TASY-ERG20 and ARS1014a as pGAL1-TASY-GFP (Walls et al., 2021). By analysing the difference in bases between the reported constructs and the results from NGS, it is more likely that the constructs inserted in ARS1114a and ARS1014a were configured as reported by Laura E. Walls et al., (2021). This conclusion highlights the potential of NGS for strain characterisation.

Overall, the chromosome location of the genes was the same for all TASY cassettes as reported in the literature. However, the insert located in ARS1622b (Figure 5.3 C) was not detected by the reference-based assembly and it was only possible to see the insertion when visualising the alignment resulting from the de novo assembly (the insert resulted in a difference of only 132 bases). This stresses the importance of combining different approaches for data analysis, something that can be backed by previous works (Lischer & Shimizu, 2017; Lu et al., 2013).

In all cases, the difference in length of the inserts reported vs detected may be due to inherent errors of the sequencing, but it could also be resulting from homologous recombination errors during gene insertion. Both scenarios could be verified by optimizing the sequencing parameters to get an even better data quality and perform a deeper analysis. A parameter that can be easily optimised is the quality and quantity of the genomic DNA used. Another strategy for optimising the results is choosing a better basecalling algorithm (which would require more computer power).

The same analysis was repeated for T5αOH+CPR, TAT, T10βOH and GAL80. According to the literature, T5αOH+CPR and TAT were inserted as shown in Figure 5.4.

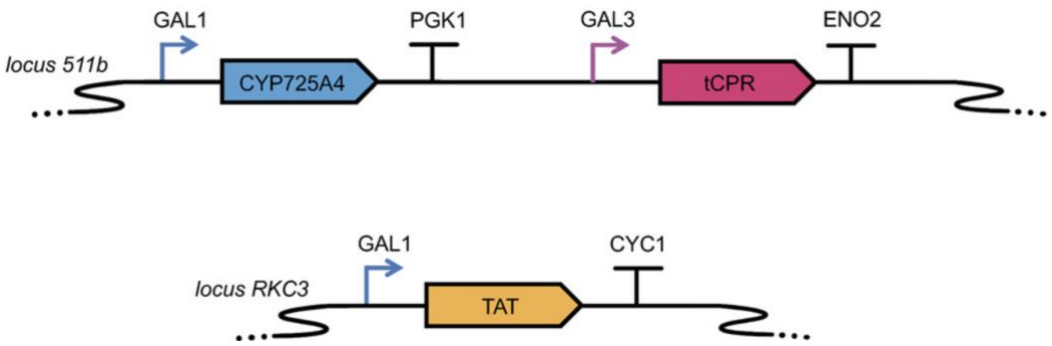
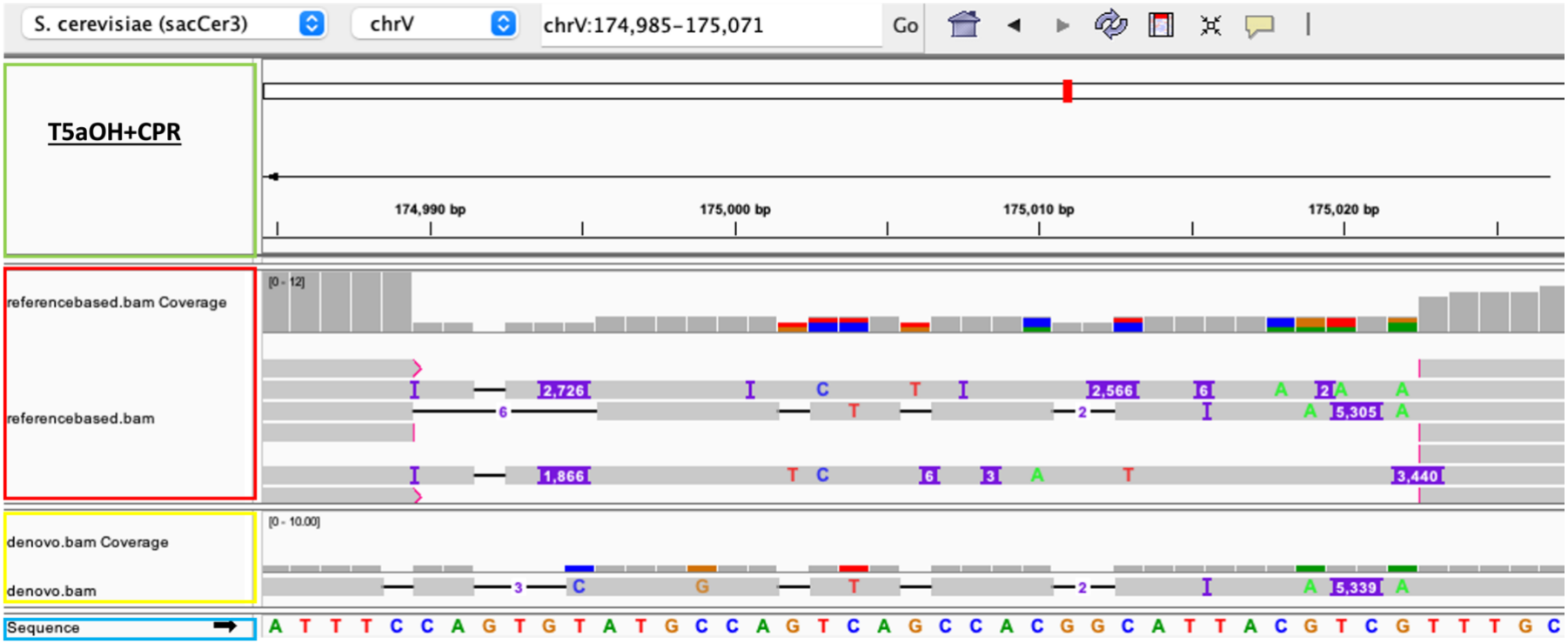


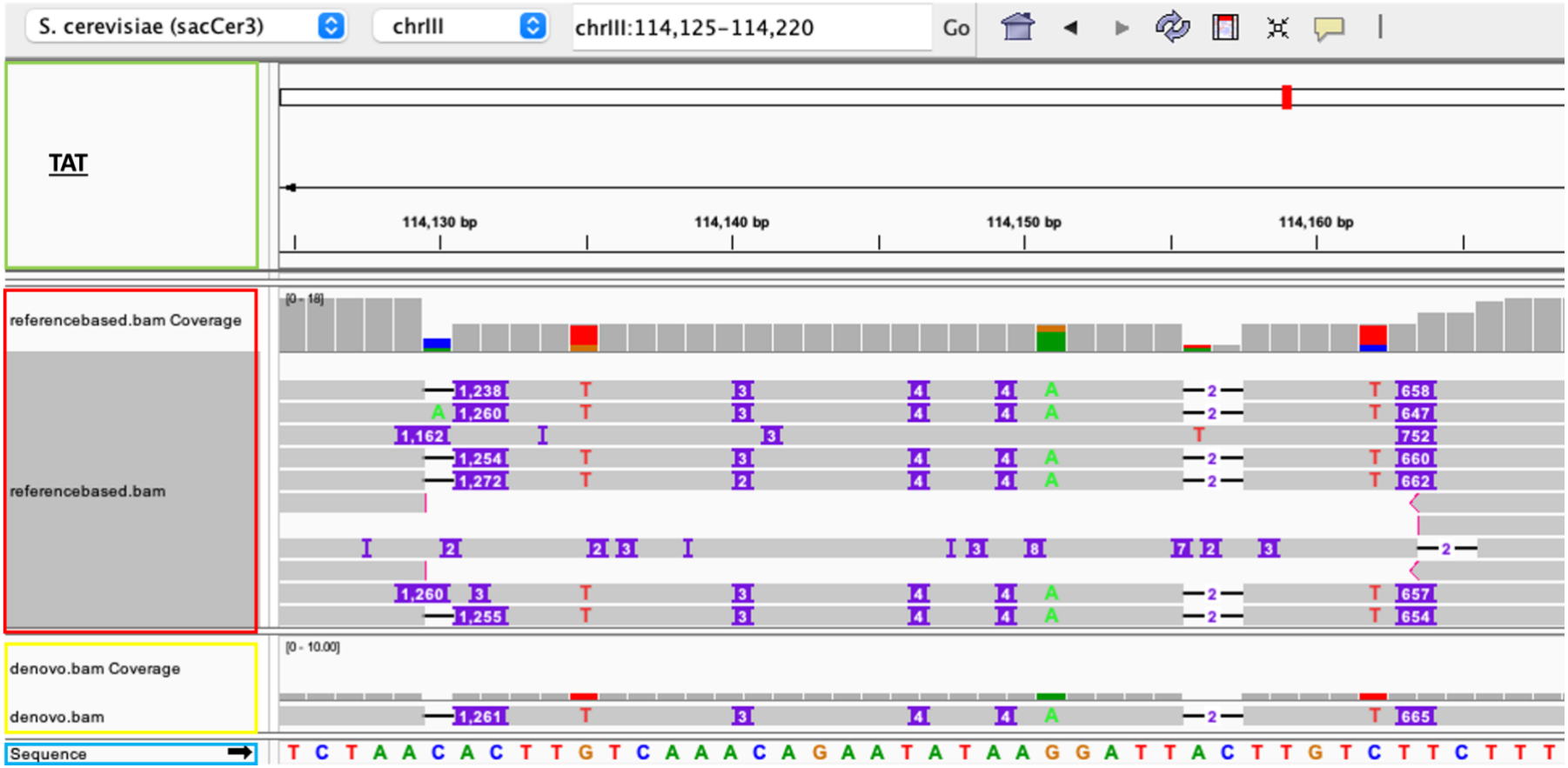
Figure 5.4. Cassettes for T5αOH+CPR and TAT and its locis. T5αOH+CPR The different promoters and terminators and the different tags for each cassette are detailed in the image. The image is taken from (Walls et al., 2021).

Based on figure 5.4, the approximate total base pair (bp) for T5αOH + CPR was 5208 while for TAT it was 1964. Once again, the number of bases signalled inside purple blocks for insertion was added with the rest of the blocks within the same alignment (horizontal lines) and the total was compared to that specified in the literature (Table 5.1). In the image below (Figure 5.5), T5αOH + CPR, TAT and GAL80 deletion are shown.

A


S. cerevisiae (sacCer3)
chrV
chrV:174,985-175,071
Go
T5aOH+CPR
174,990 bp
175,000 bp
175,010 bp
175,020 bp
referencebased.bam Coverage
[0 - 12]
referencebased.bam
2,726
2,566
5,305
1,866
3,440
denovo.bam Coverage
[0 - 10.00]
denovo.bam
5,339
Sequence
A T T T C C A G T G T A T G C C A G T C A G C C A C G G C A T T A C G T C G T T T G C

B


S. cerevisiae (sacCer3)
chrIII
chrIII:114,125-114,220
Go
TAT
114,130 bp
114,140 bp
114,150 bp
114,160 bp
referencebased.bam Coverage
referencebased.bam
denovo.bam Coverage
denovo.bam
Sequence
T C T A A C A C T T G T C A A A C A G A A T A T A A G G A T T A C T T G T C T T C T T T

C

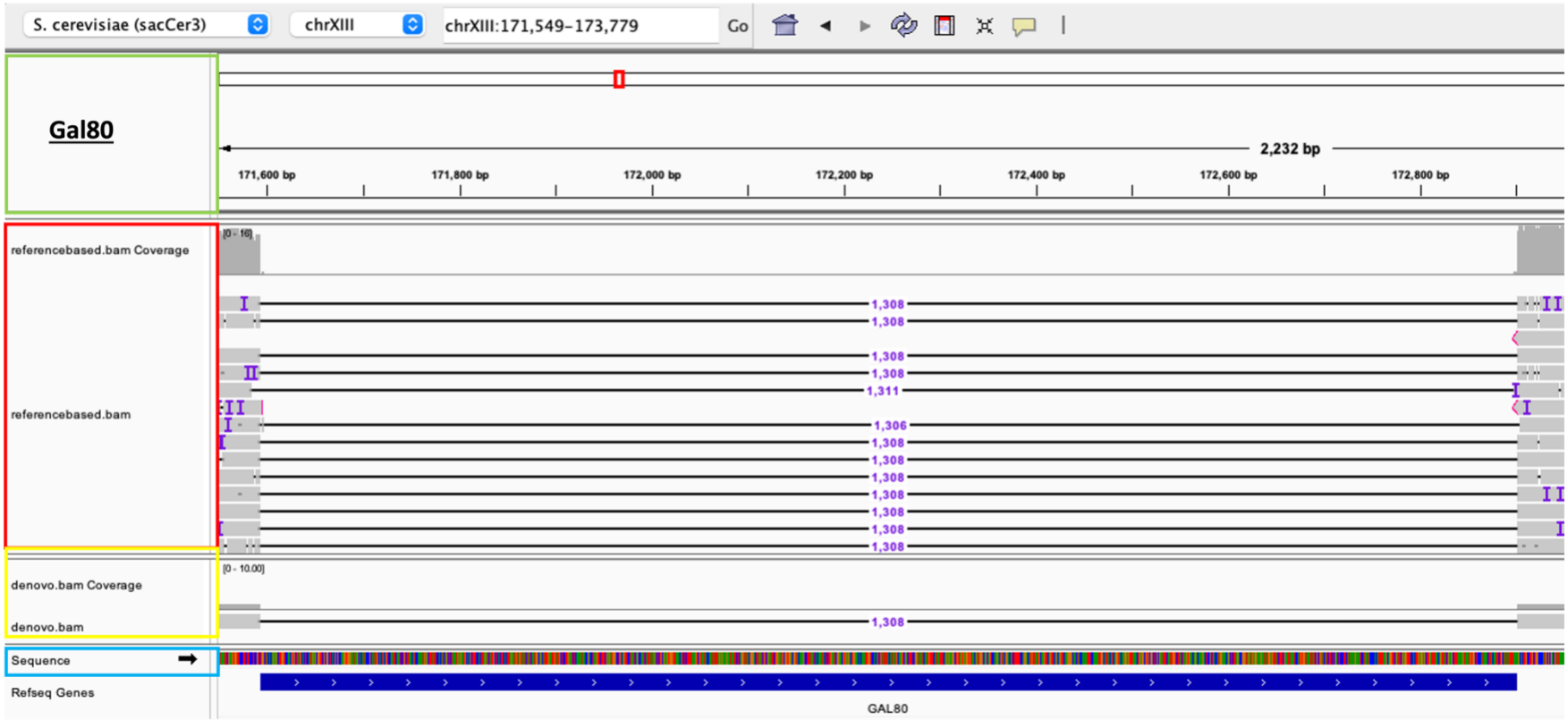


Figure 5.5. Comparison of the different paclitaxel genes in the EJ2 strain. Squares of different colours have been used to indicate the different elements in the figure. Green: gene name. Red: coverage of the alignment and alignments by the different contigs generated by the reference-based assembly. Yellow: sequenced generated by de novo assembly and alignment. Blue: reference sequence. A) Shows the insertion of T5αOH in chromosome V while B) shows the insertion of TAT in chromosome III and finally C) shows GAL80 deletion.

The length detected for T5αOH was very similar to that reported in the literature (with a difference of only 92 bases) and no changes were found in the gene location (Figure 5.5 A). On the contrary, TAT was not found in the expected location of Locus RKC3 Chr III: 119,278–119,313. Interestingly, a deeper analysis revealed another insertion in the same chromosome but in between bases 114,130-114,164 (Figure 5.5 B). The total length of this insert was approximate ~1937 which represents a difference of only 27 bases compared to the expected length size of TAT.

One of the disadvantages of the methodology used is that insertions are only visualised as the number of bases inserted, while the exact bases inserted are unknown. An approach to solve this may be to manually insert the different cassettes into the reference sequence. By doing this when the alignment is done, the sequences should match and any changes in the sequence may be easier to detect with more detail. This strategy could be implemented and tested in future experiments. Furthermore, NGS is a very versatile technology that can be greatly optimised depending on the objectives of the experiment. For the objectives of our research, the methodology followed was enough to detect the genes and their locations. However, some other protocols can detect even single nucleotide variants with a high degree of confidence (Koboldt, 2020).

Lastly, Integrative Genome Viewer indicates the deletions with respect to a reference with a black bar labelled with the number of bases deleted (Figure 5.5 C). Based on this, the total number of bases deleted was 1308, which corresponded to the length of the GAL80 gene, confirming a successful knockout (Figure 5.6 C).

Finally, Table 5.3 summarizes all the information discussed in the previous paragraphs.

Table 5.3 Comparison of the different genes inserted and deleted in EJ2. The different genes are listed along with their respective parts. The length of the parts is listed to get an estimated total of base pairs of the sequence total bp. An estimate of the insertions detected is also included for comparison against the real bp in each gene. Both the location of the genetic modification and the location detected in NGS is also included.

Gene	Parts		Expected length (bp)	Total (bp)	Total detected	Expected location	Location detected	Reference
TASY	Promoter	pTDH3	600	3970	~4290	ARS1114a ChrXI: 448,384-448,404	Same	(Nowrouzi et al., 2020; Walls et al., 2021)
	CDS	TASY	2409					
	CDS	GFP	711					
	Terminator	tADH1	250					
TASY	Promoter	pGAL1	457	5285	~3950	ARS1014a ChrX: 417,111-417,131	Same	
	CDS	MBP	1113					
	CDS	TASY	2409					
	CDS	ERG20	1056					
	Terminator	tADH1	250					
TASY	Promoter	pGAL1	457	5285	~5427	ARS1622b Chr XVI: 565,423-565,443	Same	
	CDS	MBP	1113					
	CDS	TASY	2409					
	CDS	ERG20	1056					
	Terminator	tADH1	250					
T5aOH + CPR	Promoter	pGAL1	457	5208	~5300	Locus 511b Chr V: 175,892–175,911	Same	
	CDS	T5aOH	1500					
	Terminator	PGK1	250					
	Promoter	pGal3	600					
	CDS	CPR	2151					
	Terminator	ENO2	250					
TAT	Promoter	pGAL1	457	1964	~1937	Locus RKC3 Chr III: 119,278–119,313	ChrIII: 114,130-114,164	
	CDS	TAT	1317					
	Terminator	CYC1	190					
T1BOH	Promoter	pGAL1	457	2141	Not detected	ARS1531 Chr XV: 33,841-33,860	Not detected	This study
	CDS	T10BOH	1494					
	Terminator	CYC1	190					
GAL80	GAL80		1308		Not detected	Chr XIII: 166,382-167,689	Deleted	

Despite the previous confirmation of T10βOH insertion by PCR and sanger sequencing in chapter 3, a thorough inspection of chromosome XV did not reveal any presence of the gene. The absence of this gene may explain why despite attempts for

production optimization, the product of the enzyme T10βOH was never detected in EJ2 cultures (Figures 3.9, 3.10 and 3.11) and was only detected when the gene was directly cloned in a multicopy plasmid (Figures 4.9, 4.10, 4.11 and 4.13). Further studies are needed to clarify why only this gene was lost. Any future experiment aimed to discover the next steps of the pathway should consider this and use a plasmid with the T10BOH gene.

As mentioned in Chapter 3, there was an attempt to delete the Mig1 gene to eliminate carbon catabolite repression in the presence of glucose and as consequence increase the production of taxanes in this carbon source. From the donor parts to be used for CRISPR/CAS9 deletion, the downstream homology arm was successfully amplified. On contrary, the efforts to amplify the upstream homology arm were futile as described in Chapter 3. The NGS analysis revealed the presence of 3 genes different genes very close to each other (Figure 5.6). From upstream to downstream, these genes are YGL036W, Mig1 and YGL034C. By using BLAT (blast-like alignment tool) from IGV, it was interesting to notice that the homology arms designed as donor DNA overlap with YGL036W (purple box) and YGL034C (black box). Both genes were researched in the literature and yeast databases

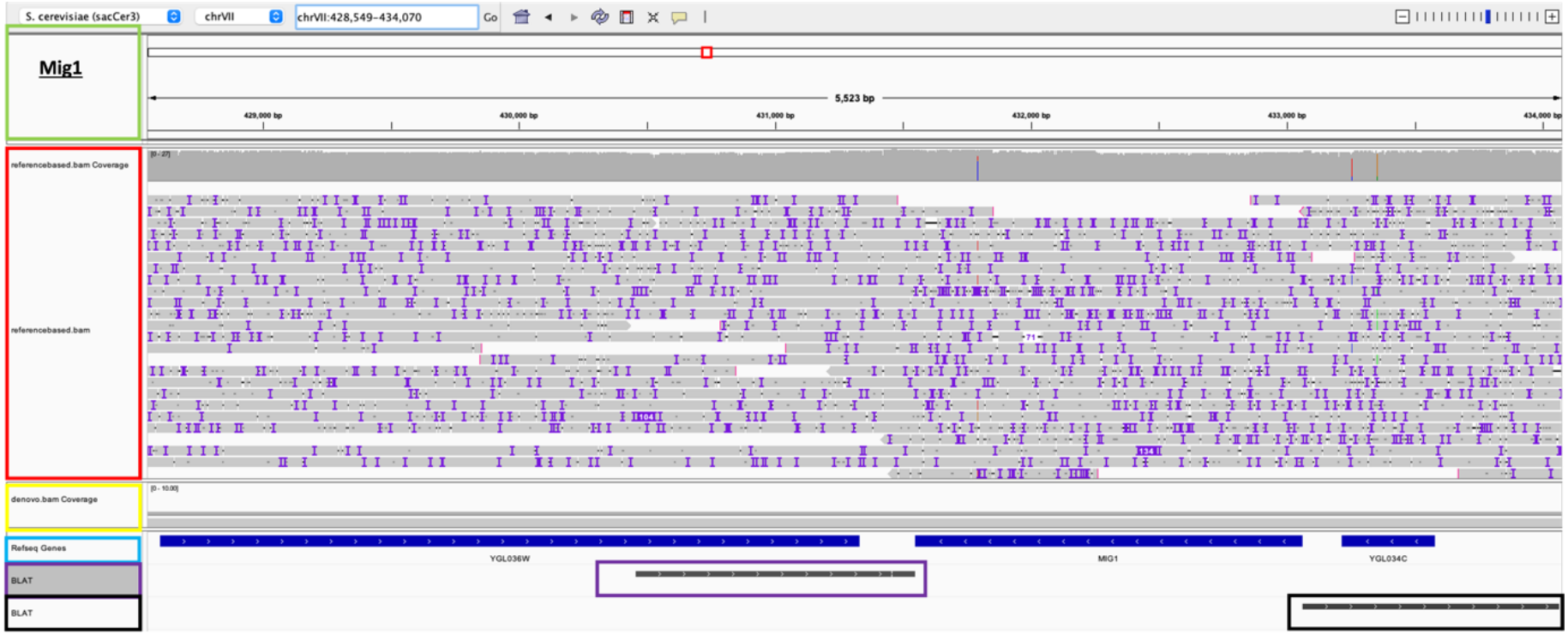


Table 5.6. Location of Mig1 in *S. cerevisiae* Chromosome VII. Squares of different colours have been used to indicate the different elements in the figure. Green: gene name. Red: coverage of the alignment and alignments by the different contigs generated by the reference-based assembly. Yellow: sequenced generated by de novo assembly and alignment. Blue: reference sequence including the genes. Purple: 1000bp upstream homology arm. Black: 1000bp downstream homology arm.

A quick search in the *Saccharomyces* genome database (SGD) revealed that YGL036W does not have a known function and is described as a non-functional gene. However, upon further research, a preprint article describing this gene was found (Ensinck et al., 2023). The new information provided describes YGL036W as part of the yeast m6A methyltransferase complexes (MTC). The same study describes MTC in yeast as critical for meiosis. Although the information already provided for YGL036W gives more context about the gene function it still doesn't explain why PCR amplification attempts were unsuccessful. It has been reported before that DNA sequences with regions of less than 30% and more than 70% GC content tend to form secondary structures that not only block the DNA polymerase but also cause mispriming and mis-annealing of the primers (Guido et al., 2016). By looking at the designed homology arms, it was noted that DHA had a GC content of 43.36% while for UHA it was 32.66% which is very close to the 30% threshold mentioned. This difference in GC content between DHA and UHA might explain why the former was easily amplified and more importantly why the attempts to amplify the latter were unsuccessful. Finally, NGS only helped to confirm that the reasons behind the unsuccessful deletion of Mig1p did not involve any complex regulatory elements and it was more likely explained by the low GC content.

For YGL034C –overlapping with the DHA— the information is scarcer than for the previous gene described. A search in SGD described this gene as a non-essential gene coding for a putative protein without a known function and conserved among *S. cerevisiae* strains. Furthermore, a report from 2013 refers to this protein as a "Dubious open reading frame unlikely to encode a functional protein" (L. Zhang et al., 2013). This potential lack of function along with its balanced GC content of 43.36% and, its length is significantly lower than YGL036W (363bp vs 2727bp) can also explain why –with unlike UHA— this gene was easily PCR amplified.

5.3.2 GC-MS analysis for T1βOH product detection and characterization

It is very important to remark that the next part of the project was done in collaboration with Ainoa Escrich from Universitat Pompeu Fabra during a small internship with Leo Rios Lab at the University of Edinburgh's Institute for Bioengineering. Part of her project consisted of the assembly of a new yeast plasmid carrying the sequence for

the candidate gene TB574 encoding for T1βOH, previously obtained from a cDNA-AFLP experiment (Ramírez-Estrada et al., 2016) and codon optimised for expression in *S. cerevisiae KM strain* (described in the methodology section). The construction of the plasmid was done by using the molecular cloning kit (MoClo-YTK) designed by John Dueber's Lab and it was acquired from Addgene (kit#1000000061). Yeast transformation resulted in the strain AE1 which carried the newly designed plasmid with T1βOH and plasmid plr49 –carrying T10βOH— from Leo Rios Lab repository. Both plasmid maps can be found in appendix 5 C. It is important to remark that the only difference between plr49 and plasmid inserted in EJ2-T10βOH (appendix 4 F) is the replacement of the Leu2 in the former by the Ura3 selection marker in the latter. The work carried out by Ainoa resulted in the construction of strain AE1 which was kindly donated for the purposes of this collaboration project. The details of the plasmid construction can be found in (Escrich-Montana, 2022), however, at the moment of writing this dissertation the work cited had a 12-month embargo.

Based on the previous results shown in Chapter 4 (Figures 4.12 and 4.13) *Saccharomyces cerevisiae* EJ2-TAT, EJ2-T10βOH and AE1 were grown in co-cultures to enhance taxane precursors production for the taxadiene-5a-acetoxy-1α,10β-diol. Two different inoculation ratios and initial ODs were used for EJ2-TAT/EJ2-T10βOH/AE1 consortia. First, a consortium with a 1:1:2 (initial ODs were 1, 1, 2) ratio was evaluated. Following the results discussed in the last chapter (Figures 4.12 and 4.13), the fermentation conditions for these experiments were done in YP media with galactose in 24 well plates at 30°C for 3 days. Since there are no reports of the mass spectrum for the T1β-diol or its retention this time, our approach was to compare the resulting chromatogram to another control consortium lacking the strain with the T1βOH (Figure 5.7 A). In this manner, we would observe the presence of new peaks potentially corresponding to the new compound T1β-diol. The comparison sparked a new peak that was detected at a retention time of 9.76 min (Figure 5.7 B) and the mass spectrum revealed a similar pattern to previously characterized taxanes. To confirm whether the new peak corresponded to a new compound, the consortium was optimized by adjusting the inoculation ratios to 1:1:6 (ODs of 0.125, 0.125 and 0.75). The same peak in the chromatogram was observed this time at 9.75 and with

more intensity (Figure 5.7 C). The comparison between the control, the 1:1:2 and 1:1:6 consortia are shown in Figure 5.7.

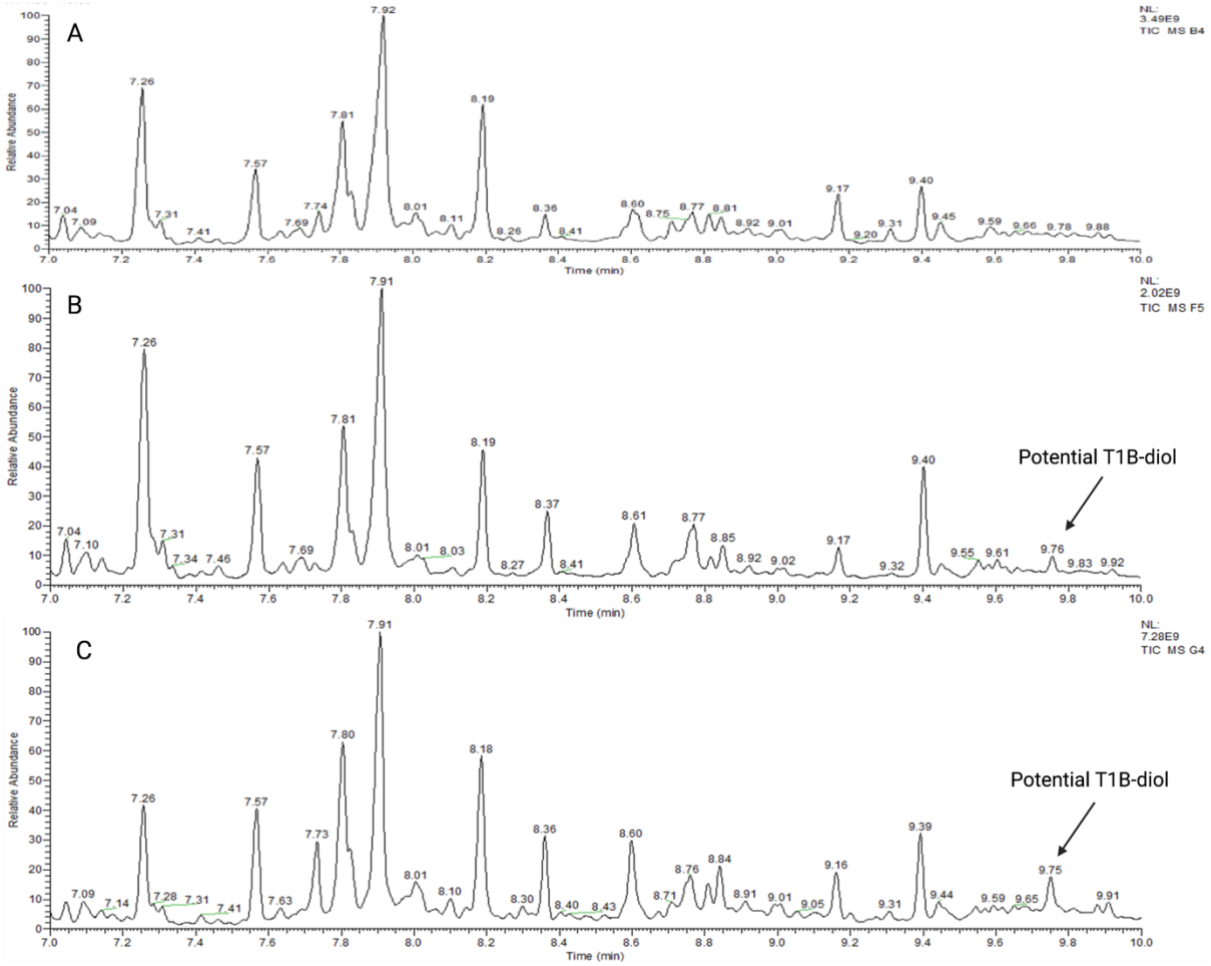


Figure 5.7. Chromatogram and mass spectrum of a consortium of 3 different *S. cerevisiae* strains. A) No peak is detected after 9.66. This chromatogram corresponds to the control without T1βOH. B) The peak corresponding to the new potential compound can be seen at 9.76. The chromatogram corresponds to the consortia with a 1:1:2 inoculation ratio. C) The peak corresponding to the new potential compound can be seen at 9.75. The chromatogram corresponds to the consortia with a 1:1:6 inoculation ratio.

The most relevant peaks in the chromatograms are explained as follows. In all panels, the first important peak at 7.26 corresponds to the precursor taxadiene while at 7.91±0.01 the side product GGOH can be observed. The compounds T5α-ol, T5αAc and T10β-ol are observed at 8.18±.01, 8.60±0.1 and 9.44±.01 respectively. Finally, the peak potentially corresponding to T1β-diol can be observed at a retention time of 9.75±.01 for panels B and C while it is missing in the control in panel A. Next, analysed the mass spectrum of the compound was at 9.76 min. Figure 5. 8 shows the mass spectrum for the potential peak of T1βOH (only a fraction of the mass spectrum

highlighting the molecular mass of the new product is shown here, the full spectrum is shown in appendix 5 D).

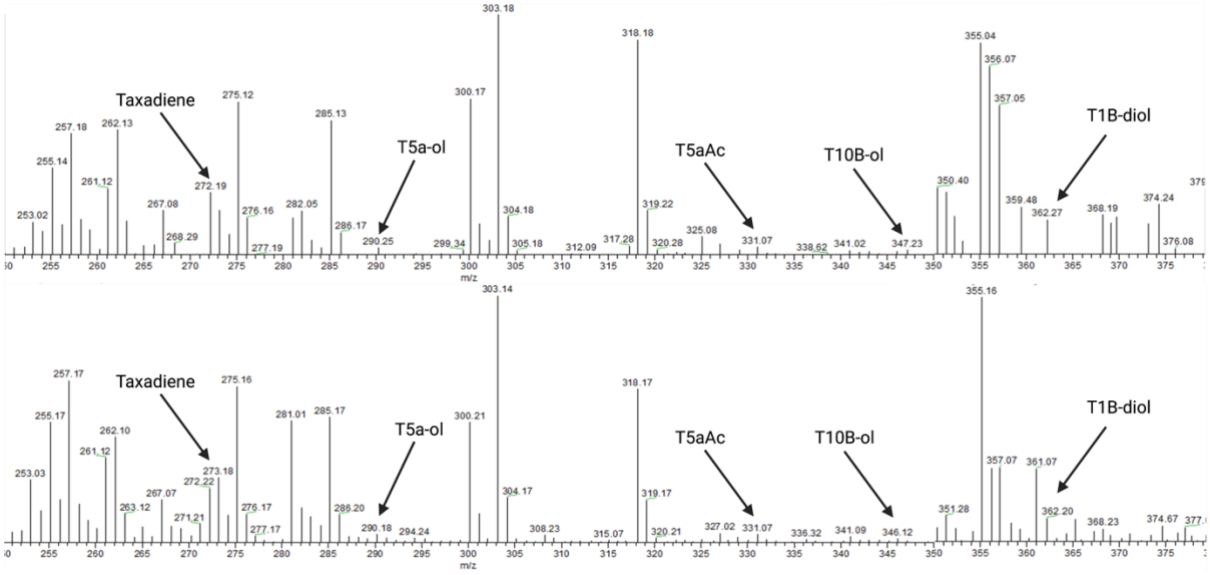


Figure 5.8 Mass spectrum of the different consortia for the potentially new T1β-diol compound. Both spectrums show many similarities. The top corresponds to a 1:1:2 inoculation ratio while the bottom corresponds to 1:1:6.

To determine if the above spectrums correspond to the new compound the first thing to consider is the molecular weight of T1β-diol. This product has a molecular weight of 362. However, as the molecule gets ionised, it can break and lose some of its functional chemical groups. Table 5.2 summarizes the different chemical groups added through each of the steps in the paclitaxel metabolic pathway. These groups can be lost during ionisation in the GC-Ms and provide a unique “fingerprint” to each molecule. The 3 different hydroxyl groups (OH) are added by the different P450s (T5αOH, T10bOH and T1βOH) while the acetate group (Ac) is added by TAT (Figure 5.1).

Table 5.4. Different paclitaxel precursors detected by GC-MS. The table shows the different precursors, their formula, molecular weight, and the different chemical groups linked to the paclitaxel skeleton. As some of these functional groups are lost during ionisation the total mass can decrease making it possible that a specific compound shows different peak patterns.

Compound	Formula	Molecular weight	Nº OH	Nº Ac
Taxadiene	C20H32	272	-	-
T5α-ol	C20H32O	288	1	-
T5αAc	C22H34O2	330	-	1
T10β-ol	C22H34O3	346	1	-
T1β-diol	C22H34O4	362	2	-

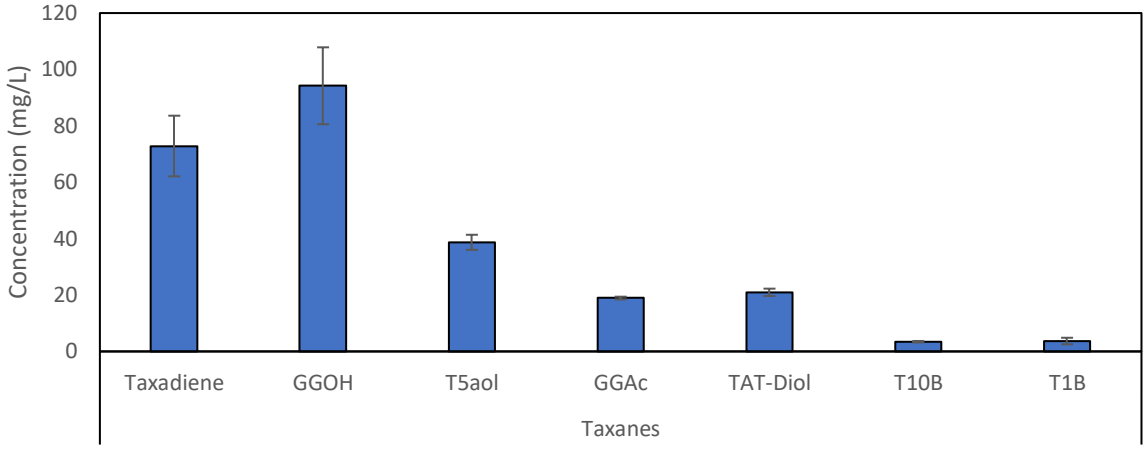
As the molecule breaks and loses some of these groups, the molecular weight of the compound also decreases. For example, when T1β-diol loses 1 OH (with a molecular weight of 16) its new molecular weight will be 346 (the same as T10β-ol) if it loses another then it will be 330 (as T5αAc) and so on. Hence, the detection of peaks at 346 ±1 and 330 ±1 is a good indicator of the presence of this compound. Furthermore, the loss of an acetate group with a molecular weight of 59 would indicate a molecular weight of 303, which also corresponds to the peaks in Figure 5.8. The evidence gathered by GC-MS suggests that the putative gene TB574 codifies for T1βOH, the next enzyme in the paclitaxel metabolic pathway. However, further experiments using nuclear Magnetic Resonance (NMR) need to be done for double confirmation of the structure of the newly discovered compound.

Finally, the discovery of the new potential compound T1β-diol is extremely relevant as it is the first time in the past 20 years that a new enzyme in the metabolic pathway is detected and functionally expressed for the conversion of its precursor since T10βOH (Schoendorf et al., 2001). Furthermore, this result could be very relevant to continue the discovery of new enzymes of precursors in the paclitaxel metabolic pathway.

5.3.3 Taxanes and T1βOH quantification

After the fermentations and detection of the compound described in the previous section, the highly potential compound T1β-diol was quantified using the same methodology described in Chapter 2 section 2.12. Figure 5.9 shows the production of different the different paclitaxel precursors including the newly characterised T1β-diol by the different consortia described in the last section.

A)



B)

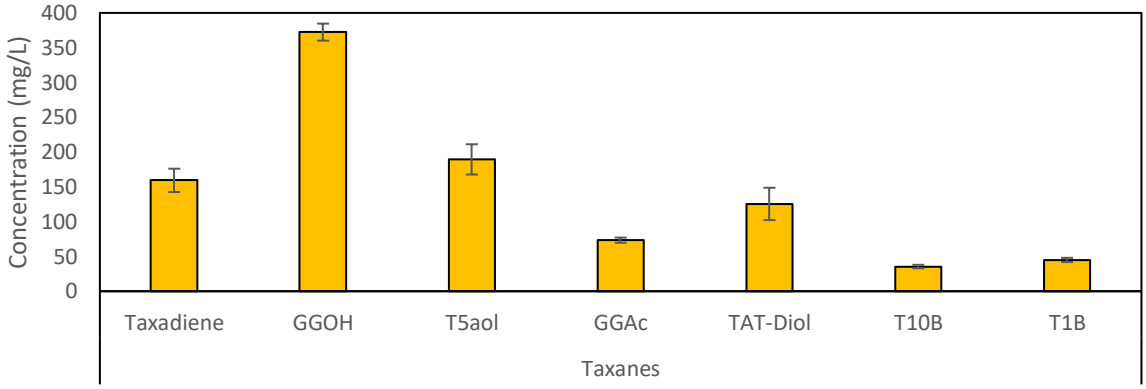


Figure 5.9. Production of different taxanes by a consortium of 3 different S. cerevisiae strains grown in YP media. The strains used were, EJ2-TAT, EJ2-T10βOH and AE1 and were present at a 1:1:2 ratio and 1:1:6 ratio for the first and second consortiums respectively. The experiment was done in 2% galactose in 24 well plates for 3 days at 30 °C degrees at 350 rpm, Error bars represent the standard deviation of the replicates (A, n=3 while B, n=4). A t-test comparing T1B production for A and B confirmed that the inoculation ratio affected the production of T1B (P value 0.05> 0.00005).

In the consortium with a 1:1:6 ratio (consortium B) and lower initial OD of 1 (Figure 5.98 A), the total amount of taxane products (not including the by-products GGOH and GGAc) was 308% higher than in the consortia with a 1:1:2 (consortium A) ratio and

higher initial OD of 4 (Figure 5.9 B). The amount of taxadiene produced by consortium B was 159± 17 mg/ L which represents an increase of 118.8 % more when compared to 73± 11 mg/L obtained by consortium A. The total production of the acetylated compounds (GGAc + T5αAc) in consortium B was 199 mg/L while for consortium A this was 40 mg/L, representing a difference of 398 %. The production of the precursor T10β-ol was also significantly lower in consortium A as the concentration was 3.5 ±0.2 mg/L while in consortium B the production was 35.5 ±2.5 mg/ L which represents a significant increase of 931%. Finally, it is also worth comparing the difference in the production of T1β-diol between both consortiums. As with the other compounds, the production was higher in consortium B with 45 ±3 mg/L versus 3.7 ± 1.2 mg/L generated by consortium A. A t-test assuming unequal variances confirmed that the inoculation ratio affected the production of the novel T1β-diol.

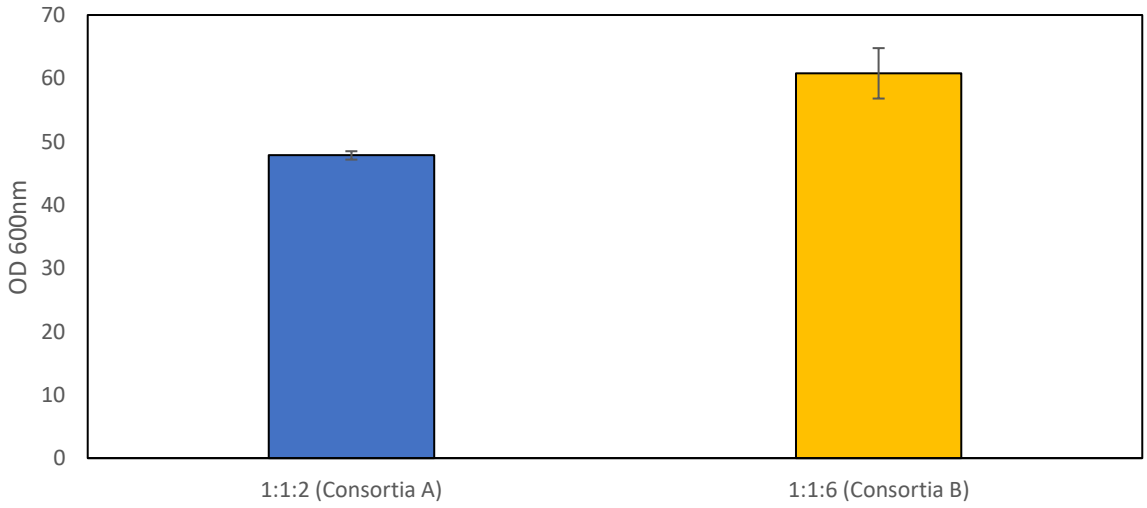


Figure 5.10 OD at 600nm corresponding to the consortiums indicated previously. The experiment was done in 24 well plates for 3 days at 30°C and 350 rpm. n=3 for 2:1:1 and n=4 for 6:1:1. Error bars represent the standard deviation of the replicates (no statistical analysis was done on this data).

The total biomass showed some interesting results as unexpectedly the consortium with the lowest initial OD (consortium B) ended up having increased biomass over the consortium with the highest initial OD (Figure 5.10). The results and production seemed to be very contrasting in both consortiums. Since the only variations between both consortiums were purely the ratio of inoculations of each strain and initial OD600, this highlights the important effect of the initial ratio and OD600.

Further validation methods might be useful to be able to determine the exact mass spectrum generated by the molecule. Further isolation for standardisation might be helpful to determine if the standard used is the most appropriate for quantification or if it can be replaced with a better one.

5.3.4 Limitations

One of the biggest limitations in this chapter is that the experiment was designed based on the results obtained in chapter 4, however, it would've been beneficial to perform a statistical design of experiments to optimise the inoculation ratio of each specie and maximise production of T1B-ol.

Another limitation is that despite the characterization of the molecule using the GC-MS equipment further validation is needed by Nuclear Magnetic Resonance (NMR) to confirm that the hydroxylation was made in the desired carbon (1β) and not in any other carbon within the molecule.

Finally, the lack of a proper standard for quantification might cause some errors when calculating the total amount of compound produced.

5.4 Conclusions

The objectives of this chapter were mainly two. The first one was to use next-generation sequencing tools to characterise our strain EJ2 and verify all deletions and insertions described in previous chapters. The second objective was also to use the microbial consortia strategy described in the previous chapters to test the conversion of the precursor T10β-ol by the new putative gene T1βOH into the product T1β-diol.

Next Generation Sequencing tools were used for the genomic characterisation of strain EJ2. Which could not reveal any presence of T10βOH insertion. This result explains the lack of production of T10β-ol despite multiple optimizations in the previous chapters. However, the rest of the heterologous enzymes inserted, and knockouts were verified in the yeast genome through the different expected chromosomes. Furthermore, the genomic analysis allowed us to understand a bit more about the reasons for the difficulties in deleting the Mig1p gene.

By using different inoculation ratios, it was possible to control the production of the precursor T10β-ol up to 35.5 mg/L. This amount was enough to achieve the conversion to a newly oxygenated taxane detected for the first time and confirm the existence of the gene TB574 as the T1βOH enzyme. The new compound T1β-diol was characterized by GC-MS and quantified reaching production values of up to 45 mg/L ± 3. It is important to mention that there is a possibility that the total amount of this compound is less than what was calculated, and that the high volume detected is the result of another compound with the same retention time being detected in the same chromatography peak used for quantification. Further experiments are needed to confirm the hydroxylation was indeed in carbon 1 of the taxadiene backbone as it can also occur that the hydroxylation of the paclitaxel skeleton was done in a different carbon.

Finally, by using a microbial consortia approach with plasmid integration and inoculum engineering, it was possible to increase the production of the T10β-ol precursor and its further conversion into T1β-diol a new compound. The production and quantification of the newly discovered compound are not only important because this is the first time that this compound is characterised, but also because high enough concentrations of it were detected which can enable the discovery of the next steps.

CHAPTER 6:

CONCLUSIONS, FUTURE WORK AND PERSPECTIVES

Chapter 6: Conclusions and future work and perspectives

6.1 Future work and perspectives

A quick search using Science Direct (https://www.sciencedirect.com) showed that since 2007, interest in engineering microbial consortia has been growing to the point at which by 2022 publication papers on this topic have increased 8-fold (Figure 6.1)

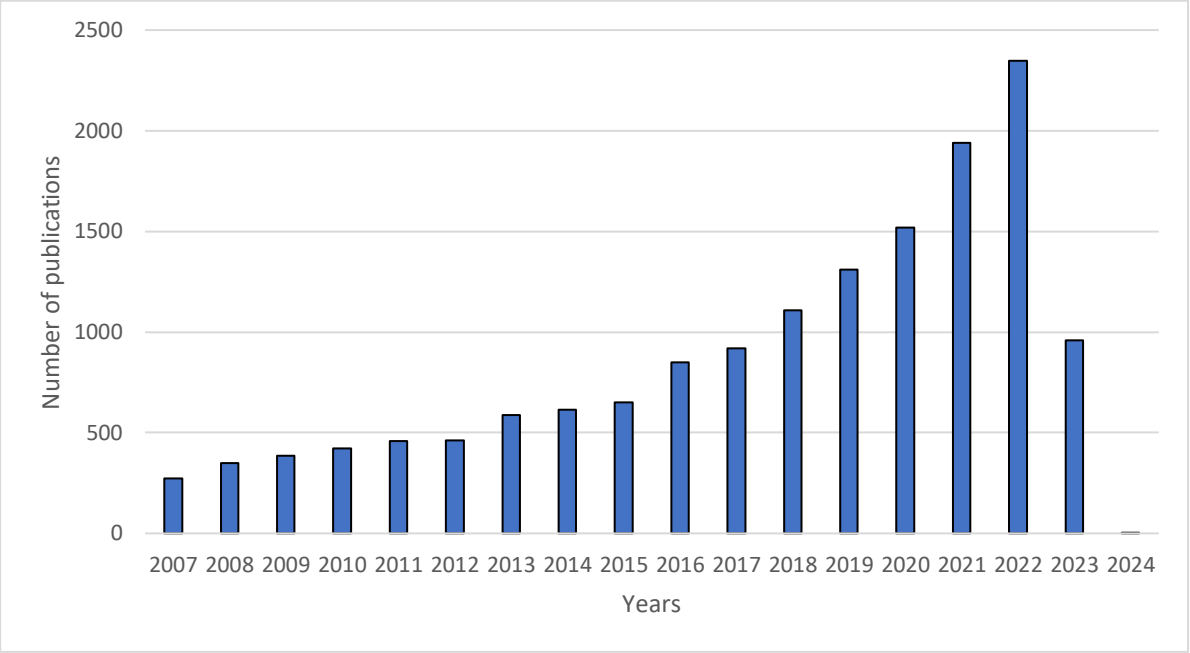


Figure 6.1. Number of publications on microbial consortia by year since 2007. The number of publications on microbial consortia has increased dramatically in the past 16 years and the trend seems to be maintained.

As systems and synthetic biology, genetics and bioinformatics develop better tools, new approaches for engineering and studying microbial consortia are arising. Furthermore, the intersection of these disciplines can enable new applications for techniques and methods already developed, thus unravelling the huge potential of synthetic microbial communities and the optimization towards their commercialisation. However, its full implementation still needs the development of new tools and the expansion of microbial chassis for strain specialization, especially for tunning populations in a more dynamic and complex manner (Sanchez et al., 2023). With the development of new genetic control mechanisms, microbial consortia promise to greatly enhance bioproduction and enable its translation into commercial settings. Finally, some of the future work in engineering synthetic microbial consortia relevant to this thesis are discussed below.

6.1.1 Scale-up fermentation of strain EJ2 using fructose

As mentioned before, GAL80 deletion enabled the production of taxane precursors in other carbon sources. Among these, fructose was one with the best results but fermentations in this sugar were only done on a micro-scale. As consequence, it would be of interest to evaluate the production of paclitaxel precursors by scaling up the fermentation using mini bioreactors. The evaluation of fructose fermentation using our *S. cerevisiae* EJ2 strain in bioreactors is very relevant because of its significantly lower price compared to other carbon sources like raffinose and galactose. A quick search on the Sigma-Aldrich website (https://www.sigmaaldrich.com/GB/en) revealed that fructose is almost 13 times lower than the price of raffinose and 10.5 lower than the price of galactose (comparison based on the prices of 100 g of the specified sugars).

6.1.2 Nuclear magnetic resonance for chemical characterization of the potential new paclitaxel precursor Taxa-4, 11-dien-5a-acetoxy-1α,10β-diol

Following the detection of the potential new compound T1β-diol it is important to verify the structure of the molecule by a different method other than GC-MS. A proposed method to validate the molecule is the use of NMR spectroscopy. This technique has proven to be successful for the characterisation of proteins as parameters like structure and dynamics can be evaluated in solution, solids and living cells (Klukowski et al., 2022). Furthermore, the use of this technique has been used before to characterise taxane diterpenoids from *Taxus* media cell suspension cultures (Fischedick et al., 2015). However, the technique is tedious and usually requires the aid of an expert (Klukowski et al., 2022).

6.1.3 Expansion into new chassis organisms

In nature, microbial communities are complex and usually can include a great number of organisms. For synthetic microbial communities to harness the capabilities of native microbial consortia, new organisms need to be tested and incorporated into new synthetic communities.

Apart from the research done in plants for the study of taxanes (especially of the genus *Taxus*) (Hao et al., 2011; Kuang et al., 2019). A thorough search in the literature could not find any research paper concerning the heterologous production of taxanes in

other organisms apart from *S. cerevisiae* and *E. coli,* however, reports show that researchers are adopting the use of non-conventional organisms for the study of microbial consortia. Among these non-conventional organisms, we can mention *Bacillus subtilis* (Det-udom et al., 2019), *Shewanella oneidensis* (Y. Liu et al., 2017), *Bacillus cereus, Brevundimonas naejangsanensis* (S. Wang et al., 2019), *Lactococcus lactis* (Kong et al., 2018), *C. glutamicum* (Sgobba et al., 2018) and *Pseudomonas putida* (Gao et al., 2021) among others. The use of novel microorganisms will allow it to harness its native capabilities and create more robust and versatile microbial consortia with new capabilities that can even be further enhanced by strain specialization. A good example of this is the use of *GGPPs* taken from *Sulfolobus acidocaldarius.* The use of this enzyme rather than the native *Taxus* enzyme enable a 40-fold increase in the taxadiene production (Engels et al., 2008).

6.1.4 New tools for real-time monitoring and balancing of consortia populations.

One of the difficulties experienced in this work was the real-time monitoring of the consortia populations. Although the offline methods were useful to provide a general picture of population dynamics, it would've been helpful to be able to monitor these changes online to be able to tune the consortia towards the desired outcome. It would also be important to distinguish between the different organisms of the consortia even if they belong to the same species (for example in an *S. cerevisiae- S. cerevisiae* consortia). Furthermore, the photographs taken and shown in appendix section 4 G belonging to different consortia did not provide enough information to estimate the final balance of the consortia. Also, it doesn't provide a real-time parameter of how the consortia population fluctuated over time.

Aligned with this, it is also important to have better tools for live fine-tuning of microbial populations. One of the most promising tools to solve this challenge is the use of optogenetics, which has proven to be an effective tool for controlling consortia populations (Lalwani et al., 2021). Different optogenetic systems have been successfully implemented in different bacteria. Some these bacteria are *E. coli* (Chang et al., 2017; Möglich et al., 2009; Ohlendorf et al., 2012; Ramakrishnan & Tabor, 2016; G. Wang et al., 2018), *P. aeruginosa* (Y. Huang et al., 2018; Pu et al., 2018; Xia et al., 2021), *V. natriegens* (Tschirhart et al., 2019), *S. meliloti* (Pirhanov et al., 2021), *B.*

subtilis (Castillo-Hair et al., 2019) and *P. putida* (Hueso-Gil et al., 2020). Despite this, to our knowledge, there's only been a few studies making use of this technique for population balance in a consortium. Furthermore, it is also of interest to highlight the study from (Lalwani et al., 2021) who combined flow cytometry for the quantification of the populations by expression of different fluorescent tags for each strain. The methods already developed for use in *S. cerevisiae* and *E. coli* could be adapted to engineer our consortia and optimise in real-time the populations for enhancement of taxane production.

Finally, the constant and ongoing expansion of the optogenetic toolkits into different microbial species will allow for better population control and provide another powerful tool for engineering synthetic consortia.

6.1.5 Incorporation of computational design

The convergence of biology and computational methods has been predicted to become wider for more than a decade (Navlakha & Bar-Joseph, 2011). Furthermore, the incorporation of artificial intelligence and computational methods will not only enable the collection, connection and analysis of data but also the construction of diverse predictive models across diverse disciplines (Hassoun et al., 2022; Wondraczek et al., 2019). This would, for example, provide methods for bioreactor control to drive the outcome of microbial communities' fermentations towards a particular goal (Treloar et al., 2020). One of the most promising tools towards the engineering of synthetic microbial consortia is the use of artificial intelligence (AI), computer simulations and models as part of the study of the microbial interactions (Fedorec et al., 2021; Kong et al., 2018; Treloar et al., 2020; Wondraczek et al., 2019; Xu, 2021). The use of modelling and AI could allow predicting how a particular consortium would behave over time based on initial conditions given. Furthermore, the control and feedback over these conditions (temperature, pH, feed ratio) could enable a new method for balancing populations towards a specific "balance goal" (Treloar et al., 2020). Finally, in a review paper published in 2021 (García-Jiménez et al., 2021) the authors summarized substantial examples of synthetic microbial consortia including the microorganisms, interaction, goal to optimise, carbon source, and yield but what's more interesting is that they listed a "descriptive microbial community

modelling methods classification" including some properties and tools used for this modelling.

6.1.6 Limitations

For the first preliminary experiment (Figure 3.5) there were no replicates. Due to this, it was not possible to perform any statistical analysis in this experiment. However, as the results expected were of qualitative nature (presence or absence of taxadiene) there was no further need for statistical analysis although these are still encouraged for future experiments.

One of the first limitations was that due to the complexity of the experiments with the bioreactors, it was only possible to do two fermentations with each carbon source (glucose and raffinose). This made it impossible to perform a statistical analysis of the results. Repeating this experiment may lead to different results until the "n" is high enough to finally conduct statistical tests to standardise the experiment. Furthermore, at some points, we had some issues with the Dissolve Oxygen probe which appeared to be damaged. However, this was solved by turning the air supply on to prevent oxygen depletion in the culture. At the same time, for an unknown reason other than a failure in the system, the temperature in the reactor was below the setpoint of 30° C during the first 24 hrs (Figure 3.8 B) which may have had a negative effect on yeast growth and subsequently also in production.

Despite the efforts in standardizing all experiments in this chapter, there still are some limitations that need to be mentioned as these can affect the reproducibility of the experiments or can be the source or errors.

In the experimental results corresponding to Figure 4.5, no viability staining was conducted which means that it is impossible to determine:

A) if decrease in the number of yeast cells after 24 hrs was due to the relative number of yeast cells diminishing in comparison to the increase in *E. coli* cells or

B) if the decrease in the number of yeast cells was a response to the toxicity in the media due to the high ethanol concentrations achieved.

A viability study performed by staining the cells with methylene blue (Matsumoto et al., 2022) could have been enough to determine the real reason behind the decrease in yeast population allowing to develop an strategy to counteract this.

The inoculation was done by using calibrated pipettes to assure that the volumes taken were accurate as any change in the volume may have influenced the final OD in each well and consequently affect the population balance. However, manual methods for liquid handling always carry an error which in this case may potentially affect the reproducibility of the experiments.

The biggest limitation in these experiments was the lack of a method to quantify with accuracy the final populations of each microbial specie as these methods combined with a statistical design of the experiment could have potentially improved the results for many of the consortia tested.

One of the biggest limitations in this chapter is that the experiment was designed based on the results obtained in chapter 4, however, it would've been beneficial to perform a statistical design of experiments to optimise the inoculation ratio of each specie and maximise production of T1B-ol.

Another limitation is that despite the characterization of the molecule using the GC-MS equipment further validation is needed by Nuclear Magnetic Resonance (NMR) to confirm that the hydroxylation was made in the desired carbon (1β) and not in any other carbon within the molecule.

Finally, the lack of a proper standard for quantification might cause some errors when calculating the total amount of compound produced.

6.2 Limitations

The first limitation presented is the number of replicates in the bioreactor runs. Bioprocessing is labour intensive and complex so its implementation can be a constraint when time limitations -like in our case- are present. Furthermore, the equipment used for online measurements such as pH and DO probes tend to be very fragile and sensitive and any failure of this can have an impact (sometimes significant) on the results. More time to work with the bioreactors would have had a positive result on the research as more repetitions would give the opportunity to perform statistical analysis to validate the results obtained. Furthermore, it would have been interesting to try on a bigger scale other carbon sources (i.e., fructose) and even some of the microbial consortia that showed the best results.

Determining an accurate distribution of the microbial populations and performing an implementing a viability method in the consortia was also an important limitation as knowing the exact viable composition of each microbial specie would have allowed to be engineered more efficiently the inoculation ratios towards the optimization in production of a particular compound of interest. A viability essay would have also allowed to determine if a particular increase or decrease of a microbial specie was the result of the accumulation of toxic compounds or a result of one species being outgrown by the other.

The inoculation was done by using calibrated pipettes to assure that the volumes taken were accurate as any change in the volume may have influenced the final OD in each well and consequently affect the population balance. However, manual methods for liquid handling always carry an error which in this case may potentially affect the reproducibility of the experiments.

The biggest limitation in these experiments was the lack of a method to quantify with accuracy the final populations of each microbial specie as these methods combined with a statistical design of the experiment could have potentially improved the results for many of the consortia tested.

Finally, the detection, characterization, and quantification of the novel T1B-ol is one of the major breakthroughs of these research. However, the section describing and discussing these results could have been benefitted with a Nuclear Magnetic Resonance (NMR) analysis to further corroborate the hydroxylation of the Carbon 1β of the paclitaxel skeleton.

6.3 Conclusions

In chapter 3 we aimed to genomically engineer diverse strains to enhance the production of T5αAc and T10β-ol. By deleting just 1 gene (GAL80) we managed to "relax" the carbon catabolite repression of GAL promoter under the presence of carbon sources other than galactose thus expanding the different sugars that can be used for bioproduction of paclitaxel intermediates by our *S. cerevisiae* strains. The production in strain EJ2 achieved levels of up to 76.12 ±15.76 of taxadiene, 18 ±2.19 of T5α-ol and 19.4 ±2.61 of T5αAc which adds up to a total taxane production of 113 mg/L which represents an increase of 18.44% (Figure 3.9 A) compared to 95.4 mg/L achieved in previous reports (Walls et al., 2021). Furthermore, the production of the acetylated product T5αAc in galactose by EJ1 and EJ2 achieved an increase of 2-fold and 4.5- respectively when compared to strain S14 without GAL80 deletion.

The experiments proved that growth and taxane production in raffinose fermentations was mainly due to the fructose monomer and that melibiose was not fermented by our strains. The results obtained from fructose fermentation by the strain EJ2 seemed to prove this theory as the total production of T5αAc was 66 mg/L compared to 73 mg/L achieved by the same strain in raffinose. This is important because the price of fructose is almost 13 times lower than the price of raffinose and 10.5 lower than the price of galactose (comparison based on the prices of 100 g of the specified sugars according to Merck) which provides a low-cost carbon source to produce taxanes.

Based on the results described previously a microbial consortia strategy was implemented in Chapter 4 to produce taxanes. The consortia were mostly based on a modular approach in which different organisms expressed different enzymes of the paclitaxel metabolic pathway. In this manner, the metabolic burden on each organism was reduced which allowed increasing the production of some paclitaxel precursors. Coupled with this strategy we also implemented an inoculum engineering approach in

which different inoculation ratios were tested to optimise the division of labour and increase the production of compounds. The modular strategy combined with the inoculum engineering approach proved the functionality of the *E. coli- S. cerevisiae* consortium and enabled –for the first time in our experiments— the detection of T10βOH with a production of ~9 ±0,3 mg/L which represents an increase of 10-fold compared to previously reported data (Kang Zhou et al., 2015).

It is also very relevant to mention that as part of our experiments, we also tested one of the most ambitious consortia that we are aware of. This involved the use of 2 different yeast strains (EJ2-TAT and EJ2-T10βOH) plus 3 different bacteria strains and it is to our knowledge one of the most successful and complex consortia including five microorganisms dividing a synthetic pathway and achieving true synergy to divide and relieve the metabolic burden among the consortia. Furthermore, different co-cultures of S. cerevisiae EJ2 with different individual plasmids (with TAT and T10βOH genes) and inoculum engineering also proved to be an efficient strategy for relieving the metabolic burden and for controlling the production of hydroxylated taxanes. Examples of this are the co-cultures of EJ2 with TAT/T10βOH plasmids to produce T5αAc and T10β-ol. In these experiments, the multi-copy nature of the plasmids coupled with inoculum engineering allowed for the highest production ever achieved of these two precursors with productions values of 110 ±3 mg/L for T5αAc and 26 ±5 mg/L fot T10βOH.

Finally, in Chapter 5 the aim was to use next-generation sequencing (NGS) tools to characterise the genome of our parental strain EJ2 and validate the insertions and deletions mentioned in the previous chapters. At the same time, a different objective was to test the same modular strategy and inoculum engineering approach described in the previous chapter to produce a new taxane compound and validate the activity of the putative gene TB574 codifying for T1βOH enzyme. The result of the NGS experiments probed the correct insertion of all the paclitaxel genes except for T10βOH. The GAL80 deletion was also successfully confirmed. The tools used for analysis probed to be very versatile for genomic characterisation of the strain. Furthermore, the results obtained from the microbial consortia allowed for the quantification and characterisation of a long-missing intermediate in the Taxol pathway (T1β-ol). Adjusting inoculation ratios had a beneficial effect on the production of the

dioxygenated product T10β-ol where production of up to 35.5 mg/L was achieved. Even more important is the production of 45 mg/L of the newly discovered compound T1β-diol. This result is very exciting as it is the first time in 20 years that a new enzyme in the metabolic pathway is detected and functionally expressed for the conversion of its precursor. Furthermore, this discovery can enable the expansion of the discovery of new enzymes and compounds in the paclitaxel metabolic pathway.

The methodology developed in this work could be applied to the discovery and expansion of new pathways to produce complex compounds. It is very important to remark that the methodology developed is simple and versatile and can be easily implemented in different microorganisms to evaluate production in diverse microbial consortia. As new computational and molecular tools arise these can be coupled to enhance and optimised the techniques here presented.

APPENDICES

Appendices for Chapter 3

Appendix 3 A List of primers used for GAL80 deletion and T10βOH insertion

Table A1: List of primers used for amplification of the DNA parts used for GAL80 deletion.

Name	Sequence 5' → 3'
NUP80-F	GCCTGTCTACAGGATAAAGACG
NUP80-R	ACTGGGGGCCAAGCACAGGGCAAGATGCTT
NUPHA80-R	ACTGGGGGCCAAGCACAGGGCAAGATGCTTGACGGGAGT GGAAAGAACGG
NDOHA-F	CAATCTCGATAGTTGGTTTCCCGTTCTTTCCACTCCCGTCA AGCATCTTG
NDO80-F	AAGCATCTTGCCCTGTG
NDO80-R	AAATATGACCCCCAATATGAGAAA
Sg RNA for GAL80 deletion *	gcaacaccttcgggtggcgaatgggactttGATGAGCGTGGTAACCGATT gttttagagctagaaatagcaagttaaaat
PosCol-PCR-F	TTTGTAGGTGTGTCGTTTATCC
PosCol-PCR-R	CCGCTTAAATAATCCTGGAAGG
NegCol-PCR-F	TGGGGTTCAAACCATCATCTCT
NegCol-PCR-R	CGGCATCGCCTTCAAGTTTC

*Lowercase letters indicate the bases overlapping the CAS9 Cassette.

Table A2: List of primers used for amplification of the DNA parts used for T10βOH insertion.

Name	Sequence 5' → 3'
N-UpHa10beta-F	GATTGAGGAACTGTTGGTAT
N-UpHa10beta-R	TCAATTCTAAGCACTCAAACCG
N-UpHa10beta-HA-	CGCTCGGCGGCTTCTAATCCGTACTTCAATTCTAAGC ACTCAAACCG
N-10b-F	AGTACGGATTAGAAGCCGCCGAGC
N-10b-R	CTTCGAGCGTCCCAAAACCTTCT
N-10b-HA-F	GCACGGTTTGAGTGCTTAGAATTGAAGTACGGATTAGAAG CCGCCGAGC
N-10b-HA-R	CAATGGGGTAGATCATACCCTAAAATGTTTCCTTCGAGCGT CCCAAAACCTTCT
Down10b-F	GAAACATTTTAGGGTATGATCT
DownHa10beta-F	CCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGAAACATT TTAGGGTATGATCT
Down10b-R	AATTTCATTTTATCTATGAGAAA
Col-PCR-10b-F	TACCACGCCATTTTAGTAGGC
Col-PCR-10b-R	TAGCTGGGGAAGATTCCAATT
Sg RNA for T10B integration *	gcaacaccttcgggtggcgaatgggactttGGCCTCGGCAACTTTCAGG Ggttttagagctagaaatagcaagttaaaat

*Lowercase letters indicate the bases overlapping the CAS9 Cassette.

Appendices for Chapter 4

4A List of codon-optimised gene sequences for expression of different P450s and reductases in *E. coli*

CPR	GGATCCAAGGAGATATACCGAAATAATTTTGTTTAACTTTAAGAAG GAGATATACATATGGCCCTGCTTCTGGCGGTCTTCAGACGTGGCGG AAGTGACACGCAGAAACCAGCAGTTCGCCCGACCCCACTGGTCAAA GAGGAAGATGAAGAGGAAGAGGACGACAGTGCAAAGAAGAAAGTT ACCATCTTCTTCGGCACACAGACGGGCACAGCAGAGGGCTTTGCGA AGGCGCTGGCGGAAGAGGCGAAAGCGCGTTATGAGAAAGCGGTAT TCAAAGTAGTGGATTTAGACAATTATGCAGCAGACGACGAGCAGTA TGAGGAGAAACTCAAGAAAGAGAAATTAGCCTTCTTCATGCTGGCA ACATACGGAGATGGCGAGCCCACAGACAATGCGGCAAGATTTTATA AATGGTTTCTGGAAGGCAAGGAGCGGGAGCCTTGGCTGAGTGATCT GACATATGGAGTCTTTGGGCTTGGAAATCGACAGTATGAGCATTTC AATAAAGTGGCAAAAGCCGTGGATGAGGTGTTGATTGAGCAGGGC GCAAAACGTCTGGTACCCGTAGGGCTGGGAGATGATGACCAGTGTA TCGAAGACGACTTTACGGCGTGGCGGGAGCAGGTGTGGCCGGAGTT AGACCAGCTCCTGCGGGACGAGGATGATGAGCCTACATCGGCAACG CCGTATACTGCCGCCATTCCCGAGTATCGTGTAGAGATATATGATAG TGTGGTCTCAGTATATGAGGAGACCCATGCACTTAAACAGAATGGG CAGGCAGTTTATGATATACATCACCCTTGCAGAAGCAACGTGGCAG TGCGCCGCGAGTTACACACCCCGCTGAGCGACCGTAGCTGCATTCA TCTGGAGTTTGATATTAGTGACACAGGACTTATTTATGAGACAGGC GATCATGTGGGAGTTCATACCGAGAATAGCATAGAGACGGTCGAGG AAGCAGCAAAACTGTTAGGCTATCAGTTAGACACAATATTTAGCGT GCATGGCGATAAAGAGGATGGAACGCCTCTGGGCGGAAGTTCATTA CCGCCGCCGTTTCCTGGGCCGTGCACACTGCGTACCGCCTTAGCACG TTATGCAGATCTGCTGAATCCCCCGCGCAAAGCGGCGTTCTTAGCCC TTGCAGCGCACGCGAGTGACCCAGCCGAAGCGGAGCGGCTGAAGTT CTTGTCCTCACCTGCGGGAAAAGATGAGTATTCACAGTGGGTCACA GCTTCACAGAGATCACTGCTCGAGATCATGGCAGAGTTTCCGAGTG CCAAACCCCCGCTCGGGGTGTTCTTTGCCGCCATCGCACCGCGTCTG CAGCCTAGATATTATTCTATCTCGTCAAGTCCTCGGTTTGCGCCCAG CCGTATACATGTGACATGTGCCTTAGTATATGGCCCGAGCCCGACA GGGAGAATTCATAAGGGAGTCTGTAGTAATTGGATGAAGAATTCGC TCCCTTCAGAGGAGACGCATGATTGCAGCTGGGCACCCGTATTTGTT CGTCAGAGTAACTTTAAGCTCCCCGCAGACAGCACCACACCGATAG TGATGGTTGGGCCCGGGACGGGGTTTGCCCCTTTCCGCGGCTTTCTG CAAGAGCGTGCAAAACTGCAGGAAGCCGGGGAGAAGTTAGGGCCG GCGGTGCTGTTCTTTGGATGCCGCAATCGCCAGATGGATTATATTTA TGAGGACGAGTTGAAAGGCTACGTTGAGAAAGGCATACTGACAAAT CTTATTGTTGCGTTTAGTCGCGAAGGCGCAACTAAAGAGTATGTCCA GCATAAGATGCTGGAGAAGGCGTCCGACACCTGGAGTCTGATAGCA CAGGGCGGATATCTTTATGTATGCGGCGACGCTAAGGGCATGGCAC GTGACGTACATCGTACGTTACACACGATCGTGCAGGAGCAGGAGTC CGTTGATTCAAGTAAAGCAGAATTCCTGGTAAAGAAACTGCAGATG GATGGGCGCTATCTGCGCGATATATGGTGA

T5αOH	ATGGCCTTATTATTAGCAGTATTCTTCTCTATAGCATTGTCAGCCATT GCAGGGATTTTATTACTTTTGTTACTGTTTCGTTCCAAACGCCACTCT TCATTAAAGCTTCCGCCGGGCAAACTGGGAATTCCGTTTATCGGAG AATCTTTTATTTTCCTGCGCGCATTACGCAGTAATTCACTGGAGCAG TTCTTTGATGAGCGTGTAAAGAAATTCGGACTGGTTTTCAAAACAA GCTTAATCGGACATCCTACCGTCGTACTCTGTGGACCGGCAGGGAA TCGACTGATACTTAGTAACGAGGAGAAACTGGTCCAGATGTCCTGG CCTGCGCAGTTTATGAAACTTATGGGCGAGAATTCAGTAGCGACTC GTCGTGGGGAAGACCATATAGTAATGCGTTCAGCATTAGCGGGGTT CTTTGGCCCGGGCGCCCTGCAGTCTTATATAGGCAAAATGAATACG GAGATACAGAGCCATATCAATGAGAAGTGGAAAGGCAAAGACGAG GTGAATGTTCTGCCCCTGGTTCGCGAGCTGGTATTTAACATAAGCGC CATACTTTTCTTTAATATTTATGATAAACAGGAACAGGACCGGCTCC ACAAACTCTTAGAGACCATCCTGGTAGGGTCGTTTGCGCTTCCGATA GACTTGCCGGGGTTCGGGTTCCATCGCGCGCTCCAGGGGCGTGCGA AATTAAATAAAATCATGCTGTCGCTTATTAAGAAAAGAAAAGAGGA CTTACAGAGCGGGTCTGCCACAGCGACACAGGATTTACTGAGCGTT CTGTTAACTTTTCGAGATGATAAAGGCACGCCCCTGACGAATGATG AGATACTCGATAACTTTAGCTCACTGCTCCATGCATCGTATGATACG ACTACAAGCCCTATGGCCCTGATATTCAAACTGCTCAGTAGTAATCC GGAGTGCTACCAGAAAGTGGTACAGGAGCAGCTTGAGATCCTGTCA AACAAAGAGGAAGGGGAAGAGATAACCTGGAAGGACCTCAAAGCC ATGAAGTACACATGGCAGGTGGCCCAGGAGACATTGAGAATGTTTC CCCCGGTCTTCGGCACTTTTCGTAAAGCCATCACTGATATCCAGTAT GATGGGTATACCATACCGAAAGGATGGAAATTATTATGGACAACCT ATTCCACACATCCTAAAGACCTTTATTTCAATGAGCCGGAGAAGTTC ATGCCGAGCCGCTTCGATCAAGAGGGAAAACATGTAGCGCCGTACA CGTTTCTTCCGTTTGGCGGCGGGCAGCGTTCATGTGTGGGCTGGGAG TTTTCCAAGATGGAGATCCTGCTCTTTGTTCATCATTTTGTAAAGAC GTTTTCAAGTTACACGCCGGTCGATCCGGATGAGAAAATTAGCGGC GACCCGTTACCGCCGCTGCCTTCCAAAGGCTTTAGTATCAAATTATT TCCTCGTCCTTGA
T10βOH	ATGGCCCTGCTTCTCGCAGTGTTCAGCCTTACACTCGCCCCGATATT AGCTATAATTCTCCTGCTTTTGTTTCGCTATAATCATCGCAGTAGTGT CAAACTCCCGCCTGGCAAGCTGGGGTTCCCTCTCATAGGGGAGACG ATACAGTTGCTCCGGACTCTGCGTAGCGAGACTCCGCAGAAATTCTT TGACGATCGGCTGAAGAAATTCGGCCCGGTGTATATGACATCGCTG ATTGGCCACCCCACGGTGGTGCTGTGCGGACCCGCCGGAAATAAGC TTGTGCTGTCGAATGAGGACAAACTGGTGGAGATGGAAGGACCTAA ATCGTTTATGAAACTCATAGGCGAGGACTCCATTGTAGCAAAACGG GGTGAGGACCATCGAATACTTCGTACCGCCCTGGCCCGATTCCTTGG GGCACAGGCCCTGCAGAATTATTTAGGGCGTATGAGCTCAGAGATA GGCCACCACTTCAATGAGAAATGGAAAGGGAAAGACGAGGTGAAA GTACTTCCGCTGGTACGGGGATTAATCTTTTCAATCGCGTCAACACT CTTCTTTGATGTGAACGACGGACACCAGCAGAAACAGCTCCATCAT CTGCTGGAGACAATCTTAGTCGGCTCACTGAGTGTACCGCTGGACTT CCCTGGAACCCGTTATCGCAAAGGCCTGCAGGCGCGACTGAAACTT GATGAGATTTTATCTAGTTTAATCAAACGCCGTCGTCGAGACTTACG TTCAGGAATCGCCTCAGACGACCAGGACCTTCTTTCGGTTCTCCTGA CATTTCGCGACGAGAAAGGGAATTCCCTGACTGACCAGGGCATTCT

	GGACAATTTTAGCGCCATGTTTCACGCAAGTTATGACACTACAGTCG CCCCTATGGCGCTGATATTTAAGCTGCTTTATAGTAATCCTGAGTAT CATGAGAAAGTATTTCAGGAGCAGCTGGAGATTATTGGAAATAAGA AAGAGGGTGAGGAGATAAGCTGGAAAGATTTAAAGTCCATGAAAT ATACCTGGCAGGCGGTTCAGGAGTCCCTGCGCATGTACCCGCCCGT ATTTGGCATCTTTCGGAAAGCGATCACAGACATTCACTATGACGGG TATACCATACCGAAAGGATGGCGAGTGCTGTGCTCACCTTACACGA CACATTTACGCGAGGAGTATTTCCCGGAGCCTGAGGAGTTTCGCCC GAGTCGGTTCGAGGACGAGGGCCGTCACGTCACCCCGTATACGTAT GTGCCTTTCGGTGGTGGATTACGCACATGCCCGGGCTGGGAGTTCA GTAAAATTGAGATACTTCTGTTTGTACATCATTTTGTCAAGAACTTT TCCAGCTATATACCTGTTGATCCCAATGAGAAGGTACTGTCAGACCC CTTACCGCCGCTGCCGGCGAACGGCTTTAGCATCAAATTATTTCCCC GTAGCTGA

4B List of primers and plasmid maps for T5αOH plasmid assembly in *E. coli* LYglc1 and *E. coli* LYxyl3

Table 4B: List of primers used for in vitro assembly of T5αOH *E. coli* plasmid gene.

Name	Sequence 5' → 3'	Description
BB-F	CCGGCTGCTAACAAAGCCC	Amplification of plasmid backbone
BB-R	CCCCTCTAGGATCCGAATTCG	
BB-F-OL	TCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCC	Addition of overlapping regions for in vitro assembly
BB-R-OL	TATCTCCTTCTTAAAGTTAAACAAAATTATTTCGAATTCGGATCCTAGAGGGG	
T5-F	ATGGCCTTATTATTAGCAGTATTCT	Amplification of T5αOH gene
T5-R	TCAAGGACGAGGAAATAATTTGATAC	
T5-F-OL	TTAACTTTAAGAAGGAGATATACATATGATGGCCTTATTATTAGCAGTATTC	Addition of overlapping regions for in vitro assembly
T5-R-OL	ATCCTCAAGGACGAGGAAATAATTTGATACTAAAGCCTTTGGAAGGC	
CPR1-F	GGATCCAAGGAGATATACCG	Amplification of Reductase Fragment 1
CPR1-R	CGTCTCTATGCTATTCTCGG	
CPR1-F-OL	TATCAAATTATTTCCTCGTCCTTGAGGATCCAAGGAGATATACCGAAATAATT	Addition of overlapping regions for in vitro assembly
CPR1-R-OL	CTGCTTCCTCGACCGTCTCTATGCTATTCTCGGTATGAACTCCCACATG	
CPR2-F	GTCGAGGAAGCAGCAAAAC	Amplification of Reductase Fragment 2
CPR2-R	TCACCATATATCGCGCAG	
CPR2-F-OL	AGAATAGCATAGAGACGGTCGAGGAAGCAGCAAAACTGTTAGGCTATCAGTTA	Addition of overlapping regions for in vitro assembly
CPR2-R-OL	CAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCTCACCATATATCGCGCAG	

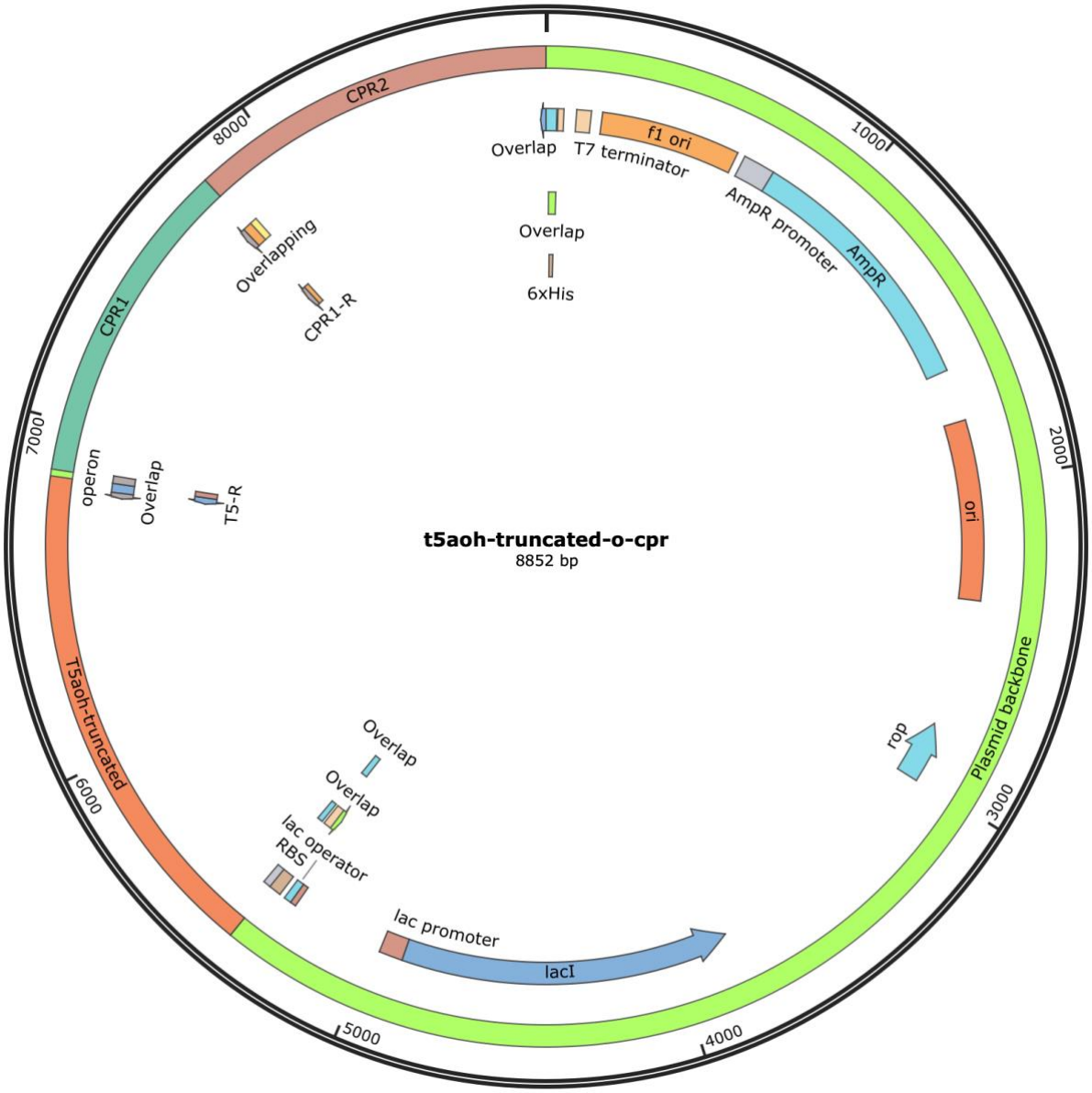


Figure 4B: Plasmid Map for *E. coli* expression of T5αOH-CPR enzyme. The system was designed using Benchling. The map was illustrated using SnapGene.

4C List of primers and plasmid maps for T10βOH plasmid assembly in *E. coli* LYglc1 and *E. coli* LYxyl3

Table 4C: List of primers used for in vitro assembly of T10βOH *E. coli* plasmid gene

Name	Sequence 5' → 3'	Description
BB-F	CCGGCTGCTAACAAAGCCC	Amplification of plasmid backbone
BB-R	CCCCTCTAGGATCCGAATTCG	
BB-F-OL	TCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCC	Addition of overlapping regions for in vitro assembly
BB-R-OL	TATCTCCTTCTTAAAGTTAAACAAAATTATTTCGAATTCGGATCCTAGAGGGG	
T10-F	ATGGCCCTGCTTCTCGC	Amplification of T10β gene
T10-R	TCAGCTACGGGGAAATAATTTG	
T10-F-OL	TTAACTTTAAGAAGGAGATATACATATGATGGCCCTGCTTCTCGCAG	Addition of overlapping regions for in vitro assembly
T10-R-OL	ATCCTCAGCTACGGGGAAATAATTTGATGCTAAAGCCGTTCGC	
CPR1-F	GGATCCAAGGAGATATACCG	Amplification of Reductase Fragment 1
CPR1-R	CGTCTCTATGCTATTCTCGG	
CPR1-F-OL	AAATTATTTCCCCGTAGCTGAGGATCCAAGGAGATATACCGAAATAATTTTG	Addition of overlapping regions for in vitro assembly
CPR1-R-OL	CTGCTTCCTCGACCGTCTCTATGCTATTCTCGGTATGAACTCCCACATG	
CPR2-F	GTCGAGGAAGCAGCAAAAC	Amplification of Reductase Fragment 2
CPR2-R	TCACCATATATCGCGCAG	
CPR2-F-OL	AGAATAGCATAGAGACGGTCGAGGAAGCAGCAAAACTGTTAGGCTATCAGTTA	Addition of overlapping regions for in vitro assembly
CPR2-R-OL	CAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCTCACCATATATCGCGCAG	

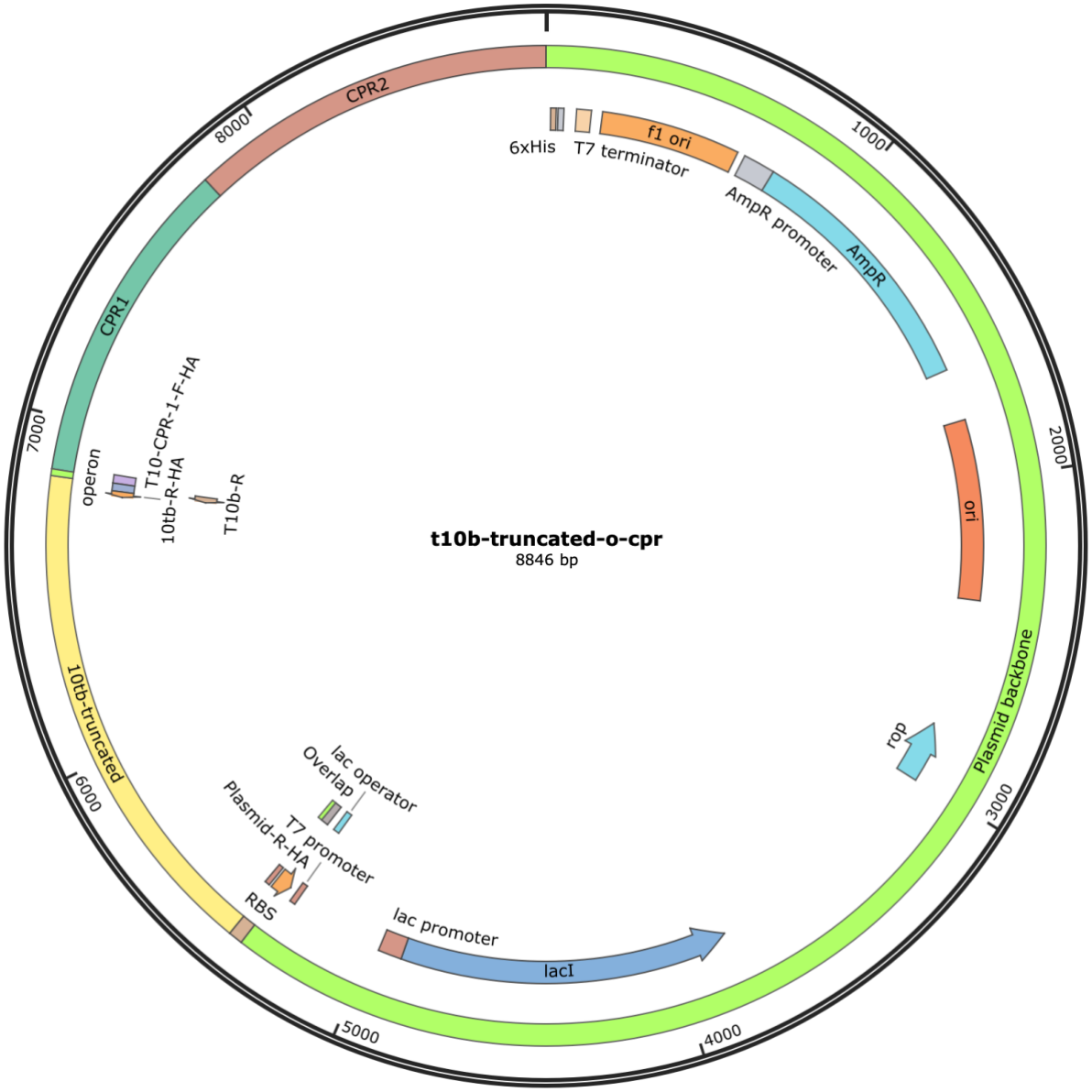


Figure 4C: Plasmid Map for *E. coli* expression of T10βOH-CPR enzyme. The system was designed using Benchling. The map was illustrated using SnapGene.

4D Plasmids map for TAT plasmid assembly in *E. coli* LYglc1 and *E. coli* LYxyl3

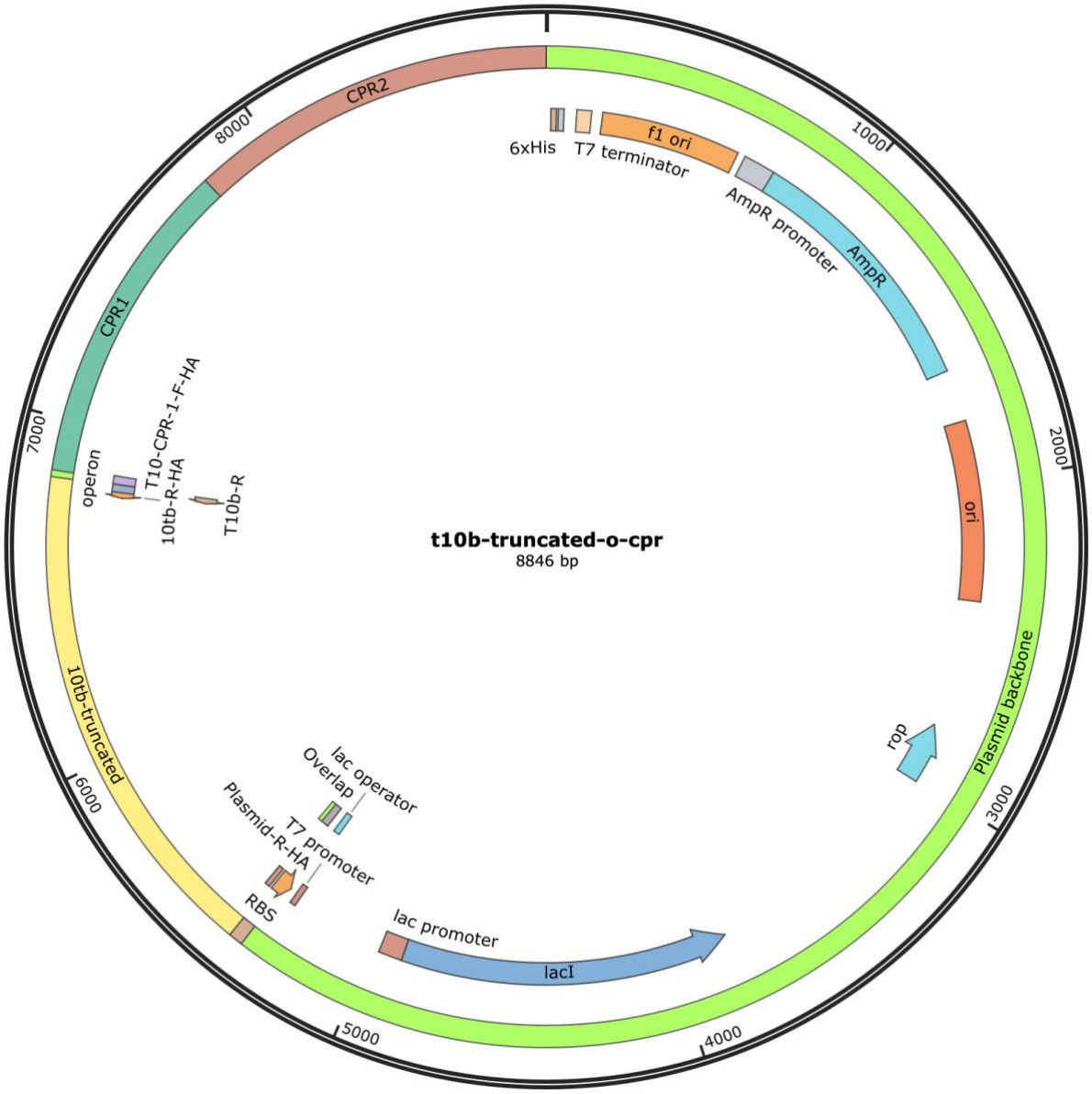


Figure 4D: Plasmid Map for *E. coli* expression of TAT enzyme. The system was designed using Benchling. The map was illustrated using SnapGene.

4E List of primers and plasmid maps for TAT plasmid assembly in *S. cerevisiae EJ2*

Table 4E: List of primers used for TAT plasmid in vivo assembly in *S. cerevisiae EJ2*

Name	Sequence 5' → 3'	Description
BB1-F	CATAGCTGTTTCCTGTGTG	Amplification of plasmid backbone 1
BB1-R	GCTCAACGTGATAAGGAAAAAG	
BB1-F-HA	GGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTG	Addition of overlapping regions to backbone 1 for in vivo assembly
BB1-R-HA	AGCATATTTGAGAAGATGCGGCCAGCAAAACTAAGCTCAACGTGATAAGGAAAAAG	
BB2-F	ACTGCGGTCAAGATATTTCTTG	Amplification of plasmid backbone 2 (Including TAT gene, promoter, and terminator)
BB2-R	ACCATGATTACGCCAAGCG	
BB2-F-HA	AATGATGAATTGAAAAGGTGGTATGGTGCACAACTGCGGTCAAGATATTTCTTG	Addition of overlapping regions to backbone 2 for in vivo assembly
BB2-R-HA	AGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCG	
URA3-F	TTAGTTTTGCTGGCCGCATC	Amplification of URA3 (including promoter)
URA3-R	GTGCACCATACCACCTTTTC	
URA3-F-HA	GTGCAATTCTTTTTCCTTATCACGTTGAGCTTAGTTTTGCTGGCCGCATC	Addition of overlapping regions to URA3 for in vivo assembly
URA3-R-HA	AGGCGCCTGATTCAAGAAATATCTTGACCGCAGTTGTGCACCATACCACCTTTTC	

Figure 4E: Plasmid Map for *S. cerevisiae EJ2* expression of TAT enzyme. The system was designed using Benchling. The map was illustrated using SnapGene.

4F List of primers and plasmid maps for T10βOH plasmid assembly in *S. cerevisiae EJ2*

Table AF: List of primers used for T10βOH plasmid in vivo assembly in *S. cerevisiae EJ2*

Name	Sequence 5' → 3'	Description
BB1-F	CATAGCTGTTTCCTGTGTG	Amplification of plasmid backbone 1
BB1-R	GCTCAACGTGATAAGGAAAAAG	
BB1-F-HA	GGTTAATTGCGCGCTTGGCGTAATCATG GTCATAGCTGTTTCCTGTGTG	Addition of overlapping regions to backbone 1 for in vivo assembly
BB1-R-HA	AGCATATTTGAGAAGATGCGGCCAGCAA AACTAAGCTCAACGTGATAAGGAAAAAG	
BB2-F	ACTGCGGTCAAGATATTTCTTG	Amplification of plasmid backbone 2 (Including TAT gene, promoter, and terminator)
BB2-R	ACCATGATTACGCCAAGCG	
BB2-F-HA	TGAAAGGGGCAGACATTAGAATGGTATA TCACTGCGGTCAAGATATTTCTTG	Addition of overlapping regions to backbone 2 for in vivo assembly
BB2-R-HA	AGCGGATAACAATTTCACACAGGAAACA GCTATGACCATGATTACGCCAAGCG	
URA3-F	TTAGTTTTGCTGGCCGCATC	Amplification of URA3 (including promoter)
URA3-R	CATTCTAATGTCTGCCCCTTTC	
URA3-F-HA	GTGCAATTCTTTTTCCTTATCACGTTGAG CTTAGTTTTGCTGGCCGCATC	Addition of overlapping regions to URA3 for in vivo assembly
URA3-R-HA	AGGCGCCTGATTCAAGAAATATCTTGAC CGCAGTT CATTCTAATGTCTGCCCCTTTC	

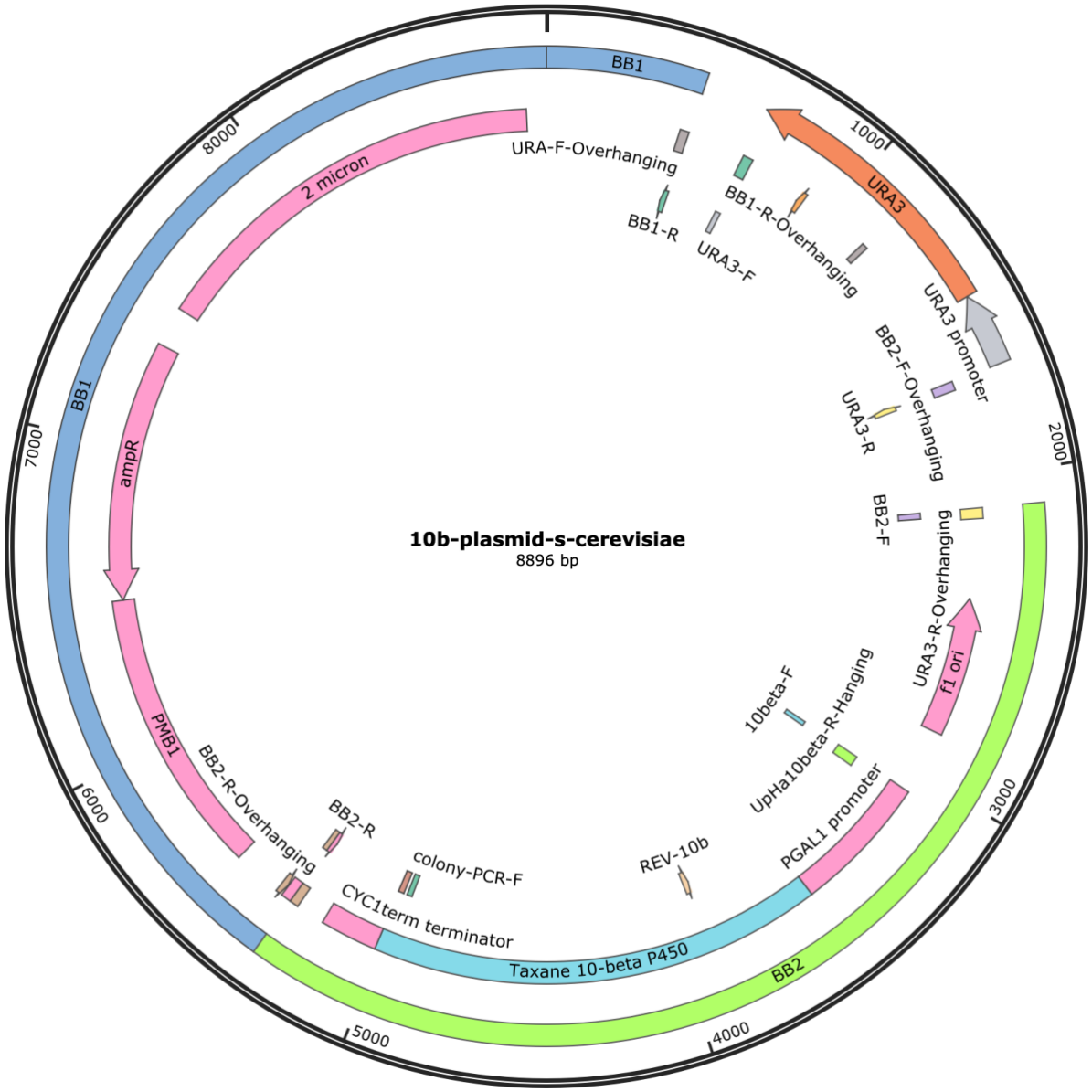


Figure 4F: Plasmid Map for *S. cerevisiae EJ2* expression of TAT enzyme. The system was designed using Benchling. The map was illustrated using SnapGene.

4G Microscope pictures showing the proportion of *S. cerevisiae* vs *E. coli* in different carbon sources

The pictures provide a visualisation of the different consortia in different carbon sources. Further analysis is needed to quantify each of the microorganisms.

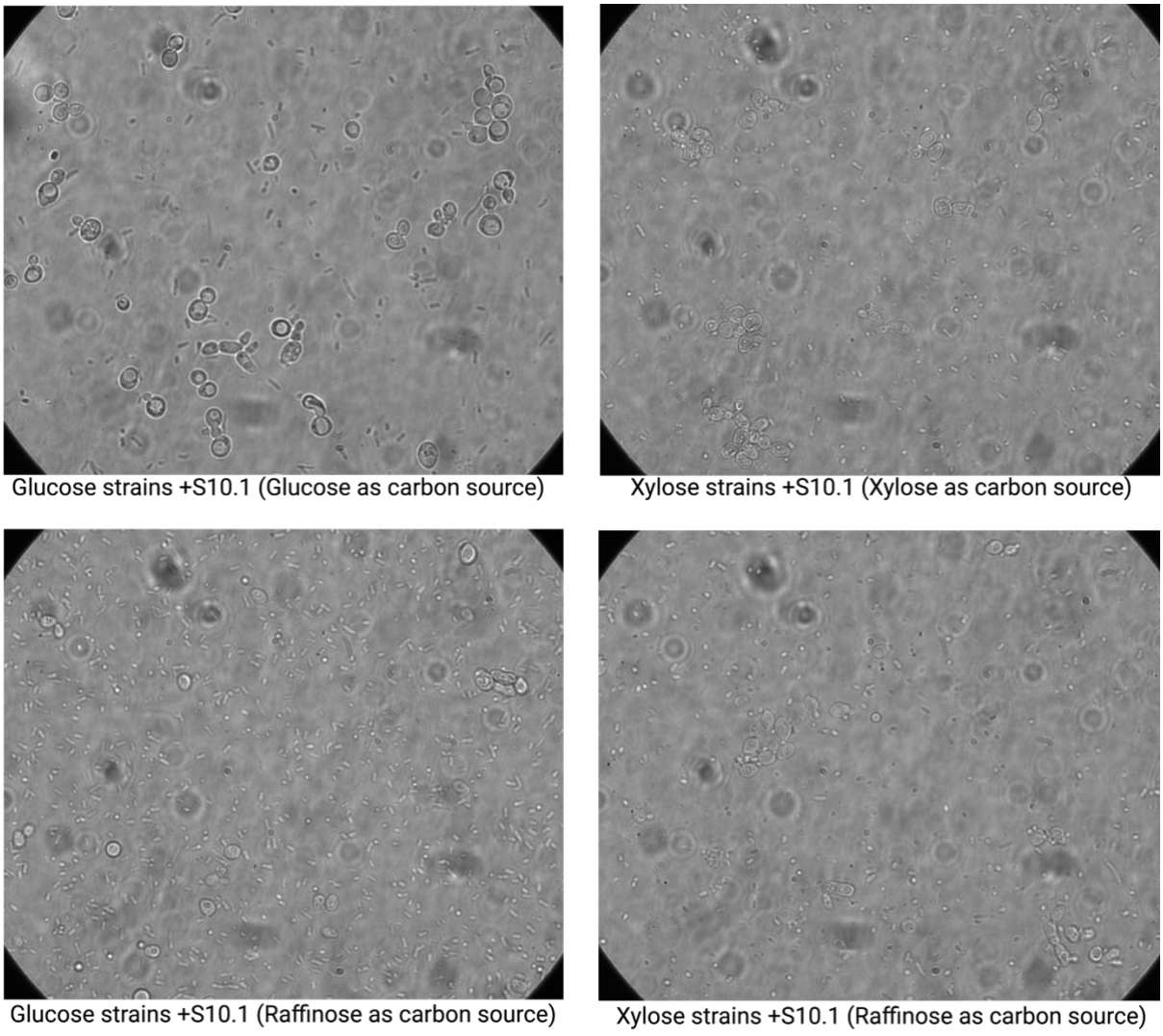


Figure 6G.1 Microscope images showing the proportion of *S. cerevisiae* and *E. coli LY strains* in different carbon sources and consortia after 72 hrs. Images correspond to the 40:1 inoculation ratio.

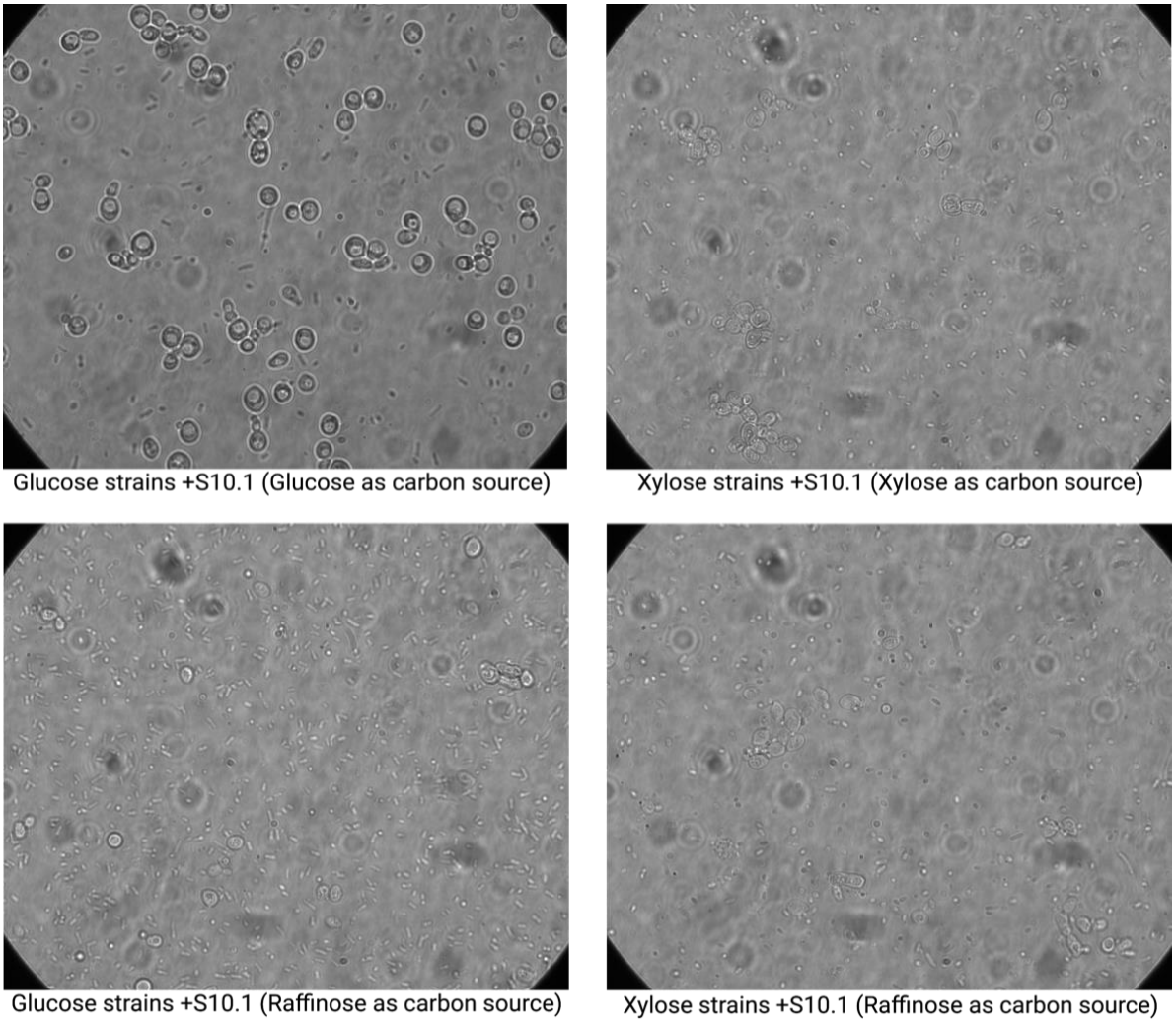
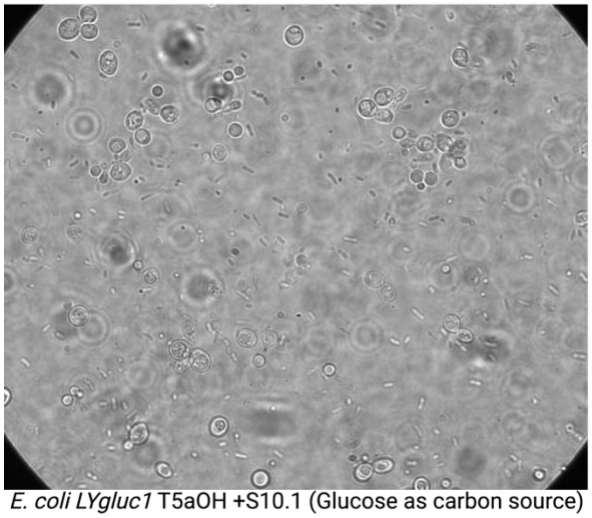
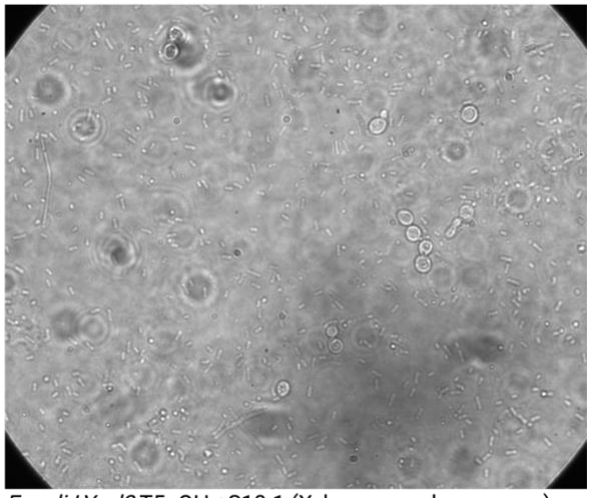


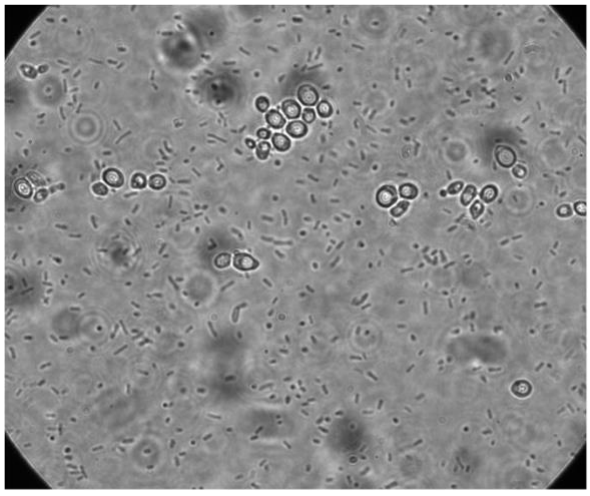
Figure 7G.2 Microscope images showing the proportion of *S. cerevisiae* and *E. coli LY strains* in different carbon sources and consortia after 72 hrs. Images correspond to the ~6.5:1 inoculation ratio.



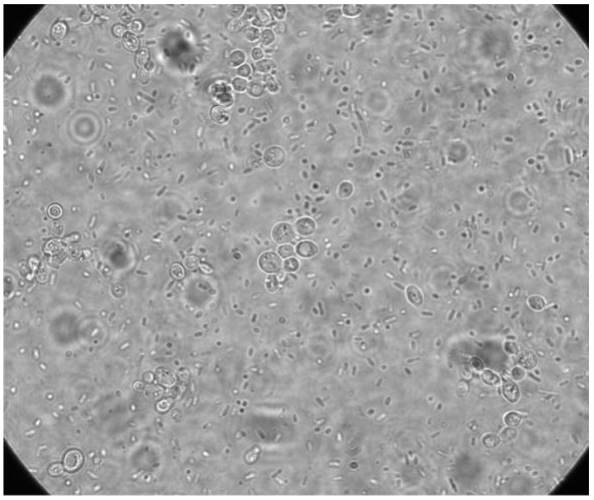
E. coli LYgluc1 T5aOH +S10.1 (Glucose as carbon source)



E. coli LYxyl3 T5aOH +S10.1 (Xylose as carbon source)



E. coli LYgluc1 T5aOH +S10.1 (Raffinose as carbon source)



E. coli LYxyl3 T5aOH +S10.1 (Raffinose as carbon source)

Figure 8G.3 Microscope images showing the proportion of *S. cerevisiae* and *E. coli LY strains* in different carbon sources and consortia after 72 hrs. Images correspond to the 20:1 inoculation ratio.

Appendices for Chapter 5

5A Comparison of the total outputs of 2 different sequencing runs

The conditions and settings for both runs were the same. The reagents from the second run (including flongle flow cell) were fresher than those for the first experiment which explains the better results.

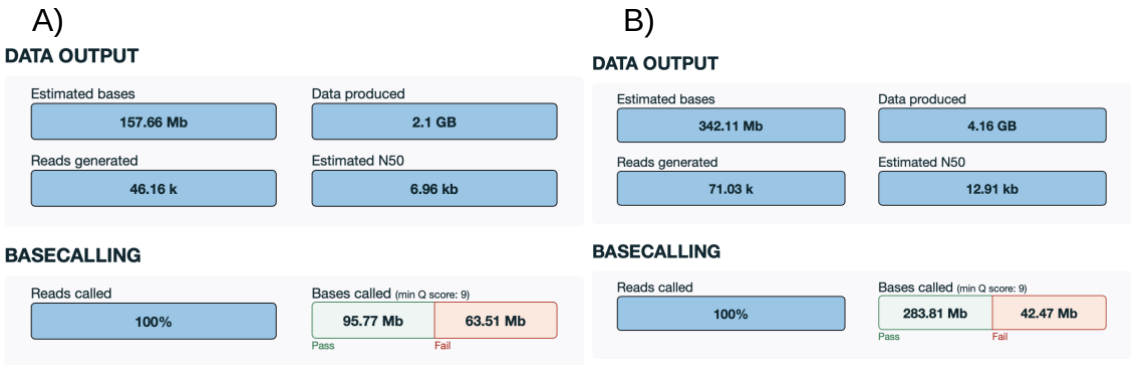


Figure 5 A.1 A and B show a summary of the data gathered by NGS. In both cases, the EJ2 strain was sequenced using Rapid Sequencing Kit using Flonge technology by Oxford Nanopore Technologies. Image B shows the replicate of the experiment showing a massive improvement in the amount of data produced, estimated bases, reads generated and bases called.

5B Code use for NGS alignment

Alternatively, the sequencing results were analysed using the command line interface provided by the macOS Big Sur ver.11.6.1 terminal. The code used is shown in the following paragraphs.

Genome Assembly
flye --nano-hq all.fastq --nano-hq all.fastq —asm-coverage 40 --genome-size 12m --threads 8

Assembly polishing a creation of consensus sequence
medaka_consensus -i all.fastq -d assembly.fasta -t 8 -m r941_min_hac_g507

Aligment against reference genome
mini_align -r reference_genome.fasta -i medaka.consensus.fasta

Sorting the Bam files generated from the alignment
samtools sort fastq.bam

Merging the Bam files
samtools merge outputsorted.bam outputsorted2.bam -o output_aligment_merge.bam

5C Plasmids maps inserted in *S. cerevisiae AE1.*

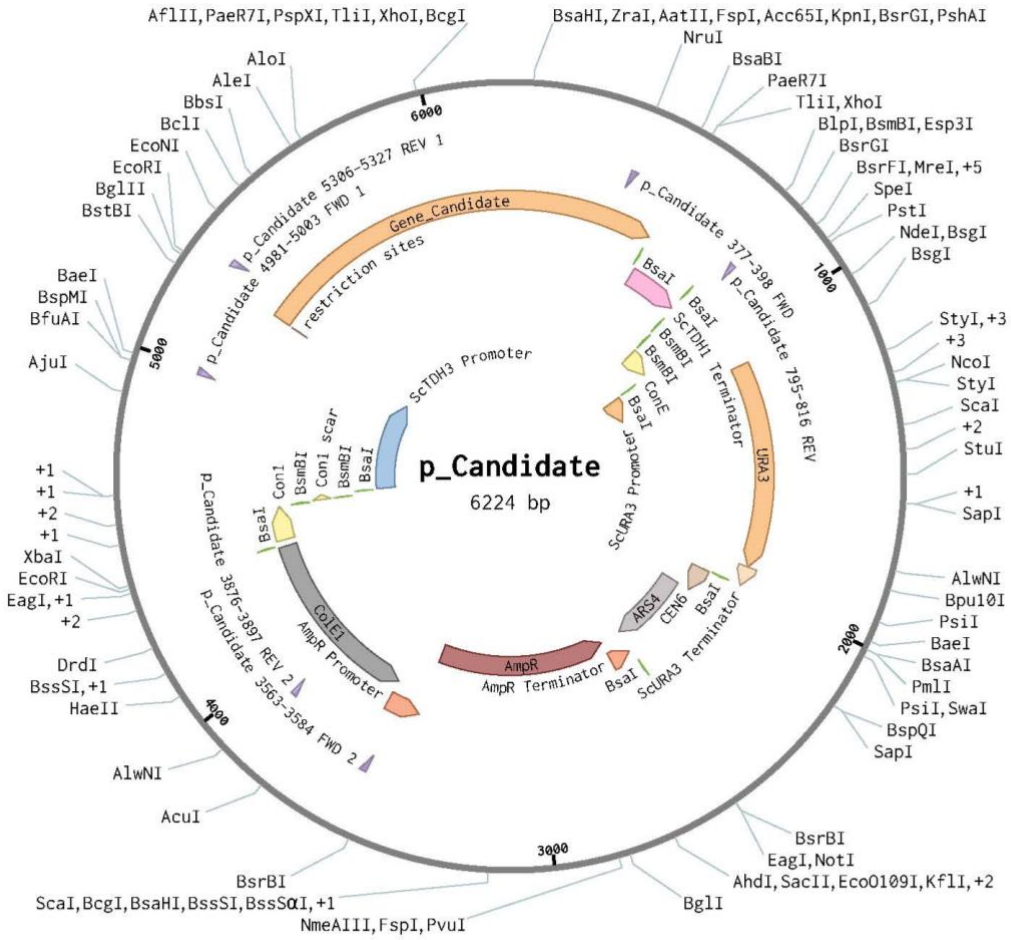


Figure 5C.1: Plasmid Map for *S. cerevisiae AE1* expression of T1βOH enzyme. The system was designed by Ainoa Escrich from Universitat Pompeu Fabra. The image was kindly provided by Ainoa.

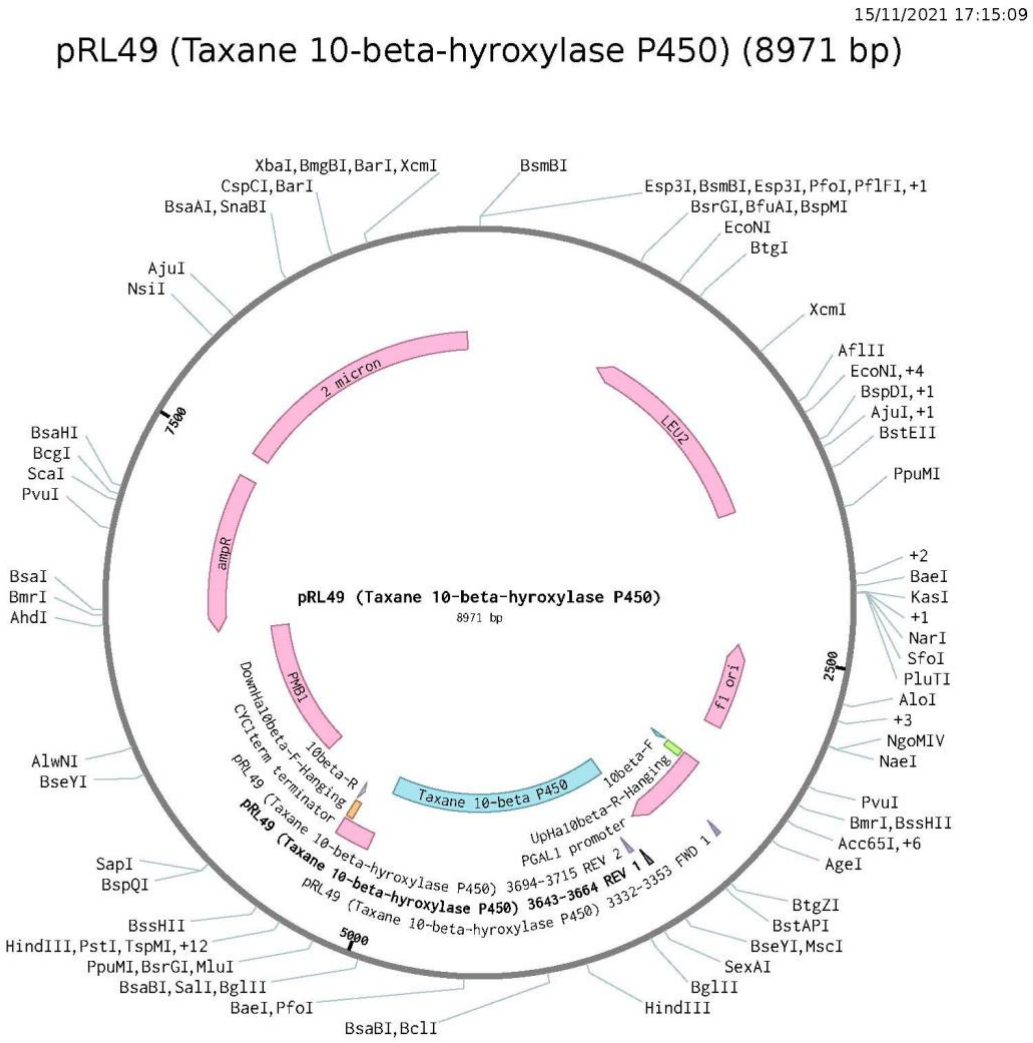
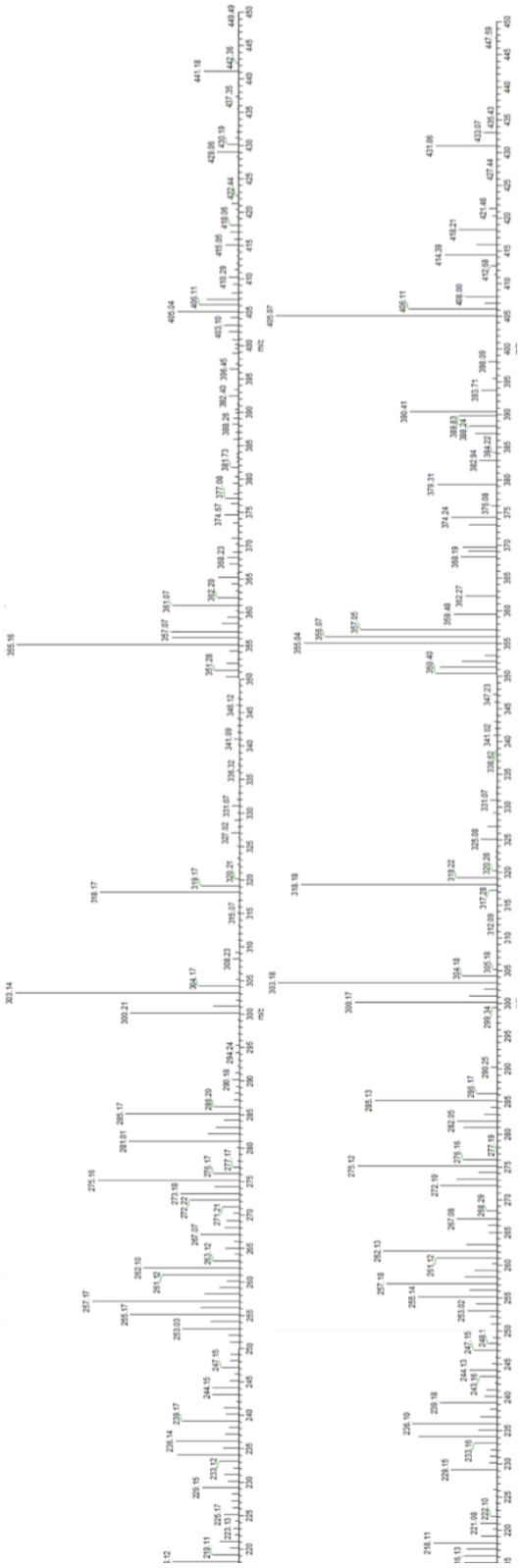


Figure 5C.2: Plasmid Map for *S. cerevisiae AE1* expression of T10βOH enzyme. The system is part of the Leo Rios Lab plasmid repository. The image was kindly provided by Ainoa.

5D List of codon-optimised gene sequences for expression of different P450s and reductases in *E. coli*



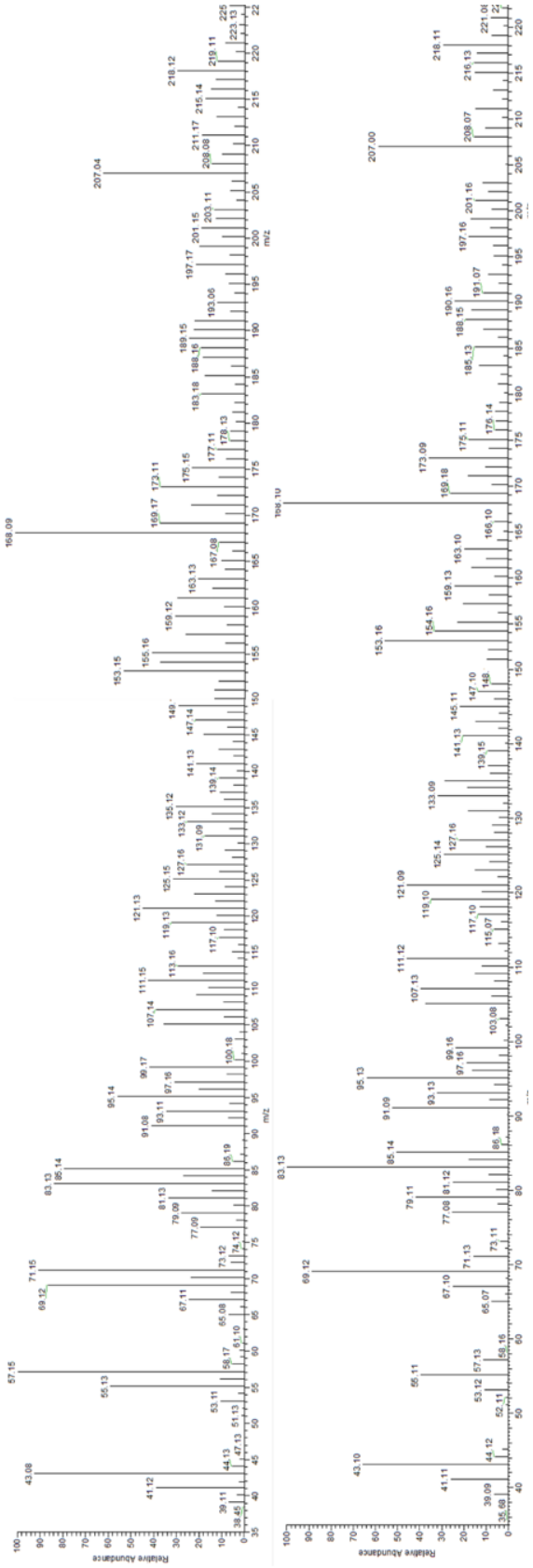


Figure 5C Complete MS spectrum for T1β-ol. The left spectrum corresponds to the consortia with 1:1:3 inoculation ratio while the right one corresponds to the 1:1:6.

REFERENCES

References

Å tagoj, M. N., Comino, A., & Komel, R. (2005). Fluorescence based assay of GAL system in yeast Saccharomyces cerevisiae. *FEMS Microbiology Letters*, *244*(1), 105–110. https://doi.org/10.1016/j.femsle.2005.01.041

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, *9*(1), 1911. https://doi.org/10.1038/s41467-018-04252-2

Ai, L., Guo, W., Chen, W., Teng, Y., & Bai, L. (2019). The gal80 Deletion by CRISPR-Cas9 in Engineered Saccharomyces cerevisiae Produces Artemisinic Acid Without Galactose Induction. *Current Microbiology*. https://doi.org/10.1007/s00284-019-01752-2

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., & Stephanopoulos, G. (2010). Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli. *Science*, *330*(6000), 70–74. https://doi.org/10.1126/science.1191652

Álvarez-Cao, M.-E., Cerdán, M.-E., González-Siso, M.-I., & Becerra, M. (2019). Optimization of Saccharomyces cerevisiae α-galactosidase production and application in the degradation of raffinose family oligosaccharides. *Microbial Cell Factories*, *18*(1), 172. https://doi.org/10.1186/s12934-019-1222-x

Batt, C. A., Caryallo, S., Easson, D. D., Akedo, M., & Sinskey, A. J. (1986). Direct evidence for a xylose metabolic pathway inSaccharomyces cerevisiae. *Biotechnology and Bioengineering*, *28*(4), 549–553. https://doi.org/10.1002/bit.260280411

Biggs, B. W., Lim, C. G., Sagliani, K., Shankar, S., Stephanopoulos, G., De Mey, M., & Ajikumar, P. K. (2016). Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *113*(12), 3209–3214. https://doi.org/10.1073/pnas.1515826113

Bittihn, P., Din, M. O., Tsimring, L. S., & Hasty, J. (2018). Rational engineering of synthetic microbial systems: from single cells to consortia. *Current Opinion in Microbiology*, *45*, 92–99. https://doi.org/10.1016/j.mib.2018.02.009

Cao, Z., Yan, W., Ding, M., & Yuan, Y. (2022). Construction of microbial consortia for microbial degradation of complex compounds. *Frontiers in Bioengineering and Biotechnology*, *10*(December), 1–14. https://doi.org/10.3389/fbioe.2022.1051233

Castillo-Hair, S. M., Baerman, E. A., Fujita, M., Igoshin, O. A., & Tabor, J. J. (2019). Optogenetic control of Bacillus subtilis gene expression. *Nature Communications*, *10*(1), 3099. https://doi.org/10.1038/s41467-019-10906-6

Chang, F., Zhang, X., Pan, Y., Lu, Y., Fang, W., Fang, Z., & Xiao, Y. (2017). Light induced expression of β-glucosidase in Escherichia coli with autolysis of cell. *BMC Biotechnology*, *17*(1), 74. https://doi.org/10.1186/s12896-017-0402-1

Che, S., & Men, Y. (2019). Synthetic microbial consortia for biosynthesis and biodegradation: promises and challenges. *Journal of Industrial Microbiology & Biotechnology*, *0123456789*. https://doi.org/10.1007/s10295-019-02211-4

Chen, T., Wang, X., Zhuang, L., Shao, A., Lu, Y., & Zhang, H. (2021). Development and optimization of a microbial co-culture system for heterologous indigo biosynthesis. *Microbial Cell Factories*, *20*(1), 154. https://doi.org/10.1186/s12934-021-01636-w

Cheng, B.-Q., Wei, L.-J., Lv, Y.-B., Chen, J., & Hua, Q. (2019). Elevating Limonene Production in Oleaginous Yeast Yarrowia lipolytica via Genetic Engineering of Limonene Biosynthesis Pathway and Optimization of Medium Composition. *Biotechnology and Bioprocess Engineering*, *24*(3), 500–506. https://doi.org/10.1007/s12257-018-0497-9

DeJong, J. M., Liu, Y., Bollon, A. P., Long, R. M., Jennewein, S., Williams, D., & Croteau, R. B. (2006). Genetic engineering of taxol biosynthetic genes inSaccharomyces cerevisiae. *Biotechnology and Bioengineering*, *93*(2), 212–224. https://doi.org/10.1002/bit.20694

Det-udom, R., Gilbert, C., Liu, L., Prakitchaiwattana, C., Ellis, T., & Ledesma-Amaro, R. (2019). Towards semi-synthetic microbial communities: enhancing soy sauce fermentation properties in B. subtilis co-cultures. *Microbial Cell Factories*, *18*(1), 101. https://doi.org/10.1186/s12934-019-1149-2

Dibrell, S. E., Tao, Y., & Reisman, S. E. (2021). Synthesis of Complex Diterpenes: Strategies Guided by Oxidation Pattern Analysis. *Accounts of Chemical Research*, *54*(6), 1360–1373. https://doi.org/10.1021/acs.accounts.0c00858

Duncker, K. E., Holmes, Z. A., & You, L. (2021). Engineered microbial consortia: strategies and applications. *Microbial Cell Factories*, *20*(1), 211. https://doi.org/10.1186/s12934-021-01699-9

Engels, B., Dahm, P., & Jennewein, S. (2008). Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metabolic Engineering*, *10*(3–4), 201–206. https://doi.org/10.1016/j.ymben.2008.03.001

Ensinck, I., Maman, A., Albihlal, W., Lassandro, M., Salzano, G., Sideri, T., Howell, S., Calvani, E., Patel, H., Guy Bushkin, G., Ralser, M., Snijders, A., Skehel, M., Casañal, A., Schwartz, S., & van Werven, F. (2023). *The yeast RNA methylation complex consists of conserved yet reconfigured components with m6A-dependent and independent roles*. bioRxiv. https://doi.org/10.1101/2023.02.10.528004

Escalante-Chong, R., Savir, Y., Carroll, S. M., Ingraham, J. B., Wang, J., Marx, C. J., & Springer, M. (2015). Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proceedings of the National Academy of Sciences*, *112*(5), 1636–1641. https://doi.org/10.1073/pnas.1418058112

Escrich-Montana, A. (2022). *Paclitaxel: New insights into its regulation, biosynthesis and biotechnological production*. Universitat Pompeu Fabra.

Faust, K., & Raes, J. (2012). Microbial interactions: from networks to models. *Nature Reviews Microbiology*, *10*(8), 538–550. https://doi.org/10.1038/nrmicro2832

Fedorec, A. J. H., Karkaria, B. D., Sulu, M., & Barnes, C. P. (2021). Single strain control of microbial consortia. *Nature Communications*, *12*(1), 1–12. https://doi.org/10.1038/s41467-021-22240-x

Fischedick, J. T., Johnson, S. R., Ketchum, R. E. B., Croteau, R. B., & Lange, B. M. (2015). NMR spectroscopic search module for Spektraris, an online resource for plant natural product identification – Taxane diterpenoids from Taxus×media cell suspension cultures as a case study. *Phytochemistry*, *113*, 87–95. https://doi.org/10.1016/j.phytochem.2014.11.020

Flores, A. D., Ayla, E. Z., Nielsen, D. R., & Wang, X. (2019a). Engineering a Synthetic, Catabolically-Orthogonal Co-Culture System for Enhanced Conversion of Lignocellulose-Derived Sugars to Ethanol [Research-article]. *ACS Synthetic Biology*, *8*(5), acssynbio.9b00007. https://doi.org/10.1021/acssynbio.9b00007

Flores, A. D., Ayla, E. Z., Nielsen, D. R., & Wang, X. (2019b). Engineering a Synthetic, Catabolically Orthogonal Coculture System for Enhanced Conversion of Lignocellulose-Derived Sugars to Ethanol [Research-article]. *ACS Synthetic Biology*, *8*, 1089–1099. https://doi.org/10.1021/acssynbio.9b00007

Gallego-Jara, J., Lozano-Terol, G., Sola-Martínez, R. A., Cánovas-Díaz, M., & de Diego Puente, T. (2020). A Compressive Review about Taxol®: History and Future Challenges. *Molecules*, *25*(24), 5986. https://doi.org/10.3390/molecules25245986

Ganesan, V., Li, Z., Wang, X., & Zhang, H. (2017). Heterologous biosynthesis of natural product naringenin by co-culture engineering. *Synthetic and Systems Biotechnology*, *2*(3), 236–242. https://doi.org/10.1016/j.synbio.2017.08.003

Gao, C.-H., Cao, H., Cai, P., & Sørensen, S. J. (2021). The initial inoculation ratio regulates bacterial coculture interactions and metabolic capacity. *The ISME Journal*, *15*(1), 29–40. https://doi.org/10.1038/s41396-020-00751-7

García-Jiménez, B., Torres-Bacete, J., & Nogales, J. (2021). Metabolic modelling approaches for describing and engineering microbial communities. *Computational and Structural Biotechnology Journal*, *19*, 226–246. https://doi.org/10.1016/j.csbj.2020.12.003

Gasmi, N., Jacques, P. E., Klimova, N., Guo, X., Ricciardi, A., Robert, F., & Turcotte, B. (2014). The switch from fermentation to respiration in Saccharomyces cerevisiae is regulated by the Ert1 transcriptional activator/repressor. *Genetics*, *198*(2), 547–560. https://doi.org/10.1534/genetics.114.168609

Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodríguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M. L., Mattanovich, D., & Villaverde, A. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microbial Cell Factories*, *7*(1), 11. https://doi.org/10.1186/1475-2859-7-11

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 31–34. https://doi.org/10.1038/nprot.2007.13

Guido, N., Starostina, E., Leake, D., & Saaem, I. (2016). Improved PCR Amplification of Broad Spectrum GC DNA Templates. *PLOS ONE*, *11*(6), e0156478. https://doi.org/10.1371/journal.pone.0156478

Hao, D. C., Ge, G., Xiao, P., Zhang, Y., & Yang, L. (2011). The First Insight into the Tissue Specific Taxus Transcriptome via Illumina Second Generation Sequencing. *PLoS ONE*, *6*(6), e21220. https://doi.org/10.1371/journal.pone.0021220

Hassoun, S., Jefferson, F., Shi, X., Stucky, B., Wang, J., & Rosa, E. (2022). Artificial

Intelligence for Biology. *Integrative and Comparative Biology*, *61*(6), 2267–2275. https://doi.org/10.1093/icb/icab188

Hayat, I. F., Plan, M., Ebert, B. E., Dumsday, G., Vickers, C. E., & Peng, B. (2021). Auxin-mediated induction of GAL promoters by conditional degradation of Mig1p improves sesquiterpene production in Saccharomyces cerevisiae with engineered acetyl-CoA synthesis. *Microbial Biotechnology*, *14*(6), 2627–2642. https://doi.org/10.1111/1751-7915.13880

Hefner, J., Rubenstein, S. M., Ketchum, R. E. B., Gibson, D. M., Williams, R. M., & Croteau, R. (1996). Cytochrome P450-catalyzed hydroxylation of taxa-4(5),11(12)-diene to taxa-4(20),11(12)-dien-5a-o1: the first oxygenation step in taxol biosynthesis. *Chemistry & Biology*, *3*(6), 479–489. https://doi.org/10.1016/S1074-5521(96)90096-4

Hitzeman, R. A., Hagie, F. E., Levine, H. L., Goeddel, D. V., Ammerer, G., & Hall, B. D. (1981). Expression of a human gene for interferon in yeast. *Nature*, *293*(5835), 717–722. https://doi.org/10.1038/293717a0

Hu, B., Wang, M., Geng, S., Wen, L., Wu, M., Nie, Y., Tang, Y.-Q., & Wu, X.-L. (2020). Metabolic Exchange with Non-Alkane-Consuming Pseudomonas stutzeri SLG510A3-8 Improves n -Alkane Biodegradation by the Alkane Degrader Dietzia sp. Strain DQ12-45-1b. *Applied and Environmental Microbiology*, *86*(8). https://doi.org/10.1128/AEM.02931-19

Huang, Q., Roessner, C. A., Croteau, R., & Scott, A. I. (2001). Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Bioorganic and Medicinal Chemistry*, *9*(9), 2237–2242. https://doi.org/10.1016/S0968-0896(01)00072-4

Huang, Y., Xia, A., Yang, G., & Jin, F. (2018). Bioprinting Living Biofilms through Optogenetic Manipulation. *ACS Synthetic Biology*, *7*(5), 1195–1200. https://doi.org/10.1021/acssynbio.8b00003

Hueso-Gil, A., Nyerges, Á., Pál, C., Calles, B., & de Lorenzo, V. (2020). Multiple-Site Diversification of Regulatory Sequences Enables Interspecies Operability of Genetic Devices. *ACS Synthetic Biology*, *9*(1), 104–114. https://doi.org/10.1021/acssynbio.9b00375

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., & Boyer, H. W. (1977). Expression in Escherichia coli of a Chemically Synthesized Gene for the

Hormone Somatostatin. *Science*, *198*(4321), 1056–1063. https://doi.org/10.1126/science.412251

Jin, Y.-S., Laplaza, J. M., & Jeffries, T. W. (2004). Saccharomyces cerevisiae Engineered for Xylose Metabolism Exhibits a Respiratory Response. *Applied and Environmental Microbiology*, *70*(11), 6816–6825. https://doi.org/10.1128/AEM.70.11.6816-6825.2004

Jonguitud-Borrego, N., Malcı, K., Anand, M., Baluku, E., Webb, C., Liang, L., Barba-Ostria, C., Guaman, L. P., Hui, L., & Rios-Solis, L. (2022). High—throughput and automated screening for COVID-19. *Frontiers in Medical Technology*, *4*. https://doi.org/10.3389/fmedt.2022.969203

Kapoore, R. V., Padmaperuma, G., Maneein, S., & Vaidyanathan, S. (2022). Co-culturing microbial consortia: approaches for applications in biomanufacturing and bioprocessing. *Critical Reviews in Biotechnology*, *42*(1), 46–72. https://doi.org/10.1080/07388551.2021.1921691

Kaspera, R., & Croteau, R. (2006). Cytochrome P450 oxygenases of Taxol biosynthesis. *Phytochemistry Reviews*, *5*(2–3), 433–444. https://doi.org/10.1007/s11101-006-9006-4

Kingston, D. G. I. (2001). Taxol, a molecule for all seasons. *Chemical Communications*, *10*, 867–880. https://doi.org/10.1039/b100070p

Kitaoka, N., Wu, Y., Xu, M., & Peters, R. J. (2015). Optimization of recombinant expression enables discovery of novel cytochrome P450 activity in rice diterpenoid biosynthesis. *Applied Microbiology and Biotechnology*, *99*(18), 7549–7558. https://doi.org/10.1007/s00253-015-6496-2

Klukowski, P., Riek, R., & Güntert, P. (2022). Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA. *Nature Communications*, *13*(1), 6151. https://doi.org/10.1038/s41467-022-33879-5

Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine*, *12*(1), 91. https://doi.org/10.1186/s13073-020-00791-w

Kong, W., Meldgin, D. R., Collins, J. J., & Lu, T. (2018). Designing microbial consortia with defined social interactions. *Nature Chemical Biology*, *14*(8), 821–829. https://doi.org/10.1038/s41589-018-0091-7

Kuang, X., Sun, S., Wei, J., Li, Y., & Sun, C. (2019). Iso-Seq analysis of the Taxus cuspidata transcriptome reveals the complexity of Taxol biosynthesis. *BMC Plant Biology*, *19*(1),

210. https://doi.org/10.1186/s12870-019-1809-8

Kusari, S., Singh, S., & Jayabaskaran, C. (2014). Rethinking production of Taxol® (paclitaxel) using endophyte biotechnology. *Trends in Biotechnology*, *32*(6), 304–311. https://doi.org/10.1016/j.tibtech.2014.03.011

Lalwani, M. A., Kawabe, H., Mays, R. L., Hoffman, S. M., & Avalos, J. L. (2021). Optogenetic Control of Microbial Consortia Populations for Chemical Production. *ACS Synthetic Biology*, *10*(8), 2015–2029. https://doi.org/10.1021/acssynbio.1c00182

Li, X., Zhou, Z., Li, W., Yan, Y., Shen, X., Wang, J., Sun, X., & Yuan, Q. (2022). Design of stable and self-regulated microbial consortia for chemical synthesis. *Nature Communications*, *13*(1), 1554. https://doi.org/10.1038/s41467-022-29215-6

Li, Z., Wang, X., & Zhang, H. (2019). Balancing the non-linear rosmarinic acid biosynthetic pathway by modular co-culture engineering. *Metabolic Engineering*, *54*, 1–11. https://doi.org/10.1016/j.ymben.2019.03.002

Liljeström, P. L. (1985). The nucleotide sequence of the yeast MEL1 gene. *Nucleic Acids Research*, *13*(20), 7257–7268. https://doi.org/10.1093/nar/13.20.7257

Lischer, H. E. L., & Shimizu, K. K. (2017). Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*, *18*(1), 474. https://doi.org/10.1186/s12859-017-1911-6

Liu, L., Mohammadifar, M., Elhadad, A., Tahernia, M., Zhang, Y., Zhao, W., & Choi, S. (2021). Spatial Engineering of Microbial Consortium for Long-Lasting, Self-Sustaining, and High-Power Generation in a Bacteria-Powered Biobattery. *Advanced Energy Materials*, *11*(22), 2100713. https://doi.org/10.1002/aenm.202100713

Liu, Y.-T., Sau, S., Ma, C.-H., Kachroo, A. H., Rowley, P. A., Chang, K.-M., Fan, H.-F., & Jayaram, M. (2014). The Partitioning and Copy Number Control Systems of the Selfish Yeast Plasmid: An Optimized Molecular Design for Stable Persistence in Host Cells. *Microbiology Spectrum*, *2*(5). https://doi.org/10.1128/microbiolspec.PLAS-0003-2013

Liu, Y., Ding, M., Ling, W., Yang, Y., Zhou, X., Li, B.-Z., Chen, T., Nie, Y., Wang, M., Zeng, B., Li, X., Liu, H., Sun, B., Xu, H., Zhang, J., Jiao, Y., Hou, Y., Yang, H., Xiao, S., et al. (2017). A three-species microbial consortium for power generation. *Energy & Environmental Science*, *10*(7), 1600–1609. https://doi.org/10.1039/C6EE03705D

Lodolo, E. J., Kock, J. L. F., Axcell, B. C., & Brooks, M. (2008). The yeast Saccharomyces

cerevisiae- the main character in beer brewing. *FEMS Yeast Research*, *8*(7), 1018–1036. https://doi.org/10.1111/j.1567-1364.2008.00433.x

Lohr, D., Venkov, P., & Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *The FASEB Journal*, *9*(9), 777–787. https://doi.org/10.1096/fasebj.9.9.7601342

Lu, B., Zeng, Z., & Shi, T. (2013). Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. *Science China Life Sciences*, *56*(2), 143–155. https://doi.org/10.1007/s11427-013-4442-z

Malcı, K., Walls, L. E., & Rios-Solis, L. (2020). Multiplex Genome Engineering Methods for Yeast Cell Factory Development. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.589468

Malcı, K., Walls, L. E., & Rios-Solis, L. (2022). Rational Design of CRISPR/Cas12a-RPA Based One-Pot COVID-19 Detection with Design of Experiments. *ACS Synthetic Biology*, *11*(4), 1555–1567. https://doi.org/10.1021/acssynbio.1c00617

Malcl, K., Jonguitud-Borrego, N., Van Der Straten Waillet, H., Puodžiū Naitė, U., Johnston, E. J., Rosser, S. J., & Rios-Solis, L. (2022). ACtivE: Assembly and CRISPR-Targeted in Vivo Editing for Yeast Genome Engineering Using Minimum Reagents and Time. *ACS Synthetic Biology*, *11*(11), 3629–3643. https://doi.org/10.1021/acssynbio.2c00175

Marchioro, V., Steinmetz, R. L. R., do Amaral, A. C., Gaspareto, T. C., Treichel, H., & Kunz, A. (2018). Poultry Litter Solid State Anaerobic Digestion: Effect of Digestate Recirculation Intervals and Substrate/Inoculum Ratios on Process Efficiency. *Frontiers in Sustainable Food Systems*, *2*. https://doi.org/10.3389/fsufs.2018.00046

Matsumoto, A., Terashima, I., & Uesono, Y. (2022). A rapid and simple spectroscopic method for the determination of yeast cell viability using methylene blue. *Yeast*, *39*(11–12), 607–616. https://doi.org/10.1002/yea.3819

McCarty, N. S., & Ledesma-Amaro, R. (2019). Synthetic Biology Tools to Engineer Microbial Communities for Biotechnology. *Trends in Biotechnology*, *37*(2), 181–197. https://doi.org/10.1016/j.tibtech.2018.11.002

Miller, G. L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, *31*(3), 426–428. https://doi.org/10.1021/ac60147a030

Mishra, A., & Ghosh, S. (2020). Saccharification of kans grass biomass by a novel fractional

hydrolysis method followed by co-culture fermentation for bioethanol production. *Renewable Energy*, *146*, 750–759. https://doi.org/10.1016/j.renene.2019.07.016

Mishra, M., Singh, S. K., & Kumar, A. (2021). Microbial consortia: approaches in crop production and yield enhancement. In *Microbiome Stimulants for Crops* (pp. 293–303). Elsevier. https://doi.org/10.1016/B978-0-12-822122-8.00013-3

Möglich, A., Ayers, R. A., & Moffat, K. (2009). Design and Signaling Mechanism of Light-Regulated Histidine Kinases. *Journal of Molecular Biology*, *385*(5), 1433–1444. https://doi.org/10.1016/j.jmb.2008.12.017

Moreno, J., & Peinado, R. (2012). Sugars: Structure and Classification. *Enological Chemistry*, 77–93. https://doi.org/10.1016/B978-0-12-388438-1.00006-6

Mutanda, I., Li, J., Xu, F., & Wang, Y. (2021). Recent Advances in Metabolic Engineering, Protein Engineering, and Transcriptome-Guided Insights Toward Synthetic Production of Taxol. *Frontiers in Bioengineering and Biotechnology*, *9*. https://doi.org/10.3389/fbioe.2021.632269

Naumov, G. I., Naumova, E. S., Turakainen, H., & Korhola, M. (1996). Identification of the α-galactosidase MEL genes in some populations of Saccharomyces cerevisiae : a new gene MEL11. *Genetical Research*, *67*(2), 101–108. https://doi.org/10.1017/S0016672300033565

Navlakha, S., & Bar-Joseph, Z. (2011). Algorithms in nature: the convergence of systems biology and computational thinking. *Molecular Systems Biology*, *7*(1), 546. https://doi.org/10.1038/msb.2011.78

Nehlin, J. O., Carlberg, M., & Ronne, H. (1991). Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. *The EMBO Journal*, *10*(11), 3373–3377. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC453065/?page=2

Nicolaou, K. C., Yang, Z., Liu, J. J., Ueno, H., Nantermet, P. G., Guy, R. K., Claiborne, C. F., Renaud, J., Couladouros, E. A., Paulvannan, K., & Sorensen, E. J. (1994). Total synthesis of taxol. *Nature*, *367*(6464), 630–634. https://doi.org/10.1038/367630a0

Nikel, P. I., & de Lorenzo, V. (2014). Robustness of Pseudomonas putida KT2440 as a host for ethanol biosynthesis. *New Biotechnology*, *31*(6), 562–571. https://doi.org/10.1016/j.nbt.2014.02.006

Nowrouzi, B., Li, R. A., Walls, L. E., d'Espaux, L., Malcı, K., Liang, L., Jonguitud-Borrego,

N., Lerma-Escalera, A. I., Morones-Ramirez, J. R., Keasling, J. D., & Rios-Solis, L. (2020). Enhanced production of taxadiene in Saccharomyces cerevisiae. *Microbial Cell Factories*, *19*(1). https://doi.org/10.1186/s12934-020-01458-2

Nowrouzi, B., Lungang, L., & Rios-Solis, L. (2022). Exploring optimal Taxol® CYP725A4 activity in Saccharomyces cerevisiae. *Microbial Cell Factories*, *21*(1), 197. https://doi.org/10.1186/s12934-022-01922-1

Ohlendorf, R., Vidavski, R. R., Eldar, A., Moffat, K., & Möglich, A. (2012). From Dusk till Dawn: One-Plasmid Systems for Light-Regulated Gene Expression. *Journal of Molecular Biology*, *416*(4), 534–542. https://doi.org/10.1016/j.jmb.2012.01.001

Paulo, J. A., O'Connell, J. D., Gaun, A., & Gygi, S. P. (2015). Proteome-wide quantitative multiplexed profiling of protein expression: carbon-source dependency in Saccharomyces cerevisiae. *Molecular Biology of the Cell*, *26*(22), 4063–4074. https://doi.org/10.1091/mbc.E15-07-0499

Pego, J. V., & Smeekens, S. C. M. (2000). Plant fructokinases: A sweet family get-together. *Trends in Plant Science*, *5*(12), 531–536. https://doi.org/10.1016/S1360-1385(00)01783-0

Peng, B., Wood, R. J., Nielsen, L. K., & Vickers, C. E. (2018). An Expanded Heterologous GAL Promoter Collection for Diauxie-Inducible Expression in Saccharomyces cerevisiae. *ACS Synthetic Biology*, *7*(2), 748–751. https://doi.org/10.1021/acssynbio.7b00355

Pirhanov, A., Bridges, C. M., Goodwin, R. A., Guo, Y.-S., Furrer, J., Shor, L. M., Gage, D. J., & Cho, Y. K. (2021). Optogenetics in Sinorhizobium meliloti Enables Spatial Control of Exopolysaccharide Production and Biofilm Structure. *ACS Synthetic Biology*, *10*(2), 345–356. https://doi.org/10.1021/acssynbio.0c00498

Pu, L., Yang, S., Xia, A., & Jin, F. (2018). Optogenetics Manipulation Enables Prevention of Biofilm Formation of Engineered Pseudomonas aeruginosa on Surfaces. *ACS Synthetic Biology*, *7*(1), 200–208. https://doi.org/10.1021/acssynbio.7b00273

Qian, X., Chen, L., Sui, Y., Chen, C., Zhang, W., Zhou, J., Dong, W., Jiang, M., Xin, F., & Ochsenreither, K. (2019). Biotechnological potential and applications of microbial consortia. *Biotechnology Advances*, *November*, 107500. https://doi.org/10.1016/j.biotechadv.2019.107500

Rajeshkannan, Mahilkar, A., & Saini, S. (2022). GAL Regulon in the Yeast S. cerevisiae is Highly Evolvable via Acquisition in the Coding Regions of the Regulatory Elements of the Network. *Frontiers in Molecular Biosciences*, *9*. https://doi.org/10.3389/fmolb.2022.801011

Ramakrishnan, P., & Tabor, J. J. (2016). Repurposing Synechocystis PCC6803 UirS–UirR as a UV-Violet/Green Photoreversible Transcriptional Regulatory Tool in E. coli. *ACS Synthetic Biology*, *5*(7), 733–740. https://doi.org/10.1021/acssynbio.6b00068

Ramírez-Estrada, K., Altabella, T., Onrubia, M., Moyano, E., Notredame, C., Osuna, L., Vanden Bossche, R., Goossens, A., Cusido, R. M., & Palazon, J. (2016). Transcript profiling of jasmonate-elicited Taxus cells reveals a β-phenylalanine-CoA ligase. *Plant Biotechnology Journal*, *14*(1), 85–96. https://doi.org/10.1111/pbi.12359

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Roell, G. W., Zha, J., Carr, R. R., Koffas, M. A., Fong, S. S., & Tang, Y. J. (2019). Engineering microbial consortia by division of labor. *Microbial Cell Factories*, *18*(1), 1–11. https://doi.org/10.1186/s12934-019-1083-3

Sanchez-Muñoz, R., Perez-Mata, E., Almagro, L., Cusido, R. M., Bonfill, M., Palazon, J., & Moyano, E. (2020). A Novel Hydroxylation Step in the Taxane Biosynthetic Pathway: A New Approach to Paclitaxel Production by Synthetic Biology. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.00410

Sanchez, A., Bajic, D., Diaz-Colunga, J., Skwara, A., Vila, J. C. C., & Kuehn, S. (2023). The community-function landscape of microbial consortia. *Cell Systems*, *14*(2), 122–134. https://doi.org/10.1016/j.cels.2022.12.011

Santoyo-Garcia, J. H., Walls, L. E., Nowrouzi, B., Galindo-Rodriguez, G. R., Ochoa-Villarreal, M., Loake, G. J., Dimartino, S., & Rios-Solis, L. (2022). In situ solid-liquid extraction enhances recovery of taxadiene from engineered Saccharomyces cerevisiae cell factories. *Separation and Purification Technology*, *290*(July 2021), 120880. https://doi.org/10.1016/j.seppur.2022.120880

Schoendorf, A., Rithner, C. D., Williams, R. M., & Croteau, R. B. (2001). Molecular cloning of a cytochrome P450 taxane 10β-hydroxylase cDNA from Taxus and functional expression in yeast. *Proceedings of the National Academy of Sciences of the United*

States of America, *98*(4), 1501–1506. https://doi.org/10.1073/pnas.98.4.1501

Schuler, M. A., & Werck-Reichhart, D. (2003). Functional Genomics of P450S. *Annual Review of Plant Biology*, *54*(1), 629–667. https://doi.org/10.1146/annurev.arplant.54.031902.134840

Sgobba, E., Stumpf, A. K., Vortmann, M., Jagmann, N., Krehenbrink, M., Dirks-Hofmeister, M. E., Moerschbacher, B., Philipp, B., & Wendisch, V. F. (2018). Synthetic Escherichia coli-Corynebacterium glutamicum consortia for l-lysine production from starch and sucrose. *Bioresource Technology*, *260*, 302–310. https://doi.org/10.1016/j.biortech.2018.03.113

Sharma, A., Bhatia, S. K., Banyal, A., Chanana, I., Kumar, A., Chand, D., Kulshrestha, S., & Kumar, P. (2022). An Overview on Taxol Production Technology and Its Applications as Anticancer Agent. *Biotechnology and Bioprocess Engineering*, *27*(5), 706–728. https://doi.org/10.1007/s12257-022-0063-3

Shong, J., Jimenez Diaz, M. R., & Collins, C. H. (2012). Towards synthetic microbial consortia for bioprocessing. *Current Opinion in Biotechnology*, *23*(5), 798–802. https://doi.org/10.1016/j.copbio.2012.02.001

Tian, F., Wang, Y., Guo, G., Ding, K., Yang, F., Wang, H., Cao, Y., & Liu, C. (2021). Enhanced azo dye biodegradation at high salinity by a halophilic bacterial consortium. *Bioresource Technology*, *326*, 124749. https://doi.org/10.1016/j.biortech.2021.124749

Toivari, M. H., Salusjärvi, L., Ruohonen, L., & Penttilä, M. (2004). Endogenous Xylose Pathway in Saccharomyces cerevisiae. *Applied and Environmental Microbiology*, *70*(6), 3681–3686. https://doi.org/10.1128/AEM.70.6.3681-3686.2004

Toyomasu, T., & Sassa, T. (2010). Diterpenes. In *Comprehensive Natural Products II* (pp. 643–672). Elsevier. https://doi.org/10.1016/B978-008045382-8.00006-X

Treloar, N. J., Fedorec, A. J. H., Ingalls, B., & Barnes, C. P. (2020). Deep reinforcement learning for the control of microbial co-cultures in bioreactors. *PLOS Computational Biology*, *16*(4), e1007783. https://doi.org/10.1371/journal.pcbi.1007783

Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology*, *7*. https://doi.org/10.3389/fbioe.2019.00420

Tschirhart, T., Shukla, V., Kelly, E. E., Schultzhaus, Z., NewRingeisen, E., Erickson, J. S.,

Wang, Z., Garcia, W., Curl, E., Egbert, R. G., Yeung, E., & Vora, G. J. (2019). Synthetic Biology Tools for the Fast-Growing Marine Bacterium Vibrio natriegens. *ACS Synthetic Biology*, *8*(9), 2069–2079. https://doi.org/10.1021/acssynbio.9b00176

Tsoi, R., Dai, Z., & You, L. (2019). Emerging strategies for engineering microbial communities. *Biotechnology Advances*, *March*, 0–1. https://doi.org/10.1016/J.BIOTECHADV.2019.03.011

Valenzuela, P., Medina, A., Rutter, W. J., Ammerer, G., & Hall, B. D. (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature*, *298*(5872), 347–350. https://doi.org/10.1038/298347a0

van Zyl, C., Prior, B. A., Kilian, S. G., & Brandt, E. V. (1993). Role of D-ribose as a cometabolite in D-xylose metabolism by Saccharomyces cerevisiae. *Applied and Environmental Microbiology*, *59*(5), 1487–1494. https://doi.org/10.1128/aem.59.5.1487-1494.1993

Van Zyl, C., Prior, B. A., Kilian, S. G., & Kock, J. L. F. (1989). D-Xylose Utilization by Saccharomyces cerevisiae. *Microbiology*, *135*(11), 2791–2798. https://doi.org/10.1099/00221287-135-11-2791

Venkataram, S., Kuo, H.-Y., Hom, E. F. Y., & Kryazhimskiy, S. (2023). Mutualism-enhancing mutations dominate early adaptation in a two-species microbial community. *Nature Ecology & Evolution*, 2021.07.07.451547. https://doi.org/10.1038/s41559-022-01923-8

Venkateswaran, K., Moser, D. P., Dollhopf, M. E., Lies, D. P., Saffarini, D. A., MacGregor, B. J., Ringelberg, D. B., White, D. C., Nishijima, M., Sano, H., Burghardt, J., Stackebrandt, E., & Nealson, K. H. (1999). Polyphasic taxonomy of the genus Shewanella and description of Shewanella oneidensis sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *49*(2), 705–724. https://doi.org/10.1099/00207713-49-2-705

Vickers, C. E., Williams, T. C., Peng, B., & Cherry, J. (2017). Recent advances in synthetic biology for engineering isoprenoid production in yeast. *Current Opinion in Chemical Biology*, *40*(Figure 1), 47–56. https://doi.org/10.1016/j.cbpa.2017.05.017

Walls, L. E., Malcı, K., Nowrouzi, B., Li, R. A., D'Espaux, L., Wong, J., Dennis, J. A., Semião, A. J. C., Wallace, S., Martinez, J. L., Keasling, J. D., & Rios-Solis, L. (2021). Optimizing the biosynthesis of oxygenated and acetylated Taxol precursors in

Saccharomyces cerevisiae using advanced bioprocessing strategies. *Biotechnology and Bioengineering*, *118*(1), 279–293. https://doi.org/10.1002/bit.27569

Walls, L. E., Martinez, J. L., & Rios-Solis, L. (2022). Enhancing Saccharomyces cerevisiae Taxane Biosynthesis and Overcoming Nutritional Stress-Induced Pseudohyphal Growth. *Microorganisms*, *10*(1), 163. https://doi.org/10.3390/microorganisms10010163

Wang, G., Lu, X., Zhu, Y., Zhang, W., Liu, J., Wu, Y., Yu, L., Sun, D., & Cheng, F. (2018). A light-controlled cell lysis system in bacteria. *Journal of Industrial Microbiology and Biotechnology*, *45*(6), 429–432. https://doi.org/10.1007/s10295-018-2034-4

Wang, J., Lu, X., Ying, H., Ma, W., Xu, S., Wang, X., Chen, K., & Ouyang, P. (2018). A Novel Process for Cadaverine Bio-Production Using a Consortium of Two Engineered Escherichia coli. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.01312

Wang, S., Tang, H., Peng, F., Yu, X., Su, H., Xu, P., & Tan, T. (2019). Metabolite-based mutualism enhances hydrogen production in a two-species microbial consortium. *Communications Biology*, *2*(1), 1–11. https://doi.org/10.1038/s42003-019-0331-8

Wang, T., Li, L., Zhuang, W., Zhang, F., Shu, X., Wang, N., & Wang, Z. (2021). Recent research progress in taxol biosynthetic pathway and acylation reactions mediated by taxus acyltransferases. *Molecules*, *26*(10), 1–13. https://doi.org/10.3390/molecules26102855

Wondraczek, L., Pohnert, G., Schacher, F. H., Köhler, A., Gottschaldt, M., Schubert, U. S., Küsel, K., & Brakhage, A. A. (2019). Artificial Microbial Arenas: Materials for Observing and Manipulating Microbial Consortia. *Advanced Materials*, *31*(24), 1900284. https://doi.org/10.1002/adma.201900284

Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., & Koffas, M. A. G. (2016). Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends in Biotechnology*, *34*(8), 652–664. https://doi.org/10.1016/j.tibtech.2016.02.010

Wu, X.-L., Bi, Y.-H., Gao, F., Xie, Z.-X., Li, X., Zhou, X., Ma, D.-J., Li, B.-Z., & Yuan, Y.-J. (2019). The effect of autonomously replicating sequences on gene expression in saccharomyces cerevisiae. *Biochemical Engineering Journal*, *149*, 107250. https://doi.org/10.1016/j.bej.2019.107250

Xia, A., Qian, M., Wang, C., Huang, Y., Liu, Z., Ni, L., & Jin, F. (2021). Optogenetic

Modification of Pseudomonas aeruginosa Enables Controllable Twitching Motility and Host Infection. *ACS Synthetic Biology*, *10*(3), 531–541. https://doi.org/10.1021/acssynbio.0c00559

Xu, P. (2021). Dynamics of microbial competition, commensalism, and cooperation and its implications for coculture and microbiome engineering. *Biotechnology and Bioengineering*, *118*(1), 199–209. https://doi.org/10.1002/bit.27562

Yadav, V. G. (2014). Unraveling the multispecificity and catalytic promiscuity of taxadiene monooxygenase. *Journal of Molecular Catalysis B: Enzymatic*, *110*, 154–164. https://doi.org/10.1016/j.molcatb.2014.10.004

Zhang, H., Boghigian, B. A., Armando, J., & Pfeifer, B. A. (2011). Methods and options for the heterologous production of complex natural products. *Nat. Prod. Rep.*, *28*(1), 125–151. https://doi.org/10.1039/C0NP00037J

Zhang, J., Hansen, L. G., Gudich, O., Viehrig, K., Lassen, L. M. M., Schrübbers, L., Adhikari, K. B., Rubaszka, P., Carrasquer-Alvarez, E., Chen, L., D'Ambrosio, V., Lehka, B., Haidar, A. K., Nallapareddy, S., Giannakou, K., Laloux, M., Arsovska, D., Jørgensen, M. A. K., Chan, L. J. G., et al. (2022). A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature*, *609*(7926), 341–347. https://doi.org/10.1038/s41586-022-05157-3

Zhang, L., Liu, N., Ma, X., & Jiang, L. (2013). The transcriptional control machinery as well as the cell wall integrity and its regulation are involved in the detoxification of the organic solvent dimethyl sulfoxide in Saccharomyces cerevisiae. *FEMS Yeast Research*, *13*(2), 200–218. https://doi.org/10.1111/1567-1364.12022

Zhang, W., Liu, H., Li, X., Liu, D., Dong, X. T., Li, F. F., Wang, E. X., Li, B. Z., & Yuan, Y. J. (2017). Production of naringenin from D-xylose with co-culture of E. coli and S. cerevisiae. *Engineering in Life Sciences*, *17*(9), 1021–1029. https://doi.org/10.1002/elsc.201700039

Zhao, E. M., Lalwani, M. A., Chen, J. M., Orillac, P., Toettcher, J. E., & Avalos, J. L. (2021). Optogenetic Amplification Circuits for Light-Induced Metabolic Control. *ACS Synthetic Biology*, *10*(5), 1143–1154. https://doi.org/10.1021/acssynbio.0c00642

Zhao, E. M., Zhang, Y., Mehl, J., Park, H., Lalwani, M. A., Toettcher, J. E., & Avalos, J. L. (2018). Optogenetic regulation of engineered cellular metabolism for microbial chemical production. *Nature*, *555*(7698), 683–687. https://doi.org/10.1038/nature26141

Zhou, K., Edgar, S., & Stephanopoulos, G. (2016). Engineering Microbes to Synthesize Plant Isoprenoids. In *Methods in Enzymology* (1st ed., Vol. 575). Elsevier Inc. https://doi.org/10.1016/bs.mie.2016.03.007

Zhou, Kang, Qiao, K., Edgar, S., & Stephanopoulos, G. (2015). Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature Biotechnology*, *33*(4), 377–383. https://doi.org/10.1038/nbt.3095